Published in final edited form as:

Psychon Bull Rev. 2014 August ; 21(4): 935–946. doi:10.3758/s13423-013-0578-x.

Is State-Trace Analysis an Appropriate Tool for Assessing the Number of Cognitive Systems?

F. Gregory Ashby

Department of Psychological & Brain Sciences University of California, Santa Barbara

Abstract

There is now much evidence that humans have multiple memory systems, and evidence is also building that other cognitive processes are mediated by multiple systems. Even so, several recent articles have questioned the existence of multiple cognitive systems and a number of these have based their arguments on results from state-trace analysis. State-trace analysis was not developed for this purpose, but rather to identify data sets that are consistent with variation in a single parameter. All previous applications have assumed that state-trace plots in which the data fall on separate curves rule out any model in which only a single parameter varies across the two tasks under study. Unfortunately, this assumption is incorrect. Models in which only one parameter varies can generate any type of state-trace plot, as can models in which two or more parameters vary. In addition, it is straightforward to show that single-system and multiple-systems models can both generate state-trace plots that are considered in the literature to be consistent with either one or multiple cognitive systems. Thus, without additional information, there is no empirical state-trace plot that supports any inferences about the number of underlying parameters or systems.

Introduction

The theory that humans have multiple memory systems became widely accepted within the field of cognitive neuroscience during the 1980's and 1990's (Eichenbaum & Cohen, 2001; Schacter, Wagner, & Buckner, 2000; Squire, 2004). Many other fields are now also debating whether multiple systems might mediate what previously was thought to be a unitary cognitive process. Included in this list are category learning (Ashby, Alfonso-Reese, Turken, & Waldron, 1998; Erickson & Kruschke, 1998), recognition memory (e.g., Yonelinas, 2002), and logical reasoning (Sloman, 1996). Although the evidence favoring multiple systems continues to grow in each of these areas, several recent articles have questioned the existence of multiple cognitive systems (e.g., Newell, Dunn, & Kalish, 2011; Nosofsky, Stanton, & Zaki, 2005; Stanton & Nosofsky, 2007). A number of these have based their arguments on results from state-trace analysis (Dunn, 2008; Dunn, Newell, & Kalish, 2012; Newell & Dunn, 2008; Newell, Dunn, & Kalish, 2010).

State-trace analysis (Bamber, 1979; Dunn & Kirsner, 1988) is a method for determining the number of cognitive processes or systems that were used to generate data from two separate

tasks or experimental conditions. The idea is to plot performance on the two tasks against one another and examine the resulting scatterplot. Based on the type of scatterplot that emerges, inferences are then made about the number of underlying processes or systems. State-trace analysis has been proposed as a more powerful alternative to dissociation logic. For example, Newell and Dunn (2008) went so far as to argue that state-trace analysis “overcomes all of the flaws of dissociation logic” (p. 285).

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. In this article, I will assume that values of $P(T_2)$ are plotted on the ordinate and values of $P(T_1)$ are on the abscissa. Current applications of state-trace analysis distinguish among four different types of plots. In a type 1 plot, the data all fall on a single strictly monotonic curve – that is, a curve that is either strictly increasing or strictly decreasing (e.g., as in Figure 1A below). In a type 2 plot, the data all fall on a single non-monotonic curve in which $P(T_2)$ is a function of $P(T_1)$ – that is, in which each value of $P(T_1)$ cooccurs with only a single value of $P(T_2)$ (e.g., as in Figure 1B below). In a type 3 plot, the data all fall on the same curve, but no restrictions are placed on the form of this curve. Finally, in a type 4 plot, the data fall on at least two separate curves (e.g., as in Figures 2C and 2D).

The origins of state-trace analysis date back to Bamber (1979), who proposed state-trace plots as a generalization of the ROC curve from signal detection theory. He showed that if only a single psychological process (or latent variable or parameter) was mediating performance in the two tasks, then all points on the state-trace plot must fall on the same (type 3) curve, whereas if two or more processes varied across tasks then any type of state trace plot is possible. Bamber did not specifically discuss curves of types 1 or 2. The focus on these was introduced by Dunn and Kirsner (1988). They argued that models in which only a single parameter or system varies almost always produce monotonic (type 1) state-trace plots and when they fail to produce a monotonic plot then they instead produce a function (type 2) plot. Thus, according to Dunn and Kirsner (1988), state-trace plots of types 1 or 2 are consistent with a single parameter or single system, whereas plots of type 4 are assumed to rule out a single varying parameter or single system. Some subsequent applications of state-trace analysis followed Bamber (1979) and attempted to test between single- (type 3) and multiple-curve (type 4) state trace plots (e.g., Loftus, 2002), but many recent state-trace studies follow Dunn and Kirsner (1988). For example, according to Prince, Brown, and Heathcote (2012) “determining whether there is one or more than one mediating latent variable is accomplished by determining whether the state-trace plot is monotonic” (p. 81).

This article shows that this logic is flawed. Although it is true that models that vary two or more parameters or systems across tasks can produce any type of state-trace plot, in the next section I show that, with empirical data, this is also true of single-system models that only vary one parameter across the two tasks. In fact, many such examples exist. The third section focuses on the question of whether state-trace analysis is appropriate for determining the number of underlying cognitive systems. There I show that the number of systems (or parameters) is irrelevant in a state-trace analysis. Single and multiple systems models can

both easily predict any type of state-trace plot. Thus, single system models that vary only one parameter, single system models that vary multiple parameters, and multiple systems models can all easily produce any type of empirical state-trace plot. As a result, in the absence of some extra information, it follows that no logical inferences about the number of underlying systems or parameters can be derived from any state-trace analysis.

State-Trace Analysis

As mentioned above, state-trace analysis plots performance in two tasks or conditions against one another. Inferences about the underlying model that generated the data are then made by examining the resulting scatterplot. To assess the validity of state-trace analysis, it is vital to understand the mathematical basis of the method. Ideally, we would like to know necessary and sufficient conditions on a model for it to produce a state-trace plot of each of the four types defined above. Unfortunately, no necessary conditions are known for any of the types. But we do know some sufficient conditions for plot types 1 and 2. First, assume that only one parameter (denoted by θ_j) in the model is allowed to vary across Tasks T_1 and T_2 . If so, then note that the model must predict that $P(T_1) = f_1(\theta_j)$ and $P(T_2) = f_2(\theta_j)$ for some mathematical functions f_1 and f_2 . Appendix 1 shows that a set of sufficient conditions for any model to produce a type 2 state-trace plot (i.e., in which performance on Task T_2 is a function of performance on Task T_1) is that only one parameter varies across the two tasks and f_1 is strictly monotonic in θ_j . In other words, the model must predict that performance on task T_1 is either a strictly increasing or strictly decreasing function of its single freely varying parameter θ_j . Appendix 1 also shows that a set of sufficient conditions for any model to produce a type 1 (i.e., monotonic) state-trace plot is that only one parameter varies across the two tasks and f_1 and f_2 are both monotonic in that parameter.

This latter result is well known (e.g., Dunn & Kirsner, 1988), but the former result is not. For example, Dunn and Kirsner's (1988) statement that "If the single-process model is true, performance on one task will be a function of performance on the other" (p. 98) implies that any single-process or single-parameter model should predict that $P(T_2)$ is a function of $P(T_1)$. This is indeed guaranteed if f_1 is monotonic, but otherwise it is not (e.g., counterexamples are given in Figures 1C and 1D). Dunn and Kirsner (1988) regarded monotonicity of f_1 and f_2 as "essential to psychological measurement" (p. 98). But in fact, many popular cognitive models include parameters that are non-monotonically related to performance. All that is needed for non-monotonicity is that the model predicts optimal performance for an intermediate value of the parameter, rather than an extreme value. Many examples of such non-monotonicities are found in the literature, including the following. 1) The response criterion in signal detection theory or the intercept of the decision bound in decision bound models (Maddox & Ashby, 1993). For example, increasing the response criterion from $-\infty$ to ∞ in signal detection models causes predicted accuracy to increase to a peak and then to decrease. 2) The response bias in the Luce-Shepard choice model (Luce, 1963; Shepard, 1957). The bias parameters in this widely used model sum to one. Unbiased responding occurs when all bias parameters are equal. Thus, in a two-alternative task, unbiased responding occurs when the bias parameter equals 0.5. As a result, in virtually all tasks, optimal accuracy will occur at some intermediate value of the bias parameter. 3) Dopamine-related parameters in almost any model that includes them (Ashby

et al., 1998; Durstewitz, & Seamans, 2002) – much data suggests that in almost all tasks affected by brain dopamine levels, performance is best when dopamine is at some intermediate level. 4) The learning rate parameter in almost all connectionist models – in almost all cases, optimal learning requires an intermediate learning rate. If the learning rate is too small, learning takes an inordinate amount of time and if the learning rate is too large the model tends to jump around so much that optimality is often missed. 5) Parameters that specify the proportion of attention allocated to each perceptual dimension, as in the generalized context model (GCM; Nosofsky, 1986) and certain multidimensional scaling models (e.g., Carroll & Chang, 1970) – in some tasks optimal performance requires attention to only one dimension, but many tasks require that some attention is allocated to two or more dimensions. In such cases, optimal performance occurs at intermediate values of these attention parameters. Many other examples could also be named. The key point is that it is very common for models to postulate optimal performance at some intermediate level of a parameter. In all such cases, the monotonicity assumption needed to guarantee a state-trace plot of types 1 or 2 is violated. Next I show that in the presence of such non-monotonicities, single-parameter models naturally predict empirical state-trace plots of type 4.

To illustrate the types of state-trace plots that can be produced by a single system model in which only one parameter varies across tasks, consider the GCM (Nosofsky, 1986). This is an exemplar model of categorization, which assumes that the stimulus in a categorization experiment is assigned to the category to which it is most similar, with category similarity defined as the sum of the similarities of the stimulus to each category exemplar. When the stimuli vary on two perceptual dimensions, the model has two attention parameters – the total amount of attention allocated to the task, denoted by c , and the proportion of the total attention allocated to dimension 1, denoted by w (so the proportion of attention allocated to dimension 2 is $1 - w$, and thus $0 < w < 1$). This is a good model to study because the GCM predicts that performance in all categorization tasks increases monotonically with c , but not necessarily with w . The predictions for w depend on the category structure. If dimension 1 is the only dimension relevant to the categorization decision, then performance will increase monotonically with w . But if both dimensions are relevant, then accuracy will increase until w reaches its optimal value and then decrease thereafter.

Figure 1 shows four different state-trace plots produced by this model when only a single parameter varies across the two tasks. In every case, T_1 and T_2 are categorization tasks in which the stimuli vary on two dimensions. The only difference between the two tasks is the category structure. In Figure 1A, both dimensions are equally important in task T_1 (so the optimal value of w is .5), whereas only dimension 1 is relevant in task T_2 (so the optimal value of w is 1). The only parameter that varies across tasks is c . Since the GCM predicts that accuracy monotonically improves in both tasks as c increases, the sufficient conditions for monotonicity are met, so as predicted, the state-trace plot is of type 1.

In Figures 1B, 1C, and 1D, c is held constant and the only parameter that varies is w . Figure 1B uses the same tasks as in Figure 1A, except now the task that requires equal attention to the two dimensions is labeled T_2 , whereas the task where only dimension 1 is relevant is labeled T_1 . Note that the state-trace plot is of type 2 (i.e., a function) because performance on task T_1 improves monotonically with w (optimal performance occurs when

$w = 1$) but performance on task T_2 does not (optimal performance occurs when $w = .5$). Figure 1C plots exactly the same data as 1B, but the axes and task names are reversed¹. Now accuracy monotonically increases with w on task T_2 , but not on task T_1 . Even though performance does increase monotonically on task T_2 with w , note that performance on Task T_2 is not a function of performance on Task T_1 [i.e., since there are many cases where the same value of $P(T_1)$ is associated with more than one value of $P(T_2)$]. Thus, Figure 1C contradicts the claim of Dunn and Kirsner (1988) that “If the single-process model is true, performance on one task will be a function of performance on the other” (p. 98). It is true that all points in the Figure 1C state-trace plot fall on the same contour, and so, as predicted by Bamber (1979), the Figure 1C state-trace plot is of type 3. Even so, in any empirical application only a fairly small set of points from the plot can be estimated. For example, suppose two groups of participants are run, that the GCM is correct, and that the only difference between the groups is that one group allocates relatively more attention to dimension 1 than the other group. The dotted line curve in Figure 1C could describe the performance of the group that allocates a greater proportion of attention to dimension 1 and the solid line curve could describe the group that allocates a smaller proportion to dimension 1. Because these two curves fall on different contours, all current applications of state-trace analysis would incorrectly conclude that these data could not have been produced by a model in which only one parameter varies across both groups.

Figure 1D illustrates what happens when only one parameter varies, but performance on both tasks is optimized at an intermediate value of that parameter. In 1D, task T_1 is identical to task T_1 from Figures 1A and 1C (and identical to task T_2 from Figure 1B). So in task T_1 the optimal value of w is $w = .5$. Task T_2 however, is different from any other tasks used in this analysis. The only difference though, is that the categories were rotated slightly so that both dimensions are relevant, but dimension 1 is more important than dimension 2 (see Appendix 2 for details on all tasks used to generate Figure 1). This means that the GCM predicts optimal performance in task T_2 when w is greater than .5 but less than 1. Note that now the curve almost closes back on itself. As before, if two groups are run that are identical except for the values of w they choose, one must almost inevitably observe two disconnected curves (e.g., as with the heavy dotted and black lines). Again, state-trace analysis would incorrectly reject a single parameter interpretation in such a case.

It is vital to point out that the troubling state-trace plots shown in Figures 1C and 1D are not due to some unique property of the GCM. For example, consider Figure 1D. Each point on the curve is associated with a different value of w . As one moves around the curve through different values of w , note that as it must, accuracy increases on each dimension and then decreases. It doesn't matter what the model is or what parameter is varying. Consider any model in which only one parameter varies and the model predicts that performance on the two tasks increases as the parameter increases to some intermediate value and then decreases thereafter. Such a model must necessarily produce a state-trace plot that is qualitatively similar to Figure 1D (i.e., where performance increases and then decreases on both

¹Stimulus presentation order was randomized in Figures 1B and 1C, so even though the tasks and model are identical, the two curves are not exact rotations of each other.

dimensions). As a result, it is quite likely that state-trace analysis would mistakenly conclude that two or more parameters must have varied to produce that plot.

State-trace Analysis and Multiple Cognitive Systems

Figure 1 makes it clear that a single-process or single-system model in which only one parameter is varying can easily generate any type of state-trace plot. But what about models that postulate multiple systems? The literature suggests that multiple systems models should always produce multiple curve (type 4) state-trace plots. This section investigates this prediction, within the context of category learning, where state-trace analysis is especially prevalent and where there are simple, multiple systems models to examine.

Most of the applications of state-trace analysis to category learning have compared performance in rule-based (RB) and information-integration (II) category-learning tasks. In RB tasks, the categories can be learned via some explicit reasoning process. Typically, the rule that maximizes accuracy (i.e., the optimal strategy) is easy to describe verbally (Ashby et al., 1998). In the most common applications, only one stimulus dimension is relevant, and the participant's task is to discover this relevant dimension and then to map the different dimensional values to the relevant categories. In II tasks, accuracy is maximized only if information from two or more stimulus components (or dimensions) is integrated perceptually (Ashby & Gott, 1988). In most cases, the optimal strategy in II tasks is difficult or impossible to describe verbally (Ashby et al., 1998). Verbal rules may be (and sometimes are) applied but they lead to suboptimal performance because they produce a maladaptive focus on only one dimension of variation. In Figure 1A, task T_2 is an RB task and task T_1 is an II task.

Briefly, when applied to category learning, state-trace analysis typically focuses on a scatterplot of accuracy in an RB task against accuracy in an II task for each learning block and condition (e.g., see Figure 2 below). As in other applications, if the data points all fall on the same curve that is either monotonically increasing or decreasing (i.e., a type 1 state-trace) or a function (type 2) then a single system is inferred and if the data fall on two separate curves (a type 4 state-trace) then multiple systems are inferred.

Illustrating this method, Newell et al. (2010) explored the effects of a concurrent working memory load on RB and II performance. Several previous studies had shown that a simultaneous concurrent load interferes more with RB learning than with II learning (Waldron & Ashby, 2001; Zeithamova & Maddox, 2006). Newell et al. began by replicating the Zeithamova and Maddox (2006) experiment and found evidence for a two-dimensional (type 4) state-trace plot, thereby replicating another empirical aspect of multiple-systems theory. But then they went on to institute a 65%-correct learning criterion that excluded many participants. Now they found evidence for a one-dimensional (type 1) state-trace plot. In additional studies they examined a less taxing concurrent task and a novel concurrent working-memory task (maintaining the 65%-accuracy learning criterion). In both cases, they argued that their results are consistent with a type 1 state-trace plot and therefore with a single category-learning system.

Dunn et al. (2012) explored the dissociation that delayed feedback impairs II category learning more than RB category learning (Maddox, Ashby, & Bohil, 2003; Maddox & Ing, 2005), and the dissociation that providing full versus minimal feedback enhances RB but not II learning (Maddox, Love, Glass, & Filoteo, 2008). Across four experiments, Dunn et al. (2012) compared the effect of a strong mask (a Gabor patch similar to the stimuli in the experiment) or weak mask (a simple pattern mask) given after the response on each trial but before the feedback. The former and latter conditions appeared to produce a two-dimensional (type 4) and a one-dimensional (type 1) state-trace plot, respectively. Dunn et al. also found that the feedback-delay effect was reduced when full feedback instead of minimal feedback was used (even with the Gabor mask). These findings are potentially quite useful, but the key question considered here is whether the state-trace analyses that were used are appropriate for answering the question of whether the RB and II tasks are learned by the same or different systems.

When attempting to determine the number of underlying cognitive systems, a state-trace analysis seems to have the most potential when the candidate multiple systems model predicts that different systems should dominate in the two tasks under study. For example, COVIS (Ashby et al., 1998) predicts that an explicit system should dominate in RB tasks and a procedural learning system should dominate in II tasks. Thus, a dual systems model could include three different types of parameters: 1) parameters that only affect performance in task T_1 ; 2) parameters that only affect performance in task T_2 ; and 3) parameters that affect performance in both tasks. In contrast, single system models should only include parameters of the third type, since such models predict that all tasks are performed in the same way. Appendix 1 shows that a dual systems model predicts a (type 1) monotonically increasing state-trace plot if the following conditions are met: 1) only one parameter is varying across tasks and that parameter affects performance in both tasks, and 2) $P(T_1)$ and $P(T_2)$ both monotonically increase with increases in the single varying parameter. If the single varying parameter affects only task T_1 , then the state-trace plot must be a single horizontal line (because under these conditions $P(T_2)$ must be a constant). In contrast, if the single varying parameter only affects task T_2 , then the state-trace plot must be a single vertical line (because under these conditions $P(T_1)$ is a constant).

Thus, multiple systems models can also easily predict monotonically increasing state-trace plots. The only extra requirement not needed for single system models is that the single parameter that varies across tasks affects performance in both tasks. Therefore, if one parameter varies across tasks, it is irrelevant whether that parameter is seated in a model that posits a single cognitive process, multiple cognitive processes within a single system, or multiple cognitive processes within multiple functionally separate systems. So long as performance on each task monotonically increases with the value of that parameter then the state-trace plot must be of type 1. Thus, even a monotonically increasing state-trace plot (type 1) provides no information about whether there are one or multiple underlying systems. In fact, it is straightforward to show that there is no state-trace result that is diagnostic of one versus multiple systems.

For example, consider Figure 2. Here I have plotted RB versus II accuracy for two different single-system and dual-systems category-learning models. The RB categories require a one-

dimensional decision rule, whereas the optimal bound in the II condition is diagonal. The single-system model is the same GCM used to generate Figure 1. The dual-systems model is a simplification of COVIS (Ashby et al., 1998). Unlike COVIS, however, this model makes a hard switch between systems – that is, it uses the explicit system on every trial until it gives up on explicit strategies. Then it switches to the procedural system². It has three parameters: a learning rate in each system, and a threshold for switching between systems. As in COVIS, the learning rate in the explicit system affects both tasks. It affects the II task because switching to the procedural system occurs later when the explicit system is responding accurately. In contrast, the procedural system learning rate and the threshold for switching affect II performance but not RB performance. In the RB task, the explicit system discovers the optimal strategy and so a switch to the procedural system never occurs. Predictions from the models are derived for two different experimental conditions – one in which the category learning occurs under single-task conditions, and one in which it occurs while participants are simultaneously engaged in a dual task (e.g., as in Waldron & Ashby, 2001). See Appendix 2 for details of these simulations.

Figures 2A and 2C show two new state trace plots predicted by the GCM. Figure 1 showed that when only a single parameter varies, a variety of different types of plots are possible. In Figure 2A, three GCM parameters were allowed to vary across tasks and conditions – the same two attention parameters described earlier (i.e., c and w) plus a parameter that measures the ability of the participant to make optimal use of the available similarity computations (i.e., γ ; Ashby & Maddox, 1993). For example, when γ is large, the participant reliably assigns the stimulus to the category with the greater summed similarity. When γ is small, the participant frequently assigns the stimulus to the category with the lower summed similarity. In Figure 2A, it was assumed that the primary effect of the dual task was to reduce γ , although it was also assumed that the dual task increased c and improved attentional learning. Thus, in this case, the dual task had discrepant effects on different parameters. The important point though, is not whether this is the best model of dual-task performance, but that even though three parameters are varying, the state trace plot is of type 1 (single, monotonic curve). Thus, Figure 2A confirms the earlier statement that a type 1 plot is not evidence that only one parameter is varying.

Two GCM parameters were manipulated to produce Figure 2C – c and w . Specifically, the dual task was assumed to reduce both discriminability (i.e., c) and attentional learning. More specifically, it was assumed that the dual task reduced the ability of the participants to discover that all attention should be allocated to dimension 1 in the RB task (i.e., so the dual-task also affected w). The curves are not on the same contour (the plot is of type 3) because the effects of reducing attentional learning are not the same in the RB and II tasks.

The right column of Figure 2 shows predictions of the dual-systems model. In the top right panel the only parameter that varied was the explicit-system learning rate, and this parameter was assumed to be reduced by the dual task. Since this parameter affects

²The assumption of a hard switch was incorporated because recent evidence (Ashby & Crossley, 2010) strongly suggests that the trial-by-trial switching assumption made by the original version of COVIS (Ashby et al., 1998) is incorrect. Instead, the evidence supports the assumption that in II tasks, participants perseverate with explicit rules until they give up on explicit strategies and switch to procedural strategies.

performance in both tasks, note that as predicted, the resulting state-trace plot is monotonically increasing (type 1). This figure verifies the statement made earlier that a state-trace plot with all points on the same contour (e.g., as in a type 1 plot) is not evidence that a single system is mediating performance in the two tasks. Note that the model did not produce a single contour by mimicking a single system. This might be the case, for example, in a dual systems model where the response blends (or averages) the outputs of the two systems. In such models the same process is used on every trial to select a response. But the dual systems model used to generate Figure 2 did not blend responses from the two systems. On each trial, only one system was used to select the response, and in the II task, control was passed from one system to the other at some intermediate point during learning. Finally, in the bottom right panel, two parameters varied across tasks: the explicit system learning rate and the threshold for switching from the explicit system to the procedural system. I assumed that the dual task slowed learning in the explicit-system and caused earlier switching to procedural strategies. Again, as expected, the model generates a type 4 state-trace plot.

Figure 2 clearly shows that there is no state-trace plot that provides empirical evidence, one way or the other, about the number of category-learning systems. A single-system model can easily predict state-trace plots on the same or different contours (i.e., types 1 or 4), and so can a dual-systems model. It is important to note that the multiple-systems model used to produce Figure 2 is hardly exotic. In fact, it is the simplest multiple-systems model that I could create. For example, COVIS (Ashby et al., 1998) is considerably more complex. Furthermore, COVIS has a number of parameters that will affect performance in both RB and II tasks (e.g., all of its explicit system parameters). So COVIS can also easily produce type 1 state-trace plots. By similar logic, it seems likely that many reasonably plausible multiple-systems models will be able to produce state-trace plots that fall on a single curve. But the number of multiple-systems models capable of producing such plots is logically irrelevant. Figure 2 shows that even a simple multiple-systems model can produce single-curve state-trace plots, and therefore the existence of a single-curve plot provides no logical basis to make any inference about the number of underlying learning systems. Figure 2 also shows that type 4 state-trace plots are easily produced by both single- and multiple-systems models. Thus, there is no state-trace plot that provides any logical basis to conclude that the data were generated by one versus multiple learning systems.

Figures 1 and 2 together show that without more information, little can be learned from a state-trace analysis. A result where state-trace plots are on different contours (i.e., a type 4 plot) could have been generated by a model in which only one parameter is varying (e.g., as in Figure 1D) or by a model in which more than one parameter is varying (as in the bottom two panels of Figure 2). Similarly, a strictly monotonic state-trace plot (type 1) could be generated by a model in which only one parameter is varying (e.g., as in Figures 1A and 2B) or by a model in which more than one parameter is varying (as in Figure 2A).

So what is the role of state-trace analysis in cognitive research? There are two possible extra steps a researcher could take to restore value to a state-trace analysis, but unfortunately, both are problematic. First, if the state-trace plot shows multiple curves, then data from new groups could be collected in an attempt to determine whether the two curves are different segments of the same curve (as in Figures 1C and 1D) or are actually two separate curves (as

in Figures 2C and 2D). This could be a challenging task. For example, in Figure 1D it would require identifying a participant group where accuracy is almost precisely 0.9 in task T_1 . Second, one could somehow try to confirm that performance in the two tasks is monotonic in the single parameter that is believed to be manipulated. Unfortunately, acquiring such knowledge is also not straightforward. Ideally, one might plot the performance of each group as a function of that single parameter on each task separately and then examine whether all of these plots are monotonically increasing or they are all monotonically decreasing. The problem of course, is that with empirical data, one would generally have no idea what that critical parameter was. And even if a candidate parameter could be identified, in order to plot performance against values of that parameter, one would need some independent method of estimating the value of the unknown parameter for each participant. Achieving this standard seems unlikely.

Suppose though, that one of these two extra efforts proved successful. In this case, the single ability of state-trace analysis would be to identify data sets that might be accounted for by a model that assumes only one parameter is varying across all the relevant tasks and conditions. In the field of categorization at least, such data sets should be extremely rare, and so even if these difficult conditions were met, the contributions of state-trace analysis would be minimal. To see this, consider again Figure 2. Here performance is plotted in a one-dimensional RB task against performance in a two-dimensional II task. Virtually all current categorization models predict that in such cases the distribution of perceptual attention to the two stimulus dimensions will vary across tasks. In particular, virtually all models predict that participants will allocate more perceptual attention to the single relevant dimension in the RB task than to the irrelevant dimension, whereas in the II task attention will be more evenly distributed across the two relevant dimensions. Thus, almost any model will need to vary at least one attentional-allocation parameter to account for the RB and II data simultaneously. In the GCM this parameter is w . In fact, one of the greatest contributions of exemplar theory, and especially of the GCM, is to show that whenever the category structure changes, one should expect changes in the allocation of perceptual attention (Nosofsky, 1986). The obvious problem with this conclusion is that, as we have already seen, categorization accuracy is non-monotonic with w in II tasks. Thus, virtually all current categorization theories predict that state-trace plots of RB versus II performance should fail to meet the sufficient conditions required for a type 1 monotonic plot – even those that assume categorization is mediated by only a single system. Instead, the best one might hope for is a plot that is qualitatively similar to Figure 1B or 1C.

Even if one ignores this problem though, there still remain serious challenges to state-trace logic when applied to categorization. Since virtually all models predict that an attention-weight parameter (e.g., w) will differ across RB and II tasks, if any other parameter varies across the tasks or conditions then state-trace logic predicts that a type 4 plot should be expected (i.e., plots that fall on different contours). Thus, we would expect plots on different contours if learning rates differed across conditions, or working memory demand, or perceptual noise, or criterial noise, or sensitivity to positive versus negative feedback, or response bias, or any one of many other possible attributes. For example, Figure 2 assumes participants completed RB and II tasks under single- and dual-task conditions. Waldron and

Ashby (2001) included this dual-task condition as a way to reduce the working memory capacity available for categorization. Thus, even if there was a single system and working memory affected RB and II tasks equally, the resulting state-trace plots should not lie on the same contour because at least two parameters would be varying across the tasks and conditions (i.e., w and at least one working memory parameter). Similarly, Maddox et al. (2003) reported results of RB and II task performance with and without a feedback delay. The delay was introduced in an attempt to reduce the salience of the trial-by-trial feedback. Presumably, any state-trace plots created from these data should produce delay and non-delay plots that do not lie on the same contour because any model capable of fitting these data should have to vary an attention-weight parameter to account for RB versus II differences and at least one parameter that accounts for feedback salience differences between the immediate and delay conditions.

State-trace analysis is most powerful in tasks that are both perceptually and cognitively simple, where it is reasonable to expect that only a single parameter might be varying across a variety of conditions. The Yes-No signal detection task is a good example. Compared to categorization, it is far less perceptually and cognitively demanding, and there is good theoretical and empirical reason to believe that if we vary payoffs or base-rates, but hold noise and signal intensity constant, then we should expect variation in only one parameter³ (i.e., the decision criterion). But this is not true in categorization. For example, Nosofsky (1986) had the same participants complete an identification task and then four separate categorization tasks that assigned the identification stimuli to two contrasting categories in four different ways. He then used model fits to the identification data to account for the results of the four different categorization tasks. Even after taking full advantage of the identification data, he still had to allow three separate GCM parameters to vary freely across the four sets of categorization data to achieve satisfactory fits. And this study did not include any separate experimental manipulation, such as repeating the categorization conditions with and without a dual task, or with and without a feedback delay. Thus, in Bayesian terminology, the prior probability that performance in two separate categorization tasks can be accounted for by a model that varies only one parameter across tasks and conditions is extremely low. In other words, state-trace analysis seems to only have the potential to disconfirm a scenario that we expect to almost never happen in human categorization.

If one accepts this argument, then why have there been several reports of type 1 categorization state-trace plots (i.e., one monotonic curve) (Newell et al., 2010, 2011; Dunn et al., 2012)? If more than one parameter is varying then state-trace analysis predicts a type 4 plot (separate curves). One possibility is that there is no established statistical test of the null hypothesis that two scatterplots fall on the same contour. Newell and Dunn (2008) have begun working on this problem, but they acknowledge that this should still be considered an open question. The method that they advocate treats the one-dimensional state-trace conclusion as the null hypothesis and uses a bootstrap technique to assess significance. But

³Signal detection theory predicts that overall accuracy is a non-monotonic function of the decision criterion, but note that an ROC curve (perhaps the best known state-trace plot) plots the proportion of hits against the proportion of false alarms. Signal detection theory predicts that these two proportions both decrease monotonically with the value of the decision criterion. Therefore, in the case of the ROC curve, signal detection theory meets the sufficient conditions required for a monotonic (type 1) state-trace plot.

the power of this technique is unknown. In fact, computing its power is a challenging statistical problem, because there is no way to know in advance what a state-trace plot should look like from a reasonable two-dimensional model. For these reasons, I suspect that the currently used methods lack statistical power. If so, then type II errors are likely. In state-trace analysis, a type II error would be incorrectly deciding that the data are consistent with variation in one parameter.

Conclusions

State-trace analysis (Bamber, 1979; Dunn & Kirsner, 1988) has recently been used to test whether various cognitive processes are mediated by one or multiple functionally separate systems. The method was not developed for this purpose, but rather to test whether a data set is consistent with variation in one or more parameters. Previous applications of state-trace analysis have always assumed that type 4 plots (i.e., separate curves) rule out models in which only one parameter varies across conditions. Figure 1 shows that this assumption is clearly false, at least when the test is made with empirical data. Furthermore, Figure 2 shows clearly that models in which more than one parameter are varying and dual systems models can both also easily predict type 1 plots (i.e., single monotonically increasing curves). Thus, unless one has some extra knowledge (e.g., about the relationship between the relevant parameters and task performance), it is unlikely that any logical conclusions can be drawn from a state-trace analysis⁴. Empirical state-trace plots of types 1, 2, or 4 can be produced by models in which only one parameter is varying, models in which more than one parameter is varying, and models with more than one system.

For these reasons, state-trace analysis is not an appropriate tool for deciding whether humans have one or multiple cognitive systems, and it is problematic even when the question is whether one or multiple parameters are varying across conditions. 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 (e.g., Dunn & Kirsner, 1988). How then, should this important question be addressed? In my opinion, the only way to answer this question is via a converging operations approach. It is vital to consider *all* available data when addressing this question. For example, suppose a multiple-systems model predicts ten new empirical dissociations *a priori* and that all ten are empirically supported. Suppose further that no single-system model is known that can account for all these results. Collectively, these ten dissociations should be interpreted as strong support for multiple systems, even though 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. Science is a cumulative endeavor. And 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.

⁴The possible exception is a plot in which the data form a curve that loops back on itself (as in Figure 1D).

Acknowledgments

I thank W. Todd Maddox and J. David Smith for helpful comments on an earlier draft of this article. This work was supported in part by AFOSR grant FA9550-12-1-0355, NIH (NINDS) Grant No. P01NS044393, and by Grant No. W911NF-07-1-0072 from the U.S. Army Research Office through the Institute for Collaborative Biotechnologies.

Appendix 1

Note that any mathematical model can be defined as a collection of mathematical functions, with one function for every experiment and dependent variable for which the model makes a prediction. Each function in this collection assigns a specific value of the dependent variable to specific values of all the model's parameters. For example, consider a model with r free parameters, $\theta_1, \theta_2, \dots, \theta_r$. Then the function f_i might predict the performance of the model in task T_i on the dependent variable of interest, which can be denoted as $P(T_i)$. In other words,

$$P(T_i) = f_i(\theta_1, \theta_2, \dots, \theta_r).$$

Note that the only assumption incorporated into this definition of a model is that each f_i is a function – that is, each f_i assigns only one value of $P(T_i)$ to each combination of $\theta_1, \theta_2, \dots, \theta_r$.

A state-trace analysis plots $P(T_2)$ (e.g., on the ordinate) against $P(T_1)$ (on the abscissa). The question is what can be learned about the underlying model from examining such plots? For example, one might ask under what conditions $P(T_2)$ is a function of $P(T_1)$ [i.e., so that each value of $P(T_1)$ occurs with only one value of $P(T_2)$]. In other words, under what conditions does there exist a function F , such that

$$P(T_2) = F[P(T_1)]?$$

And for example, under what conditions is F strictly increasing?

A general solution to this problem does not appear possible, but some strong sufficient conditions are easily derived (Dunn & Kirsner, 1988). Consider the case where all parameters are fixed save one. Assume the parameter that varies is θ_j . Thus, $P(T_1) = f_1(\theta_j)$ and $P(T_2) = f_2(\theta_j)$. Next assume that $f_1(\theta_j)$ is strictly monotonic – that is, f_1 is either a strictly increasing or strictly decreasing function of θ_j . Under these conditions, f_1 has an inverse f_1^{-1} , and the inverse is itself a function. Therefore

$$\theta_j = f_1^{-1}[P(T_1)],$$

which implies that

$$P(T_2) = f_2\{f_1^{-1}[P(T_1)]\}.$$

A function of a function is itself a function (i.e., $f_2 \circ f_1^{-1}$ is a function). Therefore, under these conditions, the state-trace plot is a function (a type 2 state-trace). Even so, the state-trace plot might not be strictly monotonic (i.e., type 1). To guarantee a state-trace plot in which all points fall on one strictly monotonic curve, it suffices to add the extra assumption that f_2 is also strictly monotonic (and therefore, f_1 , f_1^{-1} , and f_2 are all strictly monotonic). For example, if f_1 and f_2 are both strictly increasing functions then $f_2 \circ f_1^{-1}$ is a strictly increasing function and all points on the state-trace plot therefore must necessarily fall on a single strictly increasing curve.

If two parameters θ_j and θ_k are both varying then $P(T_1) = f_1(\theta_j, \theta_k)$ and $P(T_2) = f_2(\theta_j, \theta_k)$, for two functions f_1 and f_2 . In this case, there are no conditions under which f_1 has an inverse (since the inverse would have to map $\mathbb{R} \rightarrow \mathbb{R}^2$). Thus, no model in which two or more parameters are varying can meet these sufficient conditions.

Multiple systems models assume, by definition, that different cognitive systems are used in different tasks and conditions. For example, a dual systems model might assume that one system dominates in task T_1 and a different system dominates in task T_2 . Thus, a multiple systems model could be defined as a special type of mathematical model (i.e., as defined above) in which any two of the following sets are nonempty: Set 1) model parameters that only affect performance in task T_1 (denoted by $\{\alpha_1, \alpha_2, \dots\}$); Set 2) parameters that only affect performance in task T_2 (denoted by $\{\beta_1, \beta_2, \dots\}$); and Set 3) parameters that affect performance in both tasks (denoted by $\{\theta_1, \theta_2, \dots\}$). By the same analogy, a single system model should only include parameters in set 3, since such models predict that all tasks are performed in the same way.

Thus, for any dual systems model there must exist functions f_1 and f_2 such that

$$P(T_1) = f_1(\alpha_1, \alpha_2, \dots, \theta_1, \theta_2, \dots) \text{ and } P(T_2) = f_2(\beta_1, \beta_2, \dots, \theta_1, \theta_2, \dots).$$

Such a model predicts a (type 1) monotonically increasing state-trace plot if the following conditions are met: 1) Only one parameter is varying across tasks T_1 and T_2 and that parameter is a member of the set $\{\theta_1, \theta_2, \dots\}$. 2) Suppose the single varying parameter is θ_j . Then $P(T_1)$ and $P(T_2)$ both monotonically increase with increases in θ_j . The proof of this is identical to the proof given above. Note that if only one parameter is varying but it is a member of the set $\{\alpha_1, \alpha_2, \dots\}$, then the state-trace plot must be a single horizontal line (because under these conditions $P(T_2)$ must be a constant). In contrast, if the single varying parameter is a member of the set $\{\beta_1, \beta_2, \dots\}$, then the state-trace plot must be a single vertical line (because under these conditions $P(T_1)$ is a constant).

Appendix 2

Category Structures

In all panels of Figures 1 and 2, both tasks were categorization tasks with two categories A and B, each composed of three exemplars that varied on two perceptual dimensions. In the RB conditions (task T_2 in Figures 1A and 1C, task T_1 in Figure 1B, and the ordinate in all

Figure 2 panels), the A exemplars had coordinates (0,0), (0,1), and (0,2), whereas the B exemplars had coordinates (1,0), (1,1), and (1,2). In all II conditions in Figure 1 (except task T₂ in Figure 1D) and Figure 2, the A exemplars had coordinates (0,1), (1,2), and (2,3), whereas the B exemplars had coordinates (1,0), (2,1), and (3,2). Thus, the category bound has a slope of 1, so the optimal strategy is to allocate equal attention to the two perceptual dimensions. Task T₂ of Figure 1D was also an II task, but in this case the category bound had a slope of 2, so both dimensions are relevant, but the optimal strategy is to allocate more attention to dimension 1 than dimension 2. The stimuli in this condition were created by rotating the coordinates of the stimuli from the RB condition. Such a rotation guarantees that between- and within-category similarity are identical in all categorization conditions.

Single-System Model

The single-system model was the generalized context model (GCM; Nosofsky, 1986). The model had three parameters: (1) c , which could be interpreted as the total amount of attention allocated to the task, (2) an attention weight w , which is the proportion of total attention allocated to dimension 1, and (3) γ , which is a measure of response determinism. The γ parameter, introduced by Ashby and Maddox (1993), is an exponent on each summed similarity. When $\gamma = 1$ the model is the same as the original GCM, when $\gamma > 1$ it responds more deterministically, and when $\gamma < 1$ it responds more probabilistically. In all applications except Figure 2A, $\gamma = 1$. The model assumed no response bias. In Figure 1A, w was fixed to 0.5 and the only parameter that varied across tasks was c . In Figures 1B, 1C, and 1D, c was fixed at .05 and only w was allowed to vary. In Figure 2A, the dual task was assumed to decrease γ from $\gamma = 3.8$ to $\gamma = .45$, and to increase c . The attention weight w was set to .5 in all conditions except the RB task under dual task conditions, where it was set to .91. In Figure 2C, the dual task was assumed to decrease c and to impair attentional learning. Specifically, under single-task conditions, w was set to the optimal values of 1 in the RB task and 0.5 in the II task. Under dual-task conditions, w was set to 0.5 in both tasks.

Dual-Systems Model

The dual-systems model used to generate Figures 2B and 2D was a simplification of COVIS (Ashby et al., 1998). In the RB task, the model switched back-and-forth between a horizontal and a vertical decision bound. As training progressed, the use of the incorrect horizontal bound exponentially decreased and the use of the correct vertical bound exponentially increased. In the II task, the model switched back-and-forth between guessing and a vertical decision bound. As training progressed, the frequency of guessing decreased exponentially and the use of the vertical bound exponentially increased. The best one-dimensional rule (either a vertical or horizontal bound) yields an accuracy of 67% correct in the II task, so during this phase of training the model could not exceed 67% correct. After persisting for some time with this reduced accuracy, the model switched to its procedural system. After this switch trial, accuracy exponentially increased to 100%. The model has 3 parameters, a learning rate in each system (i.e., the exponential rate), and a threshold for switching from the explicit system to the procedural system (i.e., a tolerance on poor performance). In Figure 2B, it was assumed that the only effect of the dual task was to reduce the learning rate in the explicit system. In Figure 2D, the dual task was again

assumed to slow the explicit system learning rate, but now it was also assumed to reduce the threshold on poor performance. Reducing this threshold causes the model to switch to the procedural system on an earlier trial.

References

- Ashby FG, Alfonso-Reese LA, Turken AU, Waldron EM. A neuropsychological theory of multiple systems in category learning. *Psychological Review*. 1998; 105:442–481. [PubMed: 9697427]
- Ashby FG, Crossley MJ. Interactions between declarative and procedural-learning categorization systems. *Neurobiology of Learning and Memory*. 2010; 94:1–12. [PubMed: 20304078]
- Ashby FG, Gott RE. Decision rules in the perception and categorization of multidimensional stimuli. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. 1988; 14:33–53.
- Ashby FG, Maddox WT. Relations between prototype, exemplar, and decision bound models of categorization. *Journal of Mathematical Psychology*. 1993; 37:372–400.
- Bamber DA. State-trace analysis: A method of testing simple theories of causation. *Journal of Mathematical Psychology*. 1979; 19:137–181.
- Carroll JD, Chang JJ. Analysis of individual differences in multidimensional scaling via an N-way generalization of “Eckart-Young” decomposition. *Psychometrika*. 1970; 35:283–319.
- Dunn JC. The dimensionality of the remember-know task: A state-trace analysis. *Psychological Review*. 2008; 115:426. [PubMed: 18426296]
- Dunn JC, Kirsner K. Discovering functionally independent mental processes: The principle of reversed association. *Psychological Review*. 1988; 95:91–101. [PubMed: 3353477]
- Dunn J, Newell B, Kalish M. The effect of feedback delay and feedback type on perceptual category learning. *Journal of Experimental Psychology: Learning, Memory and Cognition*. 2012; 38:840–859.
- Durstewitz D, Seamans JK. The computational role of dopamine D1 receptors in working memory. *Neural Networks*. 2002; 15:561–572. [PubMed: 12371512]
- Eichenbaum, H.; Cohen, NJ. From conditioning to conscious recollection: Memory systems of the brain. Oxford University Press; New York: 2001.
- Erickson MA, Kruschke JK. Rules and exemplars in category learning. *Journal of Experimental Psychology: General*. 1998; 127:107–140. [PubMed: 9622910]
- Loftus, GR. Analysis, interpretation, and visual presentation of experimental data.. In: Pashler, H.; Wixted, J., editors. *Stevens' handbook of experimental psychology: Volume 4 Methodology in experimental psychology*. 3rd ed.. Wiley; New York: 2002. p. 339-390.
- Luce, RD. Detection and recognition.. In: Luce, RD.; Bush, RR.; Galanter, E., editors. *Handbook of mathematical psychology*. Vol. 1. Wiley; New York: 1963. p. 103-190.
- Maddox WT, Ashby FG. Comparing decision bound and exemplar models of categorization. *Perception & Psychophysics*. 1993; 53:49–70. [PubMed: 8433906]
- Maddox WT, Ashby FG, Bohil CJ. Delayed feedback effects on rule-based and information-integration category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. 2003; 29:650–662.
- Maddox WT, Ing AD. Delayed feedback disrupts the procedural-learning system but not the hypothesis testing system in perceptual category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. 2005; 31:100–107.
- Maddox WT, Love BC, Glass BD, Filoteo JV. When more is less: Feedback effects in perceptual category learning. *Cognition*. 2008; 108:578–589. [PubMed: 18455155]
- Newell BR, Dunn JC. Dimensions in data: Testing psychological models using state-trace analysis. *Trends in Cognitive Sciences*. 2008; 12:285–290. [PubMed: 18606559]
- Newell BR, Dunn JC, Kalish M. The dimensionality of perceptual category learning: A state-trace analysis. *Memory & Cognition*. 2010; 38:563–581. [PubMed: 20551337]
- Newell BR, Dunn JC, Kalish M, Ross BH. Systems of category learning: Fact or fantasy? *The Psychology of Learning & Motivation*. 2011; 54:167–215.

- Nosofsky RM. Attention, similarity, and the identification categorization relationship. *Journal of Experimental Psychology: General*. 1986; 115:39–57. [PubMed: 2937873]
- Nosofsky RM, Stanton RD, Zaki SR. Procedural interference in perceptual classification: Implicit learning or cognitive complexity? *Memory & Cognition*. 2005; 33:1256–1271. [PubMed: 16532858]
- Prince M, Brown S, Heathcote A. The design and analysis of state-trace experiments. *Psychological Methods*. 2012; 17:78–99. [PubMed: 22040373]
- Schacter, DL.; Wagner, AD.; Buckner, RL. Memory systems of 1999.. In: Tulving, E.; Craik, FIM., editors. *Oxford handbook of memory*. Oxford University Press; New York: 2000. p. 627-643.
- Shepard RN. Stimulus and response generalization: A stochastic model relating generalization to distance in psychological space. *Psychometrika*. 1957; 22:325–345.
- Sloman SA. The empirical case for two systems of reasoning. *Psychological Bulletin*. 1996; 119:3–22.
- Squire LR. Memory systems of the brain: A brief history and current perspective. *Neurobiology of Learning and Memory*. 2004; 82:171–177. [PubMed: 15464402]
- Stanton RD, Nosofsky RM. Feedback interference and dissociations of classification: Evidence against the multiple-learning-systems hypothesis. *Memory & Cognition*. 2007; 35:1747–1758. [PubMed: 18062551]
- Waldron EM, Ashby FG. The effects of concurrent task interference on category learning: Evidence for multiple category learning systems. *Psychonomic Bulletin & Review*. 2001; 8:168–176. [PubMed: 11340863]
- Yonelinas AP. The nature of recollection and familiarity: A review of 30 years of research. *Journal of Memory and Language*. 2002; 46:441–517.
- Zeithamova D, Maddox WT. Dual-task interference in perceptual category learning. *Memory & Cognition*. 2006; 34:387–398. [PubMed: 16752602]

Figure 1.

Some state-trace plots produced by a single-system model (Nosofsky's, 1986, GCM) 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) $P(T_1)$ and $P(T_2)$ both monotonically increase as the single varying parameter increases. (B) $P(T_1)$ monotonically increases as the single parameter increases, but $P(T_2)$ does not. (C) $P(T_2)$ monotonically increases as the single parameter increases, but $P(T_1)$ does not. (D) Neither $P(T_1)$ nor $P(T_2)$ monotonically increase as the single varying parameter increases. Instead $P(T_1)$ and $P(T_2)$ both peak at an intermediate value of the parameter.

Figure 2.

State-trace plots from the GCM (left column) and a simplified version of COVIS (right column). (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.