The dynamics of insight: Mathematical discovery as a phase transition

Damian G. Stephen and Rebecca A. Boncoddo
University of Connecticut, Storrs, Connecticut

and

James S. Magnuson and James A. Dixon
University of Connecticut, Storrs, Connecticut
and Haskins Laboratories, New Haven, Connecticut

In recent work in cognitive science, it has been proposed that cognition is a self-organizing, dynamical system. However, capturing the real-time dynamics of cognition has been a formidable challenge. Furthermore, it has been unclear whether dynamics could effectively address the emergence of abstract concepts (e.g., language, mathematics). Here, we provide evidence that a quintessentially cognitive phenomenon—the spontaneous discovery of a mathematical relation—emerges through self-organization. Participants solved a series of gear-system problems while we tracked their eye movements. They initially solved the problems by manually simulating the forces of the gears but then spontaneously discovered a mathematical solution. We show that the discovery of the mathematical relation was predicted by changes in entropy and changes in power-law behavior, two hallmarks of phase transitions. Thus, the present study demonstrates the emergence of higher order cognitive phenomena through the nonlinear dynamics of self-organization.

Among the most compelling issues in psychology is how the cognitive system can spontaneously leap from one structure to another. For example, as children are learning to solve addition problems (e.g., 4 + 2 = ?), they initially raise the appropriate number of fingers to represent the two addends (e.g., four on one hand, two on the other) and count them all. However, after using this strategy for a while, children spontaneously discover a new strategy: They raise the appropriate number of fingers for the smaller addend (e.g., two), and then begin counting from the larger addend (e.g., four, five, six; Siegler & Araya, 2005). Such discoveries or insights are particularly interesting phenomena, in part because the change in structure appears to be driven by the activity of the system itself. There is no external agent guiding the individual toward a new organization, nor is there an internal plan that contains the new structure in miniature. Explaining the emergence of new structures is a serious challenge for cognitive theory. Chronicle, MacGregor, and Ormerod (2004), for example, noted that information-processing approaches had not made substantial progress in explaining new structures (i.e., insights) during problem solving. The central question is, how can a functioning system can suddenly self-organize into a new configuration in the absence of any external supervision or internal blueprint.

In previous work, we found evidence for just such a spontaneous change in cognitive structure during a simple problem-solving task (Dixon & Kelley, 2006). Participants

were asked to solve gear-system problems by predicting the turning direction of the final gear, given the turning direction of the first gear (see Figure 1). After solving the problems with lower level strategies, many participants spontaneously discovered a mathematical relation—parity—that afforded a higher order solution to the problems (Dixon & Bangert, 2004; Dixon & Kelley, 2006, 2007). (The parity of the number of gears in the system [i.e., odd, even] determines whether the final gear turns in the same direction as the driving gear.) The participants rarely made errors prior to discovering parity. Furthermore, since the gear displays were static, the relation could not be extracted from the movement of the gears. The new relation appears to arise from the participants' own activity. Gentner and Namy (1999) investigated a similar phenomenon in which representational change occurs from the child's own actions. They showed that when children repeatedly compare objects during classification, they begin to detect their common dimensions, a process sometimes called progressive alignment. Gentner has proposed that repeated alignment is a central process in cognitive development (e.g., Gentner, Loewenstein, & Hung, 2007). In the present study, we address the emergence of new cognitive structure from one's own activity as an instance of self-organization.

Self-Organization of New Structure

Researchers across a wide variety of domains have grappled with the emergence of new structure (Jensen,

---

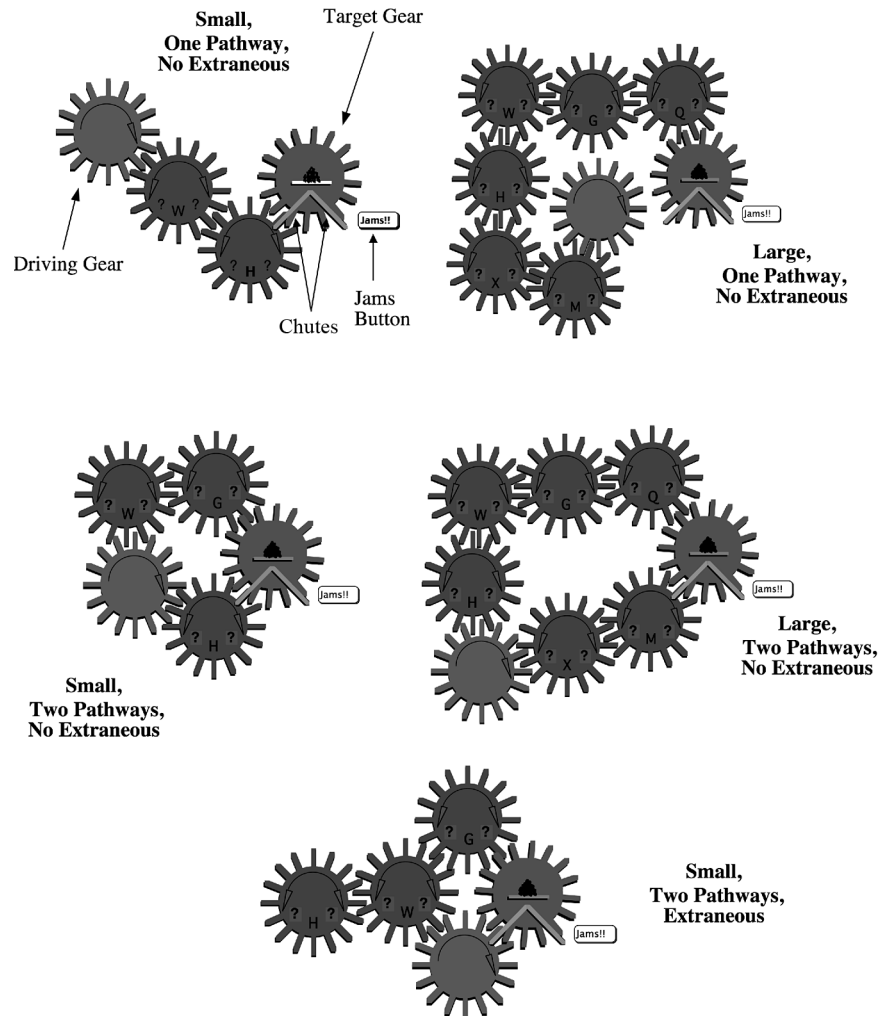D. G. Stephen, damian.stephen@uconn.edu

Figure 1. Examples of gear-system problems. The gear systems varied along three dimensions: size (small, 4 or 5 gears; large, 7 or 8 gears), number of pathways (one or two), and whether an extraneous gear was present. Extraneous gears were not part of the causal pathway from the driving gear to the target gear. Gear systems with two pathways had the potential to jam—to fail to turn because of opposing forces on the target gear; thus, a button labeled "Jams!!" constituted a third response option (in addition to clockwise and counterclockwise rotation).

1998; Webster & Goodwin, 1996). Physical and biological systems spontaneously exhibit new structures, just as cognitive systems do. For example, when a fluid is heated, the fluid molecules spontaneously form convection cells, new structures that function to keep the fluid stable (Hilborn, 1994; Lorenz, 1963). Similarly, a well-studied, single-celled organism, *Acetabularia*, can generate an assortment of novel structures (i.e., not seen in its typical developmental course) depending on the medium in which it is grown (Harrison & Hillier, 1985). Although the details of such systems vary greatly, research in nonlinear dynamics has delineated the sequence of events that results in the self-organization of new structure. In broad outline, self-organization begins when the activity of the system changes the interactions among microelements. The new interactions create new properties, leading to a different organization that appears as new, global structure. This

sequence of events is governed by a set of higher order relations, evident in system behavior, that obtain during self-organization (we discuss two such relations in detail below) (Hilborn, 1994). Because these higher order relations are grounded in the principles of nonlinear dynamics and make strong predictions about well-understood measures, they have provided a powerful approach for understanding self-organization.

In psychology, the idea that the activity of many interacting microelements undergirds cognition has had a major impact on theory through the development of connectionist computational models (McClelland, Rumelhart, & PDP Research Group, 1986). An important class of connectionist models exhibits self-organization in the sense described above. Under a variety of conditions, these models, which usually employ Hebbian and/or self-organizing map (SOM) algorithms, spontaneously

generate new, functionally appropriate structures without supervision (McClelland, 2006; Silberman, Bentin, & Miikkulainen, 2007). These algorithms provide biologically plausible accounts of changes in neural connectivity (Ashby, Ennis, & Spiering, 2007; Sullivan & de Sa, 2006). Self-organizing connectionist models have been successful in addressing the emergence of new structure in language acquisition, including aspects of vocabulary (Li, Zhao, & MacWhinney, 2007) and syntax (Hadley & Cardei, 1999). They have also proven useful for modeling the emergence of early concepts, such as physical causality (Cohen, Chaput, & Cashon, 2002; Schyns, 1991).

Attempts to understand how such networks change during learning has revealed that many of these models operate under the principles of self-organization from nonlinear dynamics (Graepel, Burger, & Obermayer, 1997; Siri, Quoy, Delord, Cessac, & Berry, 2007). The same higher order relations that govern self-organization in a wide variety of other domains, such as fluids, lasers, and ferromagnets, are exhibited by a class of connectionist models that learn via Hebbian and SOM algorithms. The alignment of nonlinear dynamics with the self-organizing behavior of this class of connectionist models has deep implications, because nonlinear dynamics makes a distinct set of predictions about the behavior of a system as it self-organizes: Nonlinear dynamics predicts that microscopic fluctuations in behavior change systematically during the emergence of new macroscopic structure. Therefore, this class of connectionist models should also make predictions about properties of very fine-grained behavior that are endemic to self-organization. Here, we focus on two key predictions from dynamics: Changes in entropy will anticipate the transition to new cognitive structure, and in addition, power-law behavior will also anticipate the transition to new cognitive structure.

Entropy and Power-Law Behavior

One way to understand why changes in entropy and changes in power-law behavior predict phase transitions is in terms of constraints, physical bonds that hold microelements together. In order for a system to function in a specific configuration, some of its microelements must be coupled or constrained. Such constraints are the underlying explanation of a system's structure. Without constraints, the microelements would act independently of each other, and the system would have no structure; that is, it would show no organized behavior. Entropy is inversely related to the number of constraints among the microelements (Kugler & Turvey, 1987). As a dynamical system approaches a phase transition, constraints among the microelements begin to break down. This results in more disordered (i.e., entropic) behavior. It also results in an increase in the degree to which interactions among its microelements dominate behavior, because microelements that were previously constrained are now free to interact. If the interactions give rise to new structure within the system, they do so by creating new constraints among the microelements. As the number of new constraints increases, entropy decreases, and the behavior of the sys-

tem again becomes more orderly (Kelso, 1995; Nicolis & Prigogine, 1977; Prigogine & Stengers, 1984).

Power-law behavior is a specific type of nonlinear relationship between the magnitude of a behavior (e.g., size of a movement, length of a response time) and its frequency (i.e., how often behaviors of that magnitude occur). Like entropy, power-law behavior is related to the degree to which the system is constrained, but in a way that reflects an additional fundamental property of the system's architecture: nested structure. All biological systems (and many nonbiological ones) are organized at multiple levels (West, Brown, & Enquist, 1999). In such nested structures, higher levels are made up of lower ones, which in turn are made up of yet lower ones; this nesting obtains across many scales. Nested structure produces power-law behavior because, as one traverses the system from higher to lower levels, the number of microelements increases proportionally to the number of levels traversed.

When such a system is operating in a stable configuration, some of the microelements are constrained; they temporarily function together much like a unitary whole or component. As the system approaches a phase transition, constraints break across multiple levels, thereby freeing previously constrained microelements. This increases the amount of activity at each level. Statistically, the increase in activity across the nested structure can be quantified by the magnitude of the power-law exponent: The greater the activity is, the greater the value of the power-law exponent will be. As constraints re-emerge and the system settles into a new structure, the power-law exponent decreases (Grebogi, Ott, Romeiras, & Yorke, 1987). Much of the interest in power-law behavior stems from the relationship between power-law exponents and phase transitions. Indeed, a major basis for classifying types of dynamical systems is their critical power-law exponent, the value of the exponent at which the phase transition begins (Hilborn, 1994).

We propose that changes in entropy and changes in power-law exponents will predict a phase transition in the cognitive system. For this purpose, we now present a simplified model of a dynamical system undergoing a phase transition and chart its entropy and power-law exponent throughout. The system that we have chosen is the Lorenz (1963) model of fluid convection. This well-known model consists of three interrelated equations: $dX/dt = s(Y - X)$, $dY/dt = rX - Y - XZ$, and $dZ/dt = XY - bZ$. The X variable indexes the intensity of the fluid flow. Y indexes the difference between ascending and descending convection currents. Z is related to the deviation of the vertical temperature profile from linearity. The three variables, X, Y, and Z, form the dimensions of the system's phase space. Phase space is simply the set of all potential states, defined as all possible combinations of X, Y, and Z values. Setting the s, r, and b parameters and initial values of X, Y, and Z determines the system's behavior. The model demonstrates a wide variety of complex behavior, including a sudden shift from one attractor to another as a control parameter, r, intended to represent the heating of the fluid, increases (Sparrow, 1982). The goal of this example is to concretely demonstrate how entropy and the power-law exponent change as a dynami-
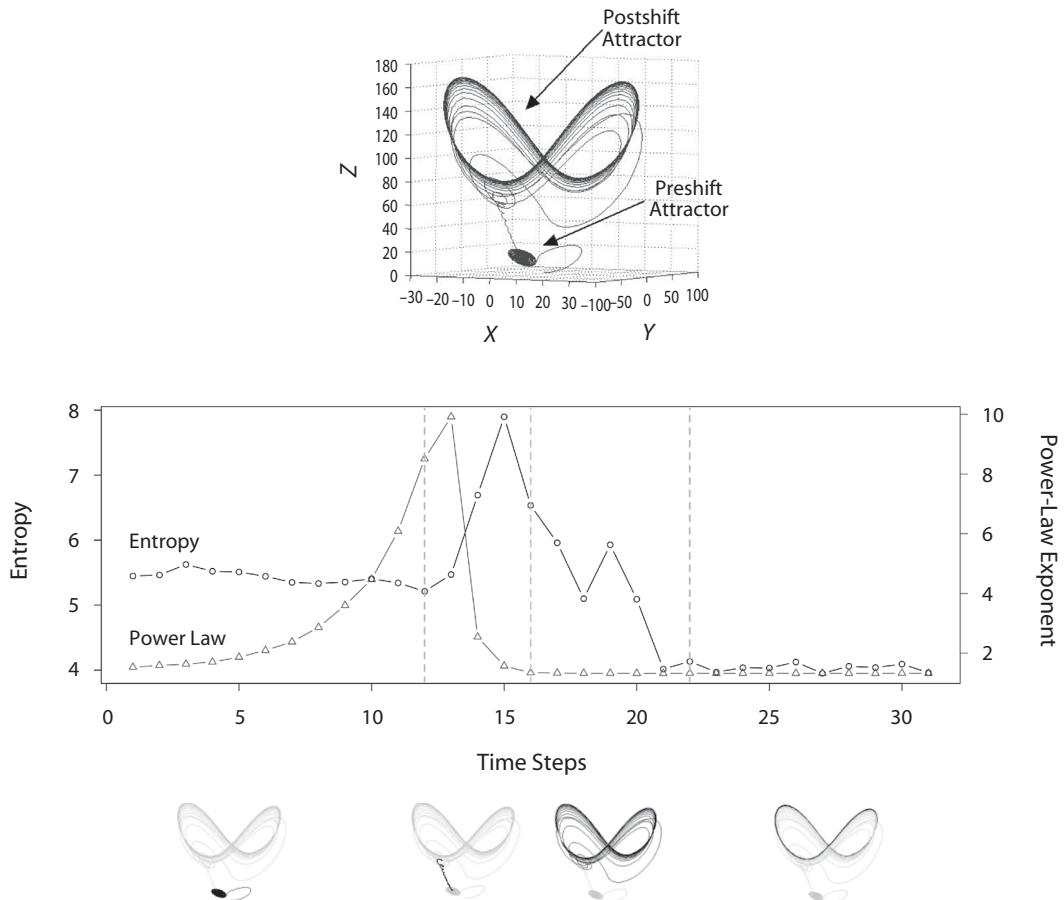
Figure 2. The top panel shows the trajectory of the Lorenz system as it undergoes a phase transition from a preshift to a postshift attractor. The middle panel shows measures of entropy and the power-law exponent over time steps as the Lorenz system approaches and goes through its phase transition. The dark gray curve shows the entropy of the phase-space trajectory over time; the light gray curve shows the power-law exponent. The panels beneath the horizontal axis show the phase-space trajectory of the Lorenz model as its control parameter is increased. The darkened region of each trajectory shows the activity of the system during the portion of the time course distinguished by the dashed vertical lines in the figure.

cal system passes through a phase transition. The control parameter within the Lorenz model is useful for changing its activity but is not a necessary aspect of such systems. Whether the cognitive system is sometimes driven by comparable variables (i.e., control parameters) is an open question, but not one that we intend to address here.

The upper panel of Figure 2 shows the trajectory of the Lorenz (1963) model across its three-dimensional phase space as a control parameter is increased. The trajectory is initially in a tight, disk-like attractor, but then abruptly leaves that regime. Subsequently, the system slowly settles into a new attractor. An *attractor* here corresponds to a region of the phase space to which the system repeatedly returns. From a more macroscopic perspective, attractors correspond to modes of behavior. In the example, the Lorenz model first repeatedly traverses a small region of very similar states. After the phase transition, the system behaves very differently, traversing a much broader region of states regularly. As can be seen in Figure 2, entropy increases as the system leaves the initial attractor and de-

creases as it settles into the new one. *Entropy*, here, is a measure of the degree of disorder in the trajectory through phase space. The power-law exponent also increases prior to the transition and decreases while the system settles into the new attractor. The power-law exponent indexes the degree of activity across the scales in the system. The activity increases and decreases around the phase transition. We explain the methods used to calculate entropy and the power-law exponent more fully in subsequent sections.

Predicting the Emergence of Cognitive Structure

In the present study, we explore the possibility that the spontaneous emergence of a new cognitive structure—a mathematical relation—reflects a phase transition within the cognitive system. On each trial, participants were given the turning direction of a driving gear that provided force to the system and were asked to predict the movement of a target gear. The gear systems were presented as static displays in the context of a game. We densely sampled the changing position of eye gaze as the participants solved

the gear-system problems. On the basis of previous work, we expected that the majority of the participants would first solve the problems using two lower level strategies. In general, previous work made two consecutive observations about participants in this task. Initially, participants traced the turning motions of the individual gears and the pushing of intermeshed teeth. Subsequently, they discovered that the gears form an alternating sequence (Dixon & Kelley, 2006; i.e., adjacent gears turn in opposite directions). Here, we focus on a further transition in which participants spontaneously discover an underlying mathematical relation: The parity of the system predicts the turning direction of the target gear (Dixon & Bangert, 2004; Schwartz & Black, 1996). In systems with an odd number of gears, the target gear turns in the same direction as the driving gear; in systems with an even number, it turns in the opposite direction.

Stephen, Dixon, and Isenhower (in press) showed that the transition from manually tracing the gears to treating the gears as an alternating sequence was predicted by changes in entropy and power-law behavior. Stephen et al. densely sampled fluctuations in behavior by tracking the motion of the participants' dominant hand as they traced the gears. Analyses of the time series from each trial showed that the transition to alternation was anticipated by changes in entropy and power-law behavior, as is described above.

In the present research, we measured the fine-grained fluctuations in behavior by tracking participants' point of gaze while they solved the gear-system task. Research from a variety of domains has shown that point-of-gaze is a very sensitive index of changes in cognition (Kowler, 1990). We used the time series of changes in point-of-gaze to test whether the transition to a mathematical relation—parity—shows the signatures of a phase transition. Specifically, we tested two converging predictions from the theory of nonlinear dynamics. First, the onset of new structure (i.e., discovery of the parity relation) should be predicted by an increase in entropy followed by a decrease. We measured the entropy of system behavior by performing recurrence quantification analysis (RQA) on the time series (Webber & Zbilut, 1994), a method that we explain below. We quantified the pattern of change in entropy across the six previous trials to predict the discovery of parity on each current trial. The second prediction from nonlinear dynamics is that the power-law exponent will increase as participants approach the transition and decrease just prior to the transition. For each trial, we estimated the power-law exponent using spectral analysis on the time series of changes in point-of-gaze (Aks, 2005). The participants should show a systematic increase and subsequent decrease in power-law behavior as they approach the discovery of parity.

METHOD

Participants

Thirty-three undergraduate students, 14 male and 19 female, completed the experiment as one option for fulfilling a course requirement.

Materials and Procedure

The participants were seated at a computer monitor with an eye-tracking camera located below it, calibrated to the left eye. Point-of-gaze relative to the computer screen was sampled at 60 Hz by an eyetracking system (ASL 6000). The gear-system task was presented on the computer monitor. The participants were asked to play a computerized game in which they would race their train against one controlled by the computer. They could increase the speed of their train by solving the gear-system problems presented at fueling stations along the racecourse.

Each gear system comprised a driving gear that turned clockwise, a variable number of intermediate gears, and a target gear. The fuel was located on a shelf on the target gear. The participants predicted whether the gear would turn clockwise or counterclockwise; their train was positioned so as to catch the fuel (if the gear turned in the correct direction). It was also possible for some gear systems to jam—to fail to turn because of opposing forces from two gear pathways. Thus, a third response option was a button labeled "Jams!!" located next to the target gear. After the participants indicated that they were satisfied with their prediction, the final gear turned appropriately, thereby providing feedback about whether their prediction was correct. The other gears were covered by a virtual screen prior to the final gear's moving. Figure 1 shows examples of the various types of gear systems. The participants completed 4 practice trials, followed by 32 standard trials. The practice trials were presented in a fixed order; the order of the standard trials was randomized for each participant.

The participants could solve the gear-system problems in any way they wished. They were asked to think aloud as they worked through each problem; the experimenter coded their strategy on each trial. The transition to the parity strategy is marked by counting and the use of odd–even designations; thus, it is very easy to identify. Reliability was very high, with 95% agreement between two independent raters across all strategies and 100% agreement with regard to the onset of parity.

Quantifying Angular Change in Point-of-Gaze

We calculated angular change in point-of-gaze by taking the arctangent of (dh/dv) for each pair of successive frames or time steps, ($t, t + 1$), where dh and dv are changes in the horizontal and vertical coordinates, respectively (Aks, Zelinsky, & Sprott, 2002). Figure 3 illustrates how angular change, θ, was computed for two hypothetical points. An angular-change time series was computed for each trial; Figure 4 shows an example from one trial.

Phase-Space Reconstruction

We used the time series of changes in point-of-gaze to reconstruct phase space on each trial. Phase space is defined by the variables that determine the state of the system; each variable forms one dimension of the space. Each point within the space specifies a particular state of the system (i.e., the combination of positions across the dimensions). Phase-space reconstruction is a powerful, widely used technique in nonlinear dynamics based on a fundamental insight by Takens (1981). Takens proved that, given a nonlinear system, key geometrical properties (i.e., the topology) of phase space could be reconstructed by using copies of the times series to stand in for dimensions that had not been explicitly measured. These copies of the original series are lagged (i.e., time shifted) but are otherwise identical to the original, univariate series. The lagged time series serve as proxies for the unmeasured dimensions of phase space, in the sense that they jointly capture important aspects of its relational structure.

Although the formal arguments underlying phase-space reconstruction are beyond our present scope (Abarbanel, 1996), an example may facilitate the understanding of this approach. The left panel of Figure 5 shows a trajectory through a three-dimensional phase space. The right panel shows the reconstruction of that phase space using only the X variable; the other two dimensions are lagged copies of the time series. As can be seen in the figure, the recon-
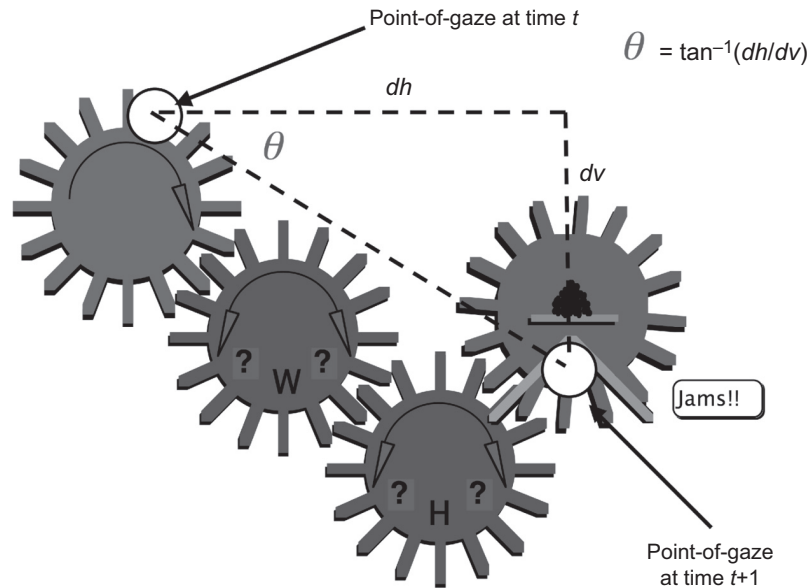
Figure 3. An example of the computation of angular change for two hypothetical, successive eye positions. Change on the *h* coordinate, *dh*, and change on the *v* coordinate, *dv*, allow the angle, *θ*, to be calculated.

structed trajectory has the same ordinal relations among its points as the actual phase-space trajectory (i.e., the ordering of the points along the dimensions is the same). Note also that the reconstruction gives the trajectory of the system in which the X variable is embedded. It is not just the trajectory of X alone, despite the fact that only the X variable was used in the reconstruction.

RQA

RQA is a method for assessing the organizational properties of phase space (Marwan, Romano, Thiel, & Kurths, 2007; Webber & Zbilut, 1994). RQA first evaluates the degree to which a system returns to points (or narrow regions) within phase space. Put more plainly, the first step in RQA is to assess when the system is in approximately the same state as it was at some previous time. As an example, assume that the trajectories in Figures 6A and 6B were repeatedly sampled, once at each point shown. For the purposes of exposition, we assume that there are only two relevant dimensions; these dimensions jointly define the space of states that the system may take. Given the location of each point in this space, identifying

whether the system is precisely revisiting a point in space is easy: The distance between the current point and the previous point is zero. More usually, recurrences are defined as coming very close to a previous point (i.e., within some specified distance); the circles surrounding the points in Figures 6C and 6D illustrate this idea. If two points fall within a specified distance (the diameter of the surrounding circles), they are considered recurrent. (The distance is set as a parameter in the analysis.)

Recurrences are the building blocks from which all other measures in RQA are constructed (i.e., these other measures assess properties of the pattern of recurrences). Of particular interest is when recurrences occur successively, because this indicates that the system is in an attractor (i.e., a region of phase space preferred by the system). The lines connecting recurrent points in Figures 6E and 6F show successive recurrences. Runs of recurrent points show that the trajectory through phase space has reconverged with its previous path (or is very close its previous path). In RQA, these runs of recurrent points are an important indicator of the organization of phase space.
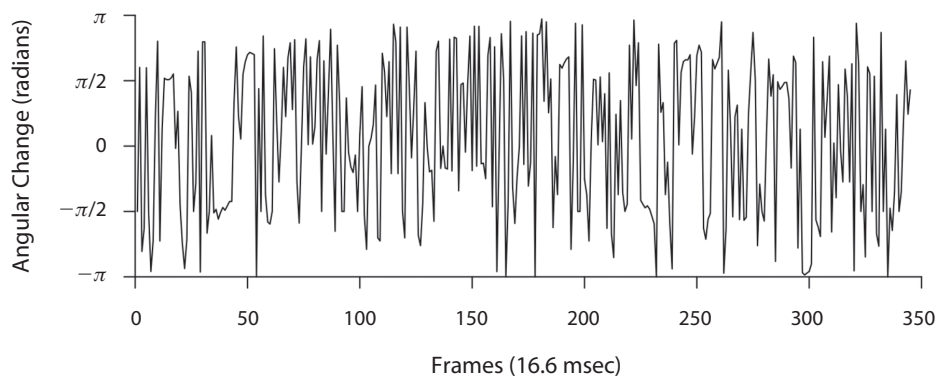

Figure 4. An example of a typical angular change time series for a single trial for a single participant. Angular change in radians is plotted as a function of time, expressed in frames; each frame was 16.6 msec.

Original Phase Space   **Reconstructed Phase Space**
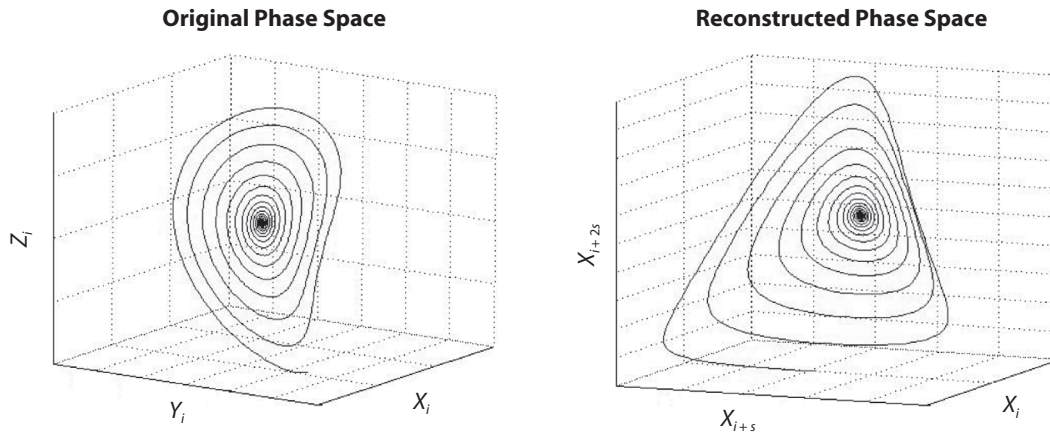

Figure 5. The panel on the left shows an example of a three-dimensional phase space for the Lorenz model. The points in the left panel (actual phase space) are given by the values of three variables: $p_i = [X_i, Y_i, Z_i]$, where i indexes time steps in the series. The panel on the right shows the reconstruction of that phase space using only the X dimension. To reconstruct phase space, the original time series of value X is lagged by s time steps ($s = 7$, in the present example) for each dimension. The points in the right panel (reconstructed phase space) are given by values of X and lagged copies of X: $p_i = [X_i, X_{i+s}, X_{i+2s}]$.

For our present purposes, we consider two measures of dynamic organization: percent recurrence and entropy. Percent recurrence is simply the number of observed recurrences divided by the number of potential recurrences (i.e., the number of unique pairs of points). Percent recurrence provides an index of the degree to which the system revisits previous regions. The trajectories in Figures 6A and 6B have 2.5% and 3.1% recurrence, respectively. Because all other measures within RQA rely on the number of recurrent points, we use percent recurrence as a covariate in our analyses. Although it is desirable to keep recurrence relatively low (~2%–4%), the greater the percentage of recurrent points is, the larger the other measures will tend to be.

RQA also provides a measure of entropy. Entropy assesses the degree of disorder in the trajectory through phase space. The more variable the trajectory is, the larger the value of entropy will be. Entropy is quantified using Shannon's (1948) equation:

$$\text{Entropy} = -\sum p(x_i)\log_2 p(x_i),$$

where $p(x_i)$ is the (nonzero) proportion of runs of length i. For example, the trajectory in Figure 6A has an entropy of 0; the trajectory in Figure 6B has an entropy of 1.58.

RESULTS

Descriptive Statistics on Performance and Strategy

As was expected, prior to discovering parity, the participants manually traced the force across the system, simulating the turning and pushing of the individual gears (on 65% of prediscovery trials). Many of the participants also used an alternation strategy in which they classified the turning direction of the gears in an alternating sequence (on 35% of prediscovery trials). Performance prior to discovery was quite accurate (84% correct). Of the 33 participants who took part in the study, 22 discovered the parity relation. The median discovery trial was 20.

Organization of Analyses

We report multiple analyses to describe the changes over time in the participants' global behavior and to address the key predictions of the study. We first employ more conventional analytical methods before taking up the complex, dynamical systems analyses. In the first section, we address reaction time and accuracy effects in two different directions: the effects that practice, discovery of parity, and individual differences had on reaction time and accuracy and the effects that reaction time and accuracy had on the onset of parity. In the second section, we address the conventional fixation measures in eye movements: changes in the number of fixations across trials and the relationships between fixations and the onset of parity. Among other findings, it will emerge that reaction time, accuracy, the number of fixations, and the duration of fixations all fail to predict the onset of parity.

We then move on to the dynamical systems analyses of eye movements. In the third section, we report descriptive statistics from the RQA of angular change in eye position. RQA is the technique that we used to generate a measure of entropy from eye movements for each trial and for each participant. In the fourth section, we report both models of measured RQA entropy and models of measured power-law exponent that successfully predict the onset of parity. It emerges from these analyses that the trajectories of entropy and the power-law exponent follow the same patterns as those illustrated by the phase transition in the Lorenz (1963) system.

In many of the inferential analyses reported below, we employed maximum likelihood estimation rather than ordinary least-squares estimation (Singer & Willett, 2003). Within this framework, testing the significance of an individual parameter or set of parameters involves evaluating the reduction in the deviance of the model. The reduction in deviance is usually reported as $-2 *$ log of the likelihood ($-2LL$), which is distributed as χ^2 with the number of degrees of freedom equal to the number of parameters added to the model. For example, a change in $-2LL$ is significant at $p < .05$ on one degree of freedom when it exceeds 3.84.
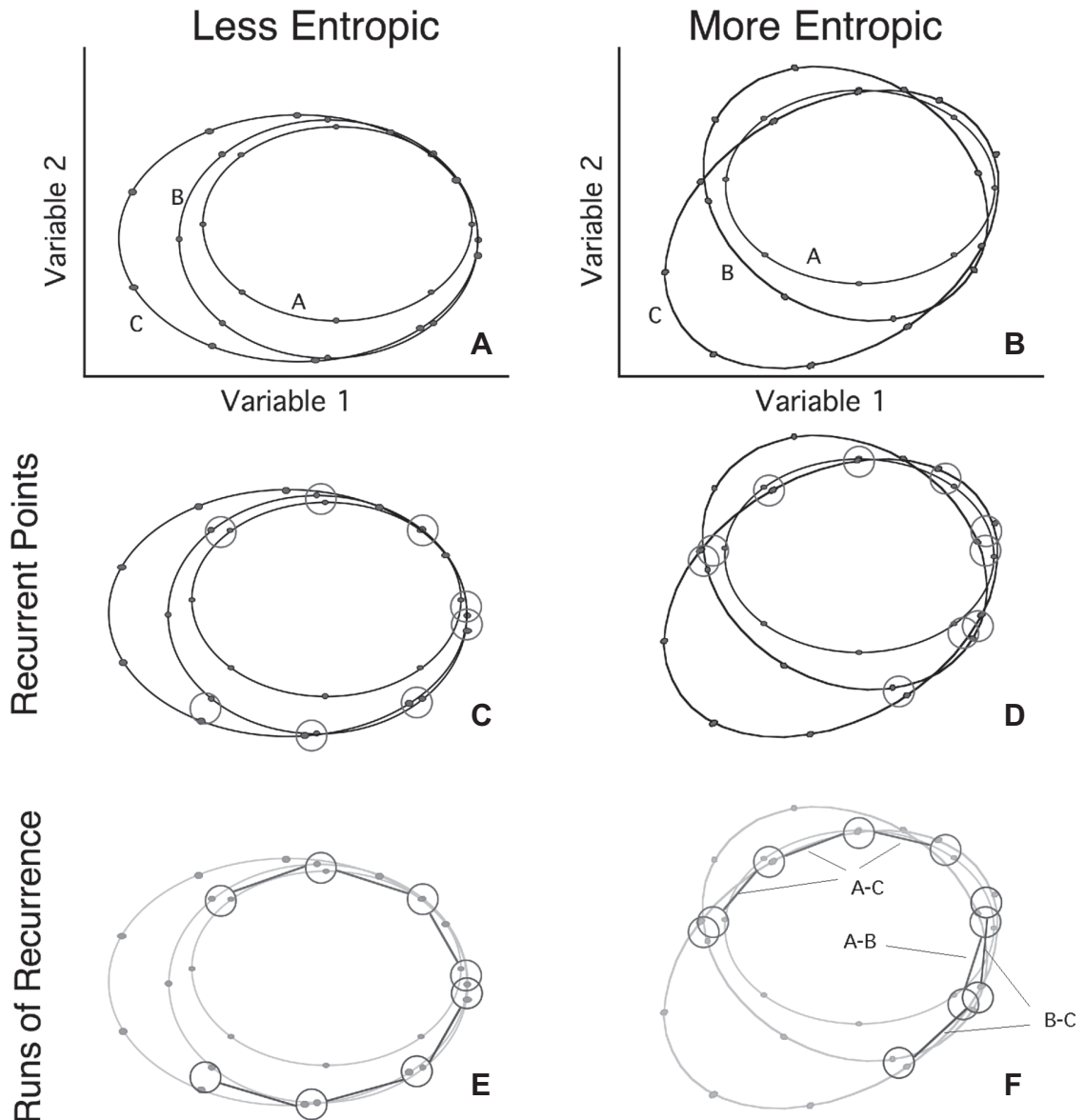
Figure 6. (A) A hypothetical trajectory projected onto a two-dimensional space. The state of the system at any moment in time is given by the values on the two variables. As the system changes, it creates trajectories through the space. The continuous behavior of the system cycles through three loops in the following order: Loop A, Loop B, Loop C, Loop A, Loop B, and so forth. The dots indicate points that were sampled during one complete circuit through the trajectory. (B) An analogous but more disordered trajectory. (C, D) The concept of recurrence. Some points in each panel fall within a specified distance of each other, as is indicated by the diameter of the circles. Each pair of points within a circle is thus considered recurrent. (E, F) Runs of recurrent points. In panel E, there are two such runs. One run, shown in the top half, consists of four recurrent pairs of points from the convergence of Loops A and B. The second run, shown in the bottom half, also consists of four recurrent pairs; here, Loops B and C are converging. Panel F also shows runs of recurrent pairs as Loops A and B are converging and as Loops B and C are converging. This trajectory also has another run of recurrent pairs as Loops A and C converge. All three runs are of different lengths; this variability is indexed by the entropy measure. Panel F also contains a recurrent pair of points that is not part of a run of recurrences: Two trajectories intersect but are not aligned for any length of time.

Response Time and Accuracy

Replicating previous results, we found that the participants solved the gear system problems quite accurately. The upper panel of Figure 7 shows the proportion of correctly solved problems over trials. A logistic regression showed that, consistent with the lack of change over trials apparent in the figure, there was no effect of trial on accuracy [$B = 0.008$, $SE = 0.01$; change in $-2LL$, $\chi^2(1) = 0.62$, n.s.]. The onset of the parity strategy significantly increased the proportion correct [$B = 1.64$, $SE = 0.38$; change in $-2LL$, $\chi^2(1) = 25.08$]. The onset of parity did not interact with the rate of change over trials [$B = 0.013$,

$SE = 0.05$; change in $-2LL$, $\chi^2(1) = 0.06$]. In other words, the onset of parity improved accuracy, but practice did not improve accuracy beyond this effect, whether for lower level strategies or for parity. More complex models that handle the potential autocorrelation and heteroscedasticity in data over time confirmed the results reported above.

The lower panel of Figure 7 shows the mean response times (in seconds) over trials. Response times did not decrease significantly over trials. A growth-curve (i.e., random coefficients) model showed no significant effect of trials [$B = -0.05$, $SE = 0.03$; change in $-2LL$, $\chi^2(1) = 1.29$, n.s.] (Mirman, Dixon, & Magnuson, 2008; Singer & Willett, 2003). The model included random effects on the intercept [$\sigma^2 = 37.24$; change in $-2LL$, $\chi^2(1) = 76.30$], on the slope (i.e., trials) [$\sigma^2 = 0.005$; change in $-2LL$, $\chi^2(1) = 0.26$], and on their correlation [$r = -.71$; change in $-2LL$, $\chi^2(1) = 6.15$]. The significant random effect on the intercept implies that there was significant between-participants variation in initial response time. The lack of a significant random effect on the slope suggests that the rate of change in response time over trials did not vary substantially between participants. The individual (i.e., per participant) parameters used to estimate these two random effects were negatively correlated. (This relationship is estimated as an additional parameter in the model.) The negative correlation indicates that individual differences in initial response time were related to individual differences in the rate of change in response time, such that initial response time usually preceded
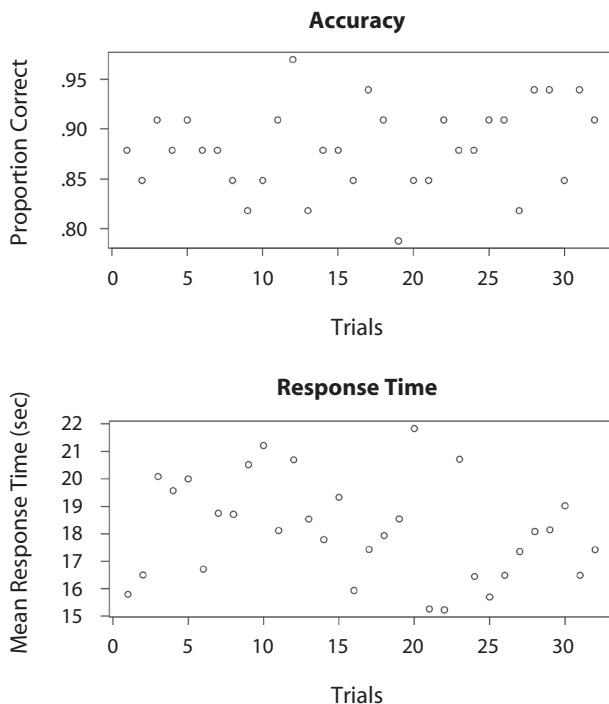

Figure 7. The top panel shows the mean proportions correct as a function of trials. The lower panel shows the mean response times as a function of trials.

a decrease in response times. This relation is often observed in growth-curve models and is usually treated as an artifact.

The onset of the parity strategy, the event of central interest, did not have a significant effect on the intercept (i.e., shifting the response time curve down) [$B = -2.03$, $SE = 1.05$; change in $-2LL$, $\chi^2(1) = 2.95$, n.s.]. However, the onset of parity did significantly interact with trials [$B = -0.20$, $SE = 0.09$; change in $-2LL$, $\chi^2(1) = 4.13$] (often called an "effect on the slope"). That is, average response time after the discovery of parity was not different from the average response time before the discovery of parity. However, the discovery of parity led to a significantly more rapid decrease in response time.

Number and Duration of Fixations

The total number of fixations per trial is shown in the top panel of Figure 8. The number of fixations decreased significantly over trials [$B = -0.34$, $SE = 0.05$; change in $-2LL$, $\chi^2(1) = 27.15$]. The model included a random effect on the intercept [$\sigma^2 = 63.30$; change in $-2LL$, $\chi^2(1) = 121.32$], on the slope (i.e., trials) [$\sigma^2 = 0.03$; change in $-2LL$, $\chi^2(1) = 6.50$], and on their correlation [$r = -.80$; change in $-2LL$, $\chi^2(1) = 19.75$]. The significant random effects imply that there was substantial individual variation in the total number of fixations on the first trial, as well as in the rate of change in the number of fixations over trials. Given the nonlinear shape of the curve, we added trial^2 to the model to capture the quadratic form [$B = 0.016$, $SE = 0.004$; change in $-2LL$, $\chi^2(1) = 20.06$]. The positive quadratic term indicates that the number of fixations decreased at a decreasing rate. The random effects remained largely unchanged with the addition of the quadratic term.

We tested for changes in number and duration of fixations with respect to the onset of parity. The onset of parity did not have a significant effect on the intercept [$B = -1.28$, $SE = 1.11$; change in $-2LL$, $\chi^2(1) = 1.32$] or the slope [$B = 0.037$, $SE = 0.12$; change in $-2LL$, $\chi^2(1) = 0.76$] of the number-of-fixations trajectory. The discovery of the parity strategy did not result in an immediate shift in the number of fixations, nor did it impact the rate of change in the number of fixations over trials. We also calculated the mean and median duration of each participant's fixations on each trial. The middle and lower panels of Figure 8 shows these two measures, respectively, as a function of trial, averaged over participants. As the figure suggests, neither of these measures changed reliably across trials [largest change in $-2LL$, $\chi^2(1) = 1.52$]. The onset of parity did not affect the intercept or the slope of these trajectories over trials [largest change in $-2LL$, $\chi^2(1) = 1.32$]. That is, the number and duration of fixations were not appreciably different before or after the onset of parity.

Descriptive Statistics for RQA Parameters

We reconstructed phase space from the angular-change time series for all trials up to the discovery of parity and performed RQA on the reconstructed phase-space tra-
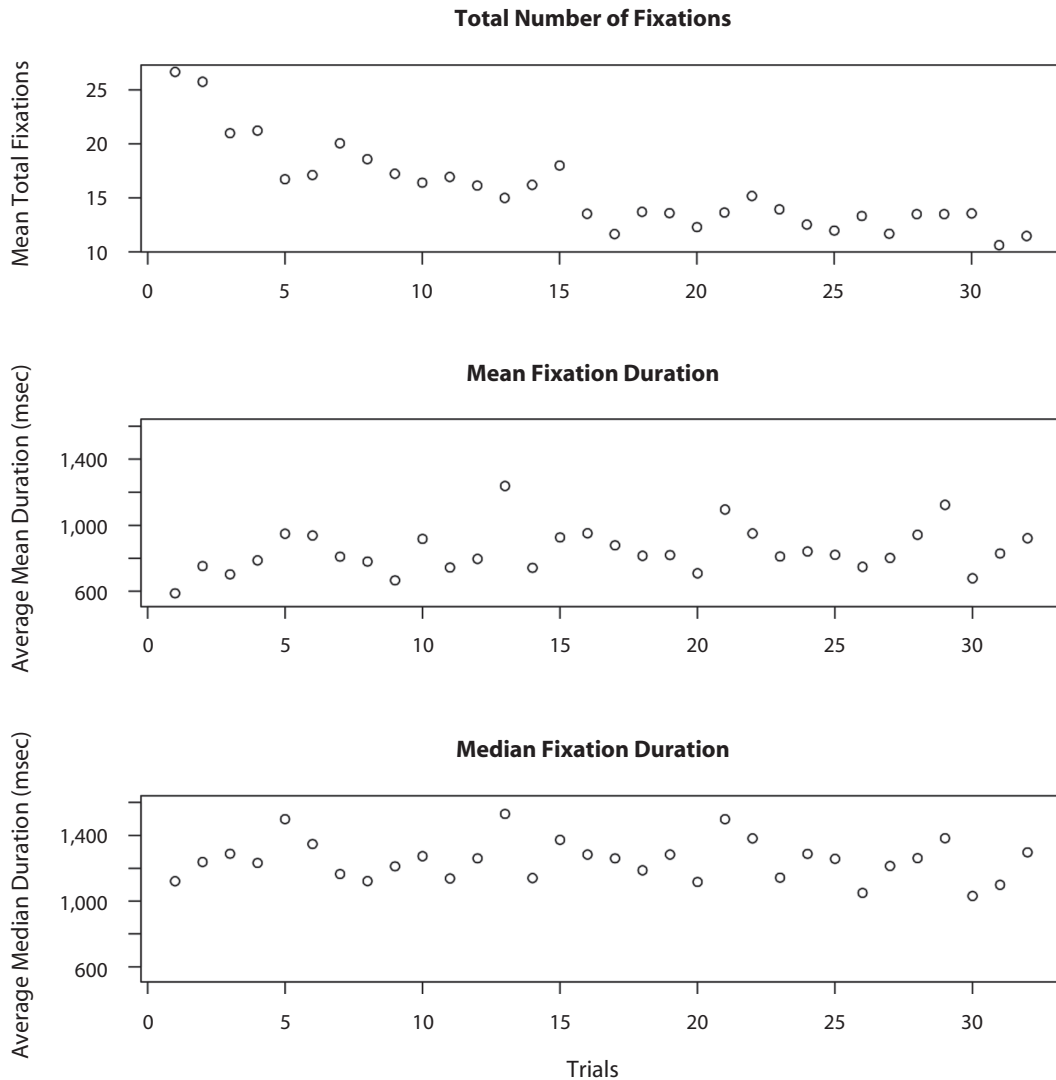
Figure 8. The top panel shows the number of fixations within each trial, averaged over participants. The middle panel shows the mean duration of fixations within each trial, averaged over participants. The bottom panel shows the median fixation duration within each trial, again averaged over participants.

jectories for each trial separately. Following Abarbanel (1996), we set the lag for each trial at the first minimum of the average mutual information function ($M = 3.46$ bits, $SD = 1.38$). *Average mutual information* is a measure (in bits) of how much one learns about the current value of a time series from a previous value. The first minimum of this function across all lags has been shown to be a good choice of lags for reconstructing phase space (see Arbarbanel, 1996, for a complete discussion). Given the length of the time series under consideration, we set the number of dimensions to four. The number of dimensions has been shown to be a noncritical parameter for this method (Webber & Zbilut, 2005). The mean values for recurrence and entropy were $M = 2.48\%$ ($SD = 3.46$) and $M = 0.85$ bits ($SD = 0.80$), respectively.

The conventional measures of eye movement behavior (i.e., total number of fixations, mean and median fixation duration) and response time jointly explained very little

variance in entropy (~2%) and recurrence (<1%). Table 1 shows the bivariate correlations between these measures and the RQA measures of phase space. The conventional measures are not associated with RQA measures, suggesting that RQA and conventional analyses provide different information (see Knöblich, Ohlsson, & Raney, 2001, for an example of conventional measures applied to insight problems).

Predicting the Discovery of Parity

Entropy. Recall that a central prediction from self-organization is that entropy should increase and decrease prior to discovery. In previous work, we found that entropy peaked and then dropped across the trials preceding discovery. The light gray curve in Figure 9 shows the mean values of entropy on the six trials prior to discovery; discovery trials are aligned on the far right side. As was predicted, entropy increased and decreased just prior

Table 1
Descriptive Fixation Statistics and Bivariate Correlations
With Entropy and Recurrence

	M	SD	Entropy r	Recurrence r
Number of fixations	17.53	14.08	.04	.06
Median fixation duration (msec)	864.24	841.13	.02	.01
Mean fixation duration (msec)	1,288.76	926.14	.07	.00
Response time (msec)	19,010.40	12,792.22	.11	.07

to discovery. The darker curve shows the mean values of entropy on the six trials preceding all nondiscovery trials (for all trials on which the participants were still at risk for discovery). We tested the different form of these observed functions more formally using a multilevel growth-curve model.

We assessed the trajectory of the entropy measure on the six trials preceding each trial (beginning, necessarily, with Trial 7). Trials that immediately precede a discovery should have a sharper rise and fall in entropy than trials that do not immediately precede a discovery. The predicted rise-and-fall pattern prior to a target trial is easily quantified as a quadratic function, consisting of an intercept, prior-trial, and prior-trial^2. We created a set of orthogonal polynomials for prior-trial and prior-trial^2 to eliminate the colinearity between these predictors (we retain the labels *prior-trial* and *prior-trial^2* for these orthogonal polynomials). The base model included the following fixed effects: intercept ($B = 0.32$, $SE = 0.14$), prior-trial ($B = 0.034$,

$SE = 0.037$), prior trial^2 ($B = -0.015$, $SE = 0.035$), target trial ($B = -0.0003$, $SE = 0.005$), and percent recurrence ($B = 0.21$, $SE = 0.005$). The model also included random effects on the intercept ($\sigma^2 = 0.001$), prior-trial ($\sigma^2 = 0.21$), and prior-trial^2 ($\sigma^2 = 0.14$). Two additional random effects were included at the level of target trial; these parameters were used to model the potential autocorrelation and heteroscedasticity across target trials (σ^2s = 0.38 and 0.0003, respectively).

The prediction of interest was whether the trajectory for entropy immediately prior to discovery trials was different from that prior to nondiscovery trials. To test this prediction, we added a variable to the model that indicated whether a discovery had occurred on that target trial. We included the effects of discovery on the prior-trial and prior-trial^2 terms. The resulting coefficients indicate whether the parameters that capture the changes in entropy are different for discovery and nondiscovery trials. Discovery had significant effects on prior-trial^2 [$B = -0.44$, $SE = 0.196$; change in $-2LL$, $\chi^2(1) = 4.99$] but not on prior-trial [$B = -0.34$, $SE = 0.21$; change in $-2LL$, $\chi^2(1) = 2.57$]. The prior-trial^2 term is more negative prior to discovery trials, indicating a steeper rise and fall, as was predicted.

In a parallel set of analyses, we took the estimated quadratic growth-curve model parameters (again using orthogonal polynomials) to capture the rise-and-fall pattern of the entropy measure over the six trials preceding each target trial. We then used these parameters as predictors of discovery in a discrete time survival analysis (where trials define the unit of discrete time) (Singer & Willett, 2003). This is conceptually equivalent to, but statistically more efficient than, performing a separate ordinary least-squares polynomial regression on the six trials preceding each target trial and then using those parameters to predict the first use of the parity strategy. The parameters quantify the predicted rise-and-fall pattern in entropy that should precede discovery. As might be expected from the growth-curve model analyses reported above, adding prior-trial and prior-trial^2 to the model significantly improved the model fit [Bs = 6.94 and -6.77, SEs = 4.65 and 3.56, respectively; change in $-2LL$, $\chi^2(2) = 6.73$]. Assessing the independent contributions of these two terms to the model showed that prior-trial^2 was primarily responsible for the effect. The model also included terms for the average height of the entropy curve (the centered intercept from the growth curve model) ($B = -1.30$, $SE = 1.20$), the intercept (of the current regression) ($B = -3.30$, $SE = 1.20$), trial ($B = 0.04$, $SE = 0.03$), and percent recurrence on the previous trial ($B = 0.12$, $SE = 0.05$). Note that the
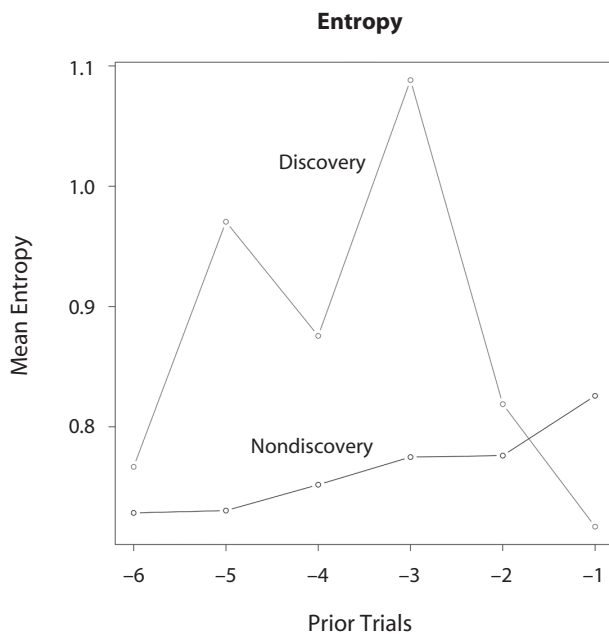

Entropy

Figure 9. The peaked gray curve shows the mean entropy values on the five trials leading up to discovery (i.e., the trial on which a participant first used parity). The darker curve shows the mean entropy values on trials preceding all other trials (i.e., those on which a discovery did not occur). The trial just prior to the current target trial is labeled -1; two trials prior is labeled -2; and so on. Entropy indexes the variability in the phase-space trajectory on each trial.
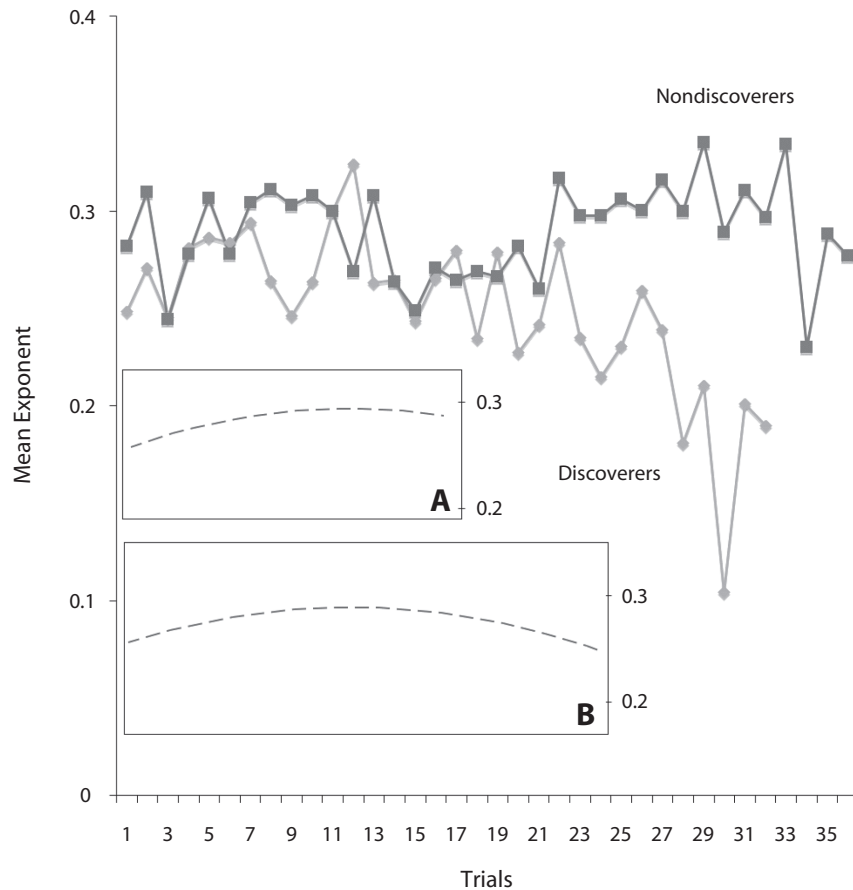
Figure 10. The main figure shows the mean power-law exponent for all trials prior to discovery, with separate curves for the participants who discovered parity (light gray line) and for those who did not (dark gray line). The participants who discovered parity contribute to the means represented by the light gray up to the trial on which they discovered parity. The participants who did not discover parity contribute to the means represented by the dark gray line on all trials. The insets illustrate the effect of timing of discovery (i.e., discovery trial) on the quadratic term. (A) An example of the predictions for relatively early discovery; discovery occurs on Trial 17. (B) Predictions for later discovery, on Trial 24.

measures of dynamic organization that we employed as predictors were computed from previous trials, not from the current one. Therefore, the rise and fall in entropy anticipates the discovery; it is not a consequence of using the new approach (i.e., parity).

Other, more traditional predictors, such as response time on the prior trial, prior accuracy, and number of saccades on the prior trial, did not contribute significantly to the model [largest change in $-2LL$, $\chi^2(1) = 1.51$, n.s.]. Similarly, the estimated effects for both prior-trial and prior-trial[2] did not change appreciably when the standard measures were added to the model.

Power-law behavior. We performed a power spectral analysis to quantify the power-law exponent for each trial separately. We used a fast Fourier transform to decompose the time series into its amplitude spectrum, a set of sine waves of varying amplitudes across a spectrum of frequencies (Aks, 2005). Frequency is inversely related to time (i.e., larger frequencies span smaller time scales, and smaller frequencies span greater time scales). The

amplitude of a sine wave at a given frequency thus represents fluctuation at a time scale corresponding to the frequency. The square of the absolute value of amplitude gives power (e.g., Handy, 2004). Hence, the square of the amplitude spectrum is the power spectrum. The slope of a log–log plot of frequency and power gives an estimate of the power-law exponent. (Following the recommendations of Edwards et al. [2007], we contrasted the fit of the power-law function with the exponential and gamma functions. For all of the participants, the power-law function provided a significantly better fit. Details of this analysis are presented in the Appendix.)

We performed this analysis on the time series of changes in point-of-gaze for each trial separately. Figure 10 shows the average power-law exponent as a function of trials for the participants who were at risk for discovering parity. The light gray line shows the participants who eventually discovered parity; the darker line shows those who did not discover it during the experiment. Note that the participants contributed to the average for all trials up to

the trial on which they discovered parity, but not beyond. As was predicted, the participants who discovered parity showed an increase in power-law behavior and a subsequent decrease over trials. That is, the degree of activity across the nested structure, indexed by the power-law exponent, first increased and then decreased across trials. Nondiscoverers did not show this systematic change in power-law exponents. Growth-curve analysis confirmed that the participants who discovered parity had a significantly larger trial^2 term ($B = -0.0001$) than those who did not [$B = -0.00004$; change in $-2LL$, $\chi^2(1) = 9.60$]. The parameters for the intercept ($B = 0.27$) and the linear effect of trial ($B = 0.002$) were not significantly different for the participants who discovered parity and those who did not (i.e., there was no effect of discovery vs. nondiscovery on these parameters). The model also included a random effect on the intercept [$\sigma^2 = 0.002$; change in $-2LL$, $\chi^2(1) = 146.00$]. The random effect of trial was estimated to be very close to zero and was thus trimmed from the model. The model would not converge when a random effect of trial^2 was included.

Of course, individual discoveries were distributed across trials ($M = 20$, $SD = 8.2$). Therefore, the rate of increase and decrease in power-law behavior, captured by the parameters for the quadratic effect of trial, should depend on the timing of discovery. To test this prediction, we used the timing of each participant's discovery of parity (i.e., the distance between the current trial and their discovery trial) as a level-two predictor in the model. This allows the model to capture changes in the power-law trajectory as a function of the impending discovery of parity. We included an effect of timing on both trial and trial^2. The effect of the timing of discovery on trial ($B = 0.0002$) was not significant [change in $-2LL$, $\chi^2(1) = 2.86$]. The form of the quadratic effect of trial depended on the timing of discovery, such that the increase and decrease occurred more rapidly for the participants who discovered parity earlier [$B = -0.0001$; change in $-2LL$, $\chi^2(1) = 4.90$]. To illustrate the results of the analysis graphically, the two insets in Figure 10 show effects of the estimated parameters for the participants who discovered parity relatively early (A) and later (B).

DISCUSSION

We showed that the discovery of a new mathematical representation of a problem can be predicted from measures of dynamical organization. Specifically, we demonstrated that a decrease in the entropy of the system predicts the discovery of the parity relation. Decreases in entropy indicate that the system has become more orderly. We also demonstrated that systematic changes in power-law behavior anticipate discovery. Power-law behavior increases as the current configuration relaxes its structural constraints and then decreases as a new configuration coalesces. These findings replicate and extend previous work in which we showed that the discovery of the alternation relationship was predicted by changes in entropy and power-law behavior (Stephen et al., in press). The present results demonstrate that the same set of relationships predicts a transition

to a mathematical rule and that eye movements can be used to assess the changes in dynamic organization.

Our results are consistent with a growing body of work indicating that cognition is a self-organizing, complex system characterized by nonlinear dynamics (Dale, Roche, Snyder, & McCall, 2008; Spivey, 2007; Spivey & Dale, 2006). In nonlinear dynamics, new structures emerge from the multifarious, continuous interactions among microscopic elements of the system. Such systems are capable of creating new structure precisely because their dynamics are dominated by these interactions, rather than the activity of isolated components (Jensen, 1998). Previous work has shown that new structures in human movement emerge through self-organization (Kugler, Kelso, & Turvey, 1982). For example, the spontaneous shift from one pattern of motion (e.g., synchronous finger tapping) to another (e.g., asynchronous finger tapping) constitutes a phase transition (Bressler & Kelso, 2001). Similarly, Van Orden, Holden, and Turvey (2003) found that performance in a simple word-naming task exhibited a signature pattern of variation, often called *1/f noise*, indicative of systems that continually self-organize to a critical point, the point just prior to a phase transition. In their study, the transition is from a state of uncertainty regarding the presented string of letters to the production of phonological mapping. On the basis of analyses of performance in a number of classic tasks (e.g., mental rotation, visual search), Gilden (2001; Thornton & Gilden, 2005) concluded that this signature $1/f$ pattern was a general property of cognition, associated with the kind of memory that arises in dynamical systems. The present study suggests that the reach of self-organization extends to the spontaneous formation of new structures at the conceptual level. Even quite abstract concepts, such as a mathematical relation, emerge according to the principles of self-organization.

Power-law behavior has been observed at other scales in cognitive performance, most notably in the relationship between practice and speed of response. In a variety of tasks, gains in speed are greatest early in learning and then tail off asymptotically as a function of practice (e.g., Lee & Anderson, 2001; Palmeri, 1999; Rickard, 1997). The present analyses find power-law relations in the microlevel fluctuations of behavior within a single trial. Although we would speculate that these two scales are ultimately connected to one another, the present findings do not allow us to address this issue. However, Lee and Anderson (2001) demonstrated that learning occurred at many different scales within a single complex task and that power-law functions at these scales had similar exponents.

The analysis offered here complements an extensive literature on self-organization in computational models of cognition. Hebbian learning and self-organizing maps (e.g., Kohonen maps) are among the most prominent learning algorithms in computational models of self-organization. A wide variety of cognitive phenomena have been addressed with these computational models, including critical periods (Munakata & McClelland, 2003), vocabulary acquisition (Li et al., 2007), and the development of categories and concepts (Cohen et al., 2002; Schyns, 1991). The ability of self-organizing models to capture the

core features of such a broad range of phenomena speaks to their potential power. Recent work shows that the principles of self-organization, outlined above, govern the behavior of these models. Thus, in addition to demonstrating the potential ability of these neurally plausible algorithms to generate interesting new structures, it is possible to test a set of behavioral predictions from the broader theory under which they operate. The present study demonstrates how two measures from nonlinear dynamics can be applied to eyetracking data to test these predictions.

We note that the approach pursued here differs substantially from most current approaches to problem solving. One major difference is that our approach assumes explicitly that cognition runs on the physical interactions in a complex system. The physical interactions causally change the way the system functions, resulting in the transition from one behavior to another. Most approaches to problem solving assume that the informational or semantic content of the system has causal status. For example, the semantic content of a goal state is evaluated against the projected outcome of a particular move (e.g., Chronicle et al., 2004). The extent of agreement between these semantic values has causal implications for the state of the system. One way to view this fundamental difference between these two approaches is that the explanations are at different levels of analysis. The microlevel fluctuations that are central to the physical-interaction account occur on much faster time scales than the internal comparisons of the information-based account.

It is currently unclear whether these different approaches can be aligned in a way that is mutually beneficial. However, as an example of how such an alignment might be pursued, consider the representational change theory (RCT) of insight problem solving (Knöblich, Ohlsson, Haider, & Rhenius, 1999; see also Jones, 2003, and Ormerod, MacGregor, & Chronicle, 2002). RCT suggests that constraint relaxation and chunk decomposition are important processes for understanding insight. Constraint relaxation involves breaking some of the semantically based restrictions on the problem solution. Chunk decomposition involves breaking experience-based organizations or groupings of the current information. Both of these processes involve breaking the bonds that support existing structures and ultimately resetting them to form new ones. This sequence of events is much like the one predicted by the theory of self-organization. Explicitly aligning these two approaches empirically might offer a starting point for bridging information-based and physical-interaction accounts.

Another way to think about the implications of the present research is to compare the relationship between the information-based and physical-interaction accounts to the relationship between Newtonian and quantum mechanics. Newtonian mechanics captures a wide variety of physical phenomena in aggregate; quantum mechanics is largely ignorable in aggregate (e.g., Simon, 2002). However, Newtonian descriptions fail to account for relatively microscopic—even quantum—fluctuations that lead to reliable deviations in individual trajectories (Nadeau & Kafatos, 1999). Stewart and Golubitsky (1992) discussed precisely this issue in the context of a falling milk droplet that splashes into a bowl of milk. The vertical trajectory of the droplet and the 24-point crown resulting from the splash may be predicted from a Newtonian description of a homogeneous sphere of fluid. However, the exact shape and orientation of the 24-point crown at the splash is outside the scope of purely Newtonian mechanics. These features are due to the precise, fine-grained, stochastic elements that a Newtonian description does not include: heterogeneities in the milk droplet, the milk surface in the bowl, and the air current in between. Newtonian mechanics sets a relatively coarse set of constraints on what we may expect in the aggregate, but Newtonian descriptions average across functionally infinite variations in the actual trajectories that unfold over time (Mandelbrot, 1982; Shlesinger, Zaslavsky, & Klafter, 1993).

We may extend the example of the milk droplet to the gear task. For this purpose, let us compare the fall of the milk droplet with practice with the gear task and the 24-point crown with the discovery of parity. Given unlimited trials of practice with the gear task, we may expect that all high-functioning cognitive systems will discovery parity, just as the spherical droplet will always produce the same 24-point crown upon splashing. Here, we have the symbolic description of the discovery of parity. The present research is an attempt to chip away at the individual-level temporal differences in this change. The conventional expectation is that relatively coarse-grained measures of performance (i.e., reaction times, accuracy, fixational eye movements) obey symbolic constraints and will mark changes in focus or attention that will herald cognitive change (e.g., van Gompel, Fischer, Murray, & Hill, 2007). This expectation provides no predictive insight into the discovery of parity, as is evidenced by the repeated null results in the Response Time and Accuracy and the Number and Duration of Fixation sections of the Results section. The symbolic description of cognitive behavior provides no ready account for this cognitive change; indeed, there is doubt whether a strictly symbolic description of cognition can account for any cognitive change (Fodor, 2000). On the other hand, dynamical systems analyses of the fine-grained fluctuations in eye movements have proven effective for predicting the discovery of alternation. Unlike the symbolic description of cognitive behavior in the gear task, dynamical systems theory has a ready account for the emergence of new structure, through the changes in microscale interactions. Symbolic descriptions of cognitive behavior may serve in the aggregate to describe overall regularities at the coarser scales, but we propose that this physical-interaction account will serve to predict the transition of the cognitive system from one representation to another.

A potentially surprising aspect of the present findings is that the time series of angular changes in point-of-gaze yielded information about the dynamics of the larger system in which the eyes are embedded. The ability to assess the dynamics of cognition from eyetracking data opens the door to dynamical analysis for a large number of problems in the field. Eyetracking has been used in a wide variety of different domains (e.g., categorization, object perception, language comprehension) and could, in prin-

ciple, be applied to any paradigm in which information is visually displayed (e.g., Henderson & Ferreira, 2004), and even in some cases without visual display (Spivey & Geng, 2001). Therefore, the present study represents a significant step toward bridging the divide between data and self-organization within cognitive science.

AUTHOR NOTE

REFERENCES

ABARBANEL, H. D. I. (1996). *Analysis of observed chaotic data*. New York: Springer.

AKS, D. J. (2005). 1/*f* dynamic in complex visual search: Evidence for self-organized criticality in human perception. In M. A. Riley & G. C. Van Orden (Eds.), *Tutorials in contemporary nonlinear methods for the behavioral sciences* (pp. 326-359). Arlington, VA: National Science Foundation. Retrieved March 1, 2006, from www.nsf.gov/sbe/bcs/pac/nmbs/nmbs.jsp.

AKS, D. J., ZELINSKY, G. J., & SPROTT, J. C. (2002). Memory across eye-movements: 1/*f* dynamic in visual search. *Nonlinear Dynamics, Psychology, & Life Sciences*, **6**, 1-25. doi:10.1023/A:1012222601935

ANDERSON, J. R., & SCHOOLER, L. J. (1991). Reflections of the environment in memory. *Psychological Science*, **2**, 396-408. doi:10.1111/j.1467-9280.1991.tb00174.x

ASHBY, F. G., ENNIS, J. M., & SPIERING, B. J. (2007). A neurobiological theory of automaticity in perceptual categorization. *Psychological Review*, **114**, 632-656. doi:10.1037/0033-295X.114.3.632

BRESSLER, S. L., & KELSO, J. A. S. (2001). Cortical coordination dynamics and cognition. *Trends in Cognitive Sciences*, **5**, 26-36. doi:10.1016/S1364-6613(00)01564-3

CHRONICLE, E. P., MACGREGOR, J. N., & ORMEROD, T. C. (2004). What makes an insight problem? The roles of heuristics, goal conception, and solution recoding in knowledge-lean problems. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **30**, 14-27. doi:10.1037/0278-7393.30.1.14

COHEN, L. B., CHAPUT, H. H., & CASHON, C. H. (2002). A constructivist model of infant cognition. *Cognitive Development*, **17**, 1323-1343. doi:10.1016/S0885-2014(02)00124-7

DALE, R., ROCHE, J., SNYDER, K., & MCCALL, R. (2008). Exploring action dynamics as an index of paired-associate learning. *PLoS One*, **3**, e1728. doi:10.1371/journal.pone.0001728

DIXON, J. A., & BANGERT, A. S. (2004). On the spontaneous discovery of a mathematical relation during problem solving. *Cognitive Science*, **28**, 433-449. doi:10.1016/j.cogsci.2003.12.004

DIXON, J. A., & KELLEY, E. A. (2006). The probabilistic epigenesis of knowledge. In R. V. Kail (Ed.), *Advances in child development and behavior* (Vol. 34, pp. 323-361). New York: Academic Press.

DIXON, J. A., & KELLEY, E. A. (2007). Theory revision and redescription: Complementary processes in knowledge acquisition. *Current Directions in Psychological Science*, **16**, 111-115. doi:10.1111/j.1467-8721.2007.00486.x

EDWARDS, A. M., PHILLIPS, R. A., WATKINS, N. W., FREEMAN, M. P., MURPHY, E. J., AFANASYEV, V., ET AL. (2007). Revisiting Lévy flight search patterns of wandering albatrosses, bumblebees and deer. *Nature*, **449**, 1044-1048. doi:10.1038/nature06199

FODOR, J. A. (2000). *The mind doesn't work that way: The scope and limits of computational psychology*. Cambridge, MA: MIT Press.

GENTNER, D., LOEWENSTEIN, J., & HUNG, B. (2007). Comparison facilitates children's learning of names for parts. *Journal of Cognition & Development*, **8**, 285-307. doi:10.1080/15248370701446434

GENTNER, D., & NAMY, L. L. (1999). Comparison in the development of categories. *Cognitive Development*, **14**, 487-513. doi:10.1016/S0885-2014(99)00016-7

GILDEN, D. L. (2001). Cognitive emissions of 1/*f* noise. *Psychological Review*, **108**, 33-56. doi:10.1037/0033-295X.108.1.33

GRAEPEL, T., BURGER, M., & OBERMAYER, K. (1997). Phase transitions in stochastic self-organizing maps. *Physical Review*, **56E**, 3876-3890. doi:10.1103/PhysRevE.56.3876

GREBOGI, C., OTT, E., ROMEIRAS, F., & YORKE, J. A. (1987). Critical exponents for crisis-induced intermittency. *Physical Review*, **36A**, 5365-5380. doi:10.1103/PhysRevA.36.5365

HADLEY, R. F., & CARDEI, V. C. (1999). Language acquisition from sparse input without error feedback. *Neural Networks*, **12**, 217-235. doi:10.1016/S0893-6080(98)00139-7

HANDY, T. C. (2004). *Event-related potentials: A methods handbook*. Cambridge, MA: MIT Press.

HARRISON, L. G., & HILLIER, N. A. (1985). Quantitative control of *Acetabularia* morphogenesis by extracellular calcium: A test of kinetic theory. *Journal of Theoretical Biology*, **114**, 177-192. doi:10.1016/S0022-5193(85)80261-7

HENDERSON, J. M., & FERREIRA, F. (EDS.) (2004). *The interface of language, vision, and action: Eye movements and the visual world*. New York: Psychology Press.

HILBORN, R. C. (1994). *Chaos and nonlinear dynamics: An introduction for scientists and engineers*. New York: Oxford University Press.

JENSEN, H. J. (1998). *Self-organized criticality: Emergent complex behavior in physical and biological systems*. Cambridge: Cambridge University Press.

JONES, G. (2003). Testing two cognitive theories of insight. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **29**, 1017-1027. doi:10.1037/0278-7393.29.5.1017

KELSO, J. A. S. (1995). *Dynamic patterns: The self-organization of brain and behavior*. Cambridge, MA: MIT Press.

KNÖBLICH, G., OHLSSON, S., HAIDER, H., & RHENIUS, D. (1999). Constraint relaxation and chunk decomposition in insight problem solving. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **25**, 1534-1555. doi:10.1037/0278-7393.25.6.1534

KNÖBLICH, G., OHLSSON, S., & RANEY, G. E. (2001). An eye movement study of insight problem solving. *Memory & Cognition*, **29**, 1000-1009.

KOWLER, E. (1990). *Eye movements and their role in visual and cognitive processes*. New York: Elsevier.

KUGLER, P. N., KELSO, J. A. S., & TURVEY, M. T. (1982). On the control and coordination of naturally developing systems. In J. A. S. Kelso & J. E. Clark (Eds.), *The development of movement control and coordination* (pp. 5-78). Chichester, U.K.: Wiley.

KUGLER, P. N., & TURVEY, M. T. (1987). *Information, natural law, and the self-assembly of rhythmic movement*. Hillsdale, NJ: Erlbaum.

LEE, F. J., & ANDERSON, J. R. (2001). Does learning a complex task have to be complex? A study in learning decomposition. *Cognitive Psychology*, **42**, 267-316. doi:10.1006/cogp.2000.0747

LI, P., ZHAO, X., & MACWHINNEY, B. (2007). Dynamic self-organization and early lexical development in children. *Cognitive Science*, **31**, 581-612. doi:10.1080/15326900701399905

LORENZ, E. N. (1963). Deterministic nonperiodic flow. *Journal of the Atmospheric Sciences*, **20**, 130-141. doi:10.1177/0309133308091948

MANDELBROT, B. B. (1982). *The fractal geometry of nature*. San Francisco: Freeman.

MARWAN, N., ROMANO, M. C., THIEL, M., & KURTHS, J. (2007). Recurrence plots for the analysis of complex systems. *Physics Reports*, **438**, 237-329. doi:10.1016/j.physrep.2006.11.001

MCCLELLAND, J. L. (2006). How far can you go with Hebbian learning, and when does it lead you astray? In Y. Munakata & M. Johnson (Eds.), *Processes of change in brain and cognitive development: Attention and performance XXI* (pp. 33-69). Oxford: Oxford University Press.

MCCLELLAND, J. L., RUMELHART, D. E., & PDP RESEARCH GROUP (EDS.) (1986). *Parallel distributed processing: Explorations in the microstructure of cognition* (Vol. 1). Cambridge, MA: MIT Press.

MIRMAN, D., DIXON, J. A., & MAGNUSON, J. S. (2008). Statistical and computational models of the visual world paradigm: Growth curves and individual differences. *Journal of Memory & Language*, **59**, 475-494. doi:10.1016/j.jml.2007.11.006

MUNAKATA, Y., & MCCLELLAND, J. L. (2003). Connectionist models of

development. *Developmental Science*, **6**, 413-429. doi:10.1111/1467
-7687.00296

Nadeau, R. L., & Kafatos, M. (1999). *The non-local universe*. Oxford:
Oxford University Press.

Nicolis, G., & Prigogine, I. (1977). *Self-organization in nonequilib-rium systems: From dissipative structures to order through fluctua-tions*. New York: Wiley.

Ormerod, T. C., MacGregor, J. N., & Chronicle, E. P. (2002). Dy-namics and constraints in insight problem solving. *Journal of Experi-mental Psychology Learning, Memory, & Cognition*, **28**, 791-799. doi:10.1037/0278-7393.28.4.791

Palmeri, T. J. (1999). Theories of automaticity and the power law of practice. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **25**, 543-551. doi:10.1037/0278-7393.25.2.543

Prigogine, I., & Stengers, I. (1984). *Order out of chaos*. New York: Bantam.

Rickard, T. C. (1997). Bending the power law: A CMPL theory of strategy shifts and the automatization of cognitive skills. *Journal of Experimental Psychology: General*, **126**, 288-311. doi:10.1037/0096 -3445.126.3.288

Schwartz, D. L., & Black, J. B. (1996). Shuttling between depictive models and abstract rules: Induction and fallback. *Cognitive Science*, **20**, 457-497. doi:10.1016/S0364-0213(99)80012-3

Schyns, P. G. (1991). A modular neural network model of concept acquisition. *Cognitive Science*, **15**, 461-508. doi:10.1016/0364 -0213(91)80016-X

Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, **27**, 379-423, 623-656.

Shlesinger, M. F., Zaslavsky, G. M., & Klafter, J. (1993). Strange kinetics. *Nature*, **363**, 31-37. doi:10.1038/363031a0

Siegler, R. S., & Araya, R. (2005). A computational model of con-scious and unconscious strategy discovery. In R. V. Kail (Ed.), *Ad-vances in child development and behavior* (Vol. 33, pp. 1-42). Oxford: Elsevier.

Silberman, Y., Bentin, S., & Miikkulainen, R. (2007). Semantic boost on episodic associations: An empirically-based computational model. *Cognitive Science*, **31**, 645-671. doi:10.1080/15326900701399921

Simon, H. A. (2002). Near decomposability and the speed of evolu-tion. *Industrial & Corporate Change*, **11**, 587-599. doi:10.1093/icc/ 11.3.587

Sims, D. W., Southall, E. J., Humphries, N. E., Hays, G. C., Brad-shaw, C. J. A., Pitchford, J. W., et al. (2008). Scaling laws of ma-rine predator search behavior. *Nature*, **451**, 1098-1102. doi:10.1038/ nature06518

Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analy-sis*. New York: Oxford University Press.

Siri, B., Quoy, M., Delord, B., Cessac, B., & Berry, H. (2007). Ef-fects of Hebbian learning on the dynamics and structure of random networks with inhibitory and excitatory neurons. *Journal of Physiol-ogy*, **101**, 136-148. doi:10.1016/j.jphysparis.2007.10.003

Sparrow, C. (1982). *Lorenz equations: Bifurcations, chaos, and strange attractors*. New York: Springer.

Spivey, M. J. (2007). *The continuity of mind*. New York: Oxford Uni-versity Press.

Spivey, M. J., & Dale, R. (2006). Continuous dynamics in real-time cognition. *Current Directions in Psychological Science*, **15**, 207-211. doi:10.1111/j.1467-8721.2006.00437.x

Spivey, M. J., & Geng, J. J. (2001). Oculomotor mechanisms activated by imagery and memory: Eye movements to absent objects. *Psycho-logical Research*, **65**, 235-241. doi:10.1007/s004260100059

Stephen, D. G., Dixon, J. A., & Isenhower, R. W. (in press). Dynam-ics of representational change: Entropy, action, and cognition. *Journal of Experimental Psychology: Human Perception & Performance*.

Stewart, I., & Golubitsky, M. (1992). *Fearful symmetry: Is God a geometer?* Oxford: Blackwell.

Sullivan, T. J., & de Sa, V. R. (2006). Homeostatic synaptic scaling in self-organizing maps. *Neural Networks*, **19**, 734-743. doi:10.1016/j .neunet.2006.05.006

Takens, F. (1981). Detecting strange attractors in turbulence. In D. Rand & L. S. Young (Eds.), *Lecture notes in mathematics*: *Dynamical sys-tems and turbulence* (Vol. 898, pp. 366-381). Berlin: Springer.

Thornton, T. L., & Gilden, D. L. (2005). Provenance of correlations in psychological data. *Psychonomic Bulletin & Review*, **12**, 409-441.

van Gompel, R. P. G., Fischer, M. H., Murray, W. S., & Hill, R. L. (Eds.) (2007). *Eye movements: A window on mind and brain*. Oxford: Elsevier.

Van Orden, G. C., Holden, J. G., & Turvey, M. T. (2003). Self-organization of cognitive performance. *Journal of Experimental Psy-chology: General*, **132**, 331-350. doi:10.1037/0096-3445.132.3.331

Webber, C. L., Jr., & Zbilut, J. P. (1994). Dynamical assessment of physiological systems and states using recurrence plot strategies. *Journal of Applied Physiology*, **76**, 965-973.

Webber, C. L., Jr., & Zbilut, J. P. (2005). Recurrence quantifica-tion analysis of nonlinear dynamical systems. In M. A. Riley & G. C. Van Orden (Eds.), *Tutorials in contemporary nonlinear meth-ods for the behavioral sciences* (pp. 26-94). Arlington, VA: Na-tional Science Foundation. Retrieved March 1, 2006, from www .nsf.gov/sbe/bcs/pac/nmbs/nmbs.jsp.

Webster, G., & Goodwin, B. (1996). *Form and transformation: Gen-erative and relational principles in biology*. Cambridge: Cambridge University Press.

West, G. B., Brown, J. H., & Enquist, B. J. (1999). The fourth dimen-sion of life: Fractal geometry and allometric scaling of organisms. *Sci-ence*, **284**, 1677-1679. doi:10.1126/science.284.5420.1677

APPENDIX

The presence or absence of power-law behavior can be assessed by the structure in the power spectrum of a time series. Power and frequency would be uncorrelated (i.e., completely horizontal) for Gaussian noise. How-ever, in the case of $1/f$ noise, power P is an inverse power law of frequency f:

$$P \propto \frac{1}{f^{\beta}}, \tag{A1}$$

where β is the power-law exponent defining the steepness of the curve (see the top panel of Figure A1).

The power law specifies an invariant relationship between power and frequency. That is, the change in power (i.e., fluctuation) is always the same proportion of the change in frequency (i.e., inversely, time scale), no matter the frequency. It becomes especially clear when we plot the power law on logarithmic axes (see the bottom panel of Figure A1). The power law becomes a line with slope $-\beta$.

There are, of course, other ways to generate a power spectrum with higher power in the lower frequencies. For example, consider the exponential function

$$P \propto \frac{\lambda}{e^{\lambda f}}, \tag{A2}$$

where λ is a rate parameter specifying the change in power. Whereas the unmodulated β in Equation A1 ensures that power-law behavior is scale invariant, the modulation of rate λ by f in Equation A2 makes exponential be-havior scale dependent. The exponential curve bears passing resemblance to the power law on regular axes (see the top panel of Figure A2). However, whereas the power law exhibits invariantly linear structure on logarithmic

APPENDIX (Continued)

axes, the curvilinear structure of the exponential curve on logarithmic axes shows its sensitivity to scale (see the bottom panel of Figure A2).

Asserting that a particular data set demonstrates power-law behavior thus requires testing the power-law fit against the fit of other skewed distributions, such as the exponential function and the gamma function. Traditionally, a comparison of the ordinary least-squares r^2 of competing distributional fits has sufficed (e.g., Anderson & Schooler, 1991). However, recent research in statistical mechanics has demonstrated that maximum likelihood
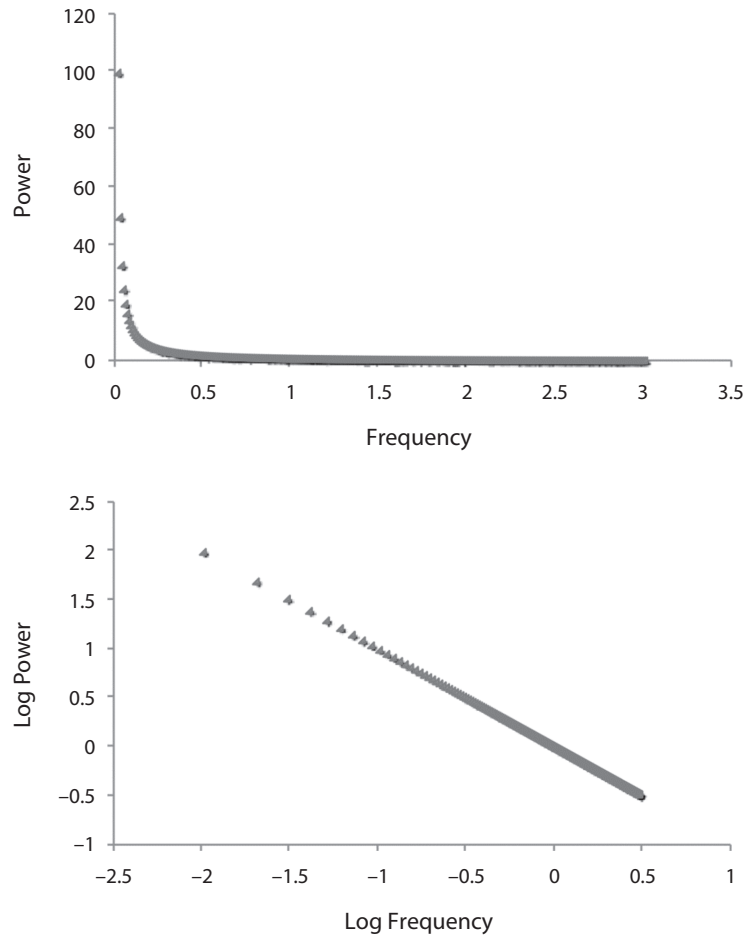

Figure A1. The top panel shows an example of a power-law relationship in original scales. Power is plotted as a function of frequency. The lower panel shows the same relationship plotted in log–log scales.

estimation is a more appropriate strategy for assessing the relative likelihood of power laws over other skewed distributions (Edwards et al., 2007; Sims et al., 2008). Maximum likelihood estimation generates the log likelihood of each candidate distributional fit for a given data set, and the best-fitting candidate distribution will have the highest log likelihood.

In the present study, we used the fast Fourier transform to produce the power spectrum of the time series of angular change in eye movements for each trial. Because each participant completed 36 trials, we calculated 36 different power spectra for each participant. We used maximum likelihood estimation to fit a power-law (i.e., generalized Pareto) distribution, an exponential distribution, and a gamma distribution to each power spectrum. We compared the log likelihood for power-law fit with respect to those for gamma fit and exponential fit of each power spectrum. For each participant, we ran two paired-samples t tests over 36 power-spectrum fits to test the difference between exponential and power-law log likelihoods and the difference between the gamma and power-law log likelihoods. For all of the participants, power-law log likelihoods were greater than exponential log likelihoods [$t(35)$s ranging from 3.58, $p < .001$, to 11.25, $p < .0001$, with a median of 4.94, $p < .0001$]. For

APPENDIX (Continued)

all of the participants, power-law log likelihoods were also greater than gamma log likelihoods [$t(35)$s ranging from 3.26, $p < .01$, to 10.26, $p < .0001$, with a median of 5.58, $p < .0001$]. In brief, for all of the participants, the power-law log likelihoods were significantly greater than those for the other candidate distributions.
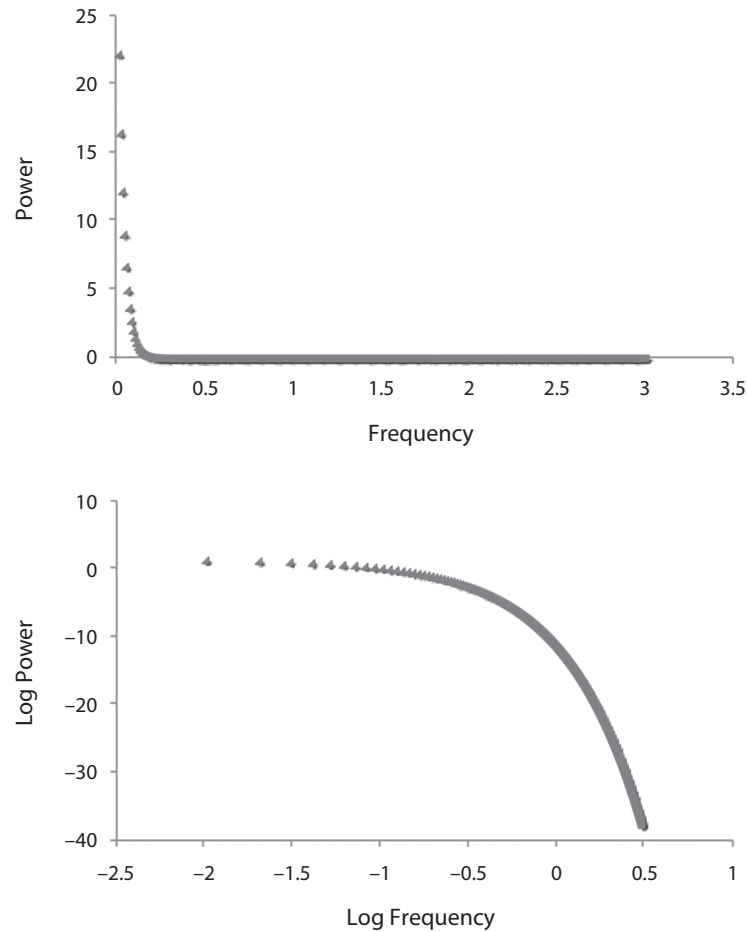
Figure A2. The top panel shows an example of an exponential relationship in original scales. Power is plotted as a function of frequency. The lower panel shows the same relationship plotted in log–log scales.