6-13-2015

THE EMERGENCE OF OUTLIERS IN ENTREPRENEURSHIP: A SELF-ORGANIZED CRITICALITY FRAMEWORK

G. Christopher Crawford
Ohio University, USA, christopher.crawford@ohio.edu

Recommended Citation
Available at: http://digitalknowledge.babson.edu/fer/vol35/iss16/1

This Paper is brought to you for free and open access by the Entrepreneurship at Babson at Digital Knowledge at Babson. It has been accepted for inclusion in Frontiers of Entrepreneurship Research by an authorized editor of Digital Knowledge at Babson. For more information, please contact digitalknowledge@babson.edu.
THE EMERGENCE OF OUTLIERS IN ENTREPRENEURSHIP: A SELF-ORGANIZED CRITICALITY FRAMEWORK

G. Christopher Crawford, Ohio University, USA

Abstract
At the beginning of 2015, in an inductive investigation of four entrepreneurship-related data sets (N=12,000+), Crawford, Aguinis, Lichtenstein, Davidsson, and McKelvey discovered power law distributions in all resource-, cognition-, action-, and environmental-based input variables, as well as all revenue-, employee-, and growth-based outcome variables. Given the large number of outliers in these distributions, the findings call for new theory and method to identify the mechanism that generates all observations. Grounded in a complexity science paradigm, I propose the mechanism to be self-organized criticality—where thresholds in the distribution have the potential to produce large, nonlinear, and cascading effects—and empirically demonstrate that outliers are most likely to emerge when venture endowments and expectations for future growth are above critical values.

Introduction
How did Facebook become Facebook? While an N=1 study describing the astronomical growth of a firm like this may lack generalizability across all new ventures, there is great interest in the domain to explain and predict the emergence of outliers from the entire population of entrepreneurs (Acs, 2008). Unfortunately, theory development on firm growth has been extremely fragmented due to conflicting empirical results (Simon, 2009), inconsistent growth measures across studies cross-sectional studies that exhibit survival bias, or small samples that are of limited generalizability to the full realm of entrepreneurial activity (Aldrich & Ruef, 2006; Shepherd & Wiklund, 2009).

This is an important problem for entrepreneurship scholars. The consistently inconsistent empirical findings and the lack of a dominant theoretical framework to explain the full range of antecedents and consequences of entrepreneurial action at multiple levels reduces the field’s ability to make claims of causal inference—claims that would be of benefit to practice, policy, pedagogy, and domain legitimacy (Busenitz et al., 2003; Shane, 2012). Moreover, the skewed distributions mentioned above exacerbate this problem because the domain’s principal hypothesis testing techniques are based on Gaussian assumptions of normal distributions, independent observations, and linear relationships among constructs (Dean, Shook, & Payne, 2007). Most importantly, these techniques cannot accurately model emergent phenomena like entrepreneurship, where systems transition from non-existence to existence (Katz & Gartner, 1988; Lichtenstein et al., 2007). Though these theoretical and methodological difficulties have pervaded the domain of entrepreneurship research since its inception, new findings may proffer an empirical foundation upon which a comprehensive theory could be built.

Most recently, in an inductive empirical investigation of four entrepreneurship-related data sets—including the Panel Study of Entrepreneurial Dynamics II, Comprehensive Australian Study
ENTREPRENEUR OTHER

of Entrepreneurial Emergence, Kauffman Firm Survey, and Inc. 5000 (N=12,000+)—Crawford, Aguinis, Lichtenstein, Davidsson, & McKelvey (2015) found power law distributions (PLDs) in all resource-, cognition-, action-, and environmental-based input variables, as well as all revenue-, employee-, and growth-based outcome measures. These distributions are important to entrepreneurship scholars because the theoretical constructs that explain them are scale-free, where system characteristics, processes, behaviors, and outcomes appear in self-similar (i.e., fractal) patterns across many orders of magnitude (Andriani & McKelvey, 2009). These distributions are indicators of critical, threshold-based phenomena (Dooley & Van de Ven, 1999) and, moreover, it is possible that a single mechanism drives the emergence of PLDs in each of the theoretical constructs (Newman, 2005).

The purpose of this paper is to investigate a causal mechanism purported to generate PLDs in other domains: self-organized criticality (Bak, 1996), which describes a stable point of disequilibrium, where the addition of one input has the potential to create cascading, nonlinear change at all interconnected levels. Grounded in the complexity science paradigm and seminal entrepreneurship research, I propose four meta-constructs within which self-organized criticality operated. These constructs can apply to every venture, regardless of scale: endowments, expectations, engagement, and environments. I proffer that these constructs provide a mutually exclusive—yet collectively exhaustive—framework for explaining how outliers emerge in entrepreneurship. I develop a hypothesis of autogenesis (Drazin & Sandelands, 1992), where feedback loops from potential stakeholders at the micro-level drive outcome expectations at the nascent organizing stage and, subsequently, exhibit self-similar patterns of outcomes at more aggregated levels of analysis.

The following section begins to develop a theory that can sufficiently describe the presence of PLDs in inputs and outcomes in entrepreneurship and explain the mechanism generating these distributions. I first outline the general dynamics underlying the formation of PLDs, and subsequently argue how simple rules of interaction based on the entrepreneur’s expected outcomes can result in distinct patterns of aggregated behavior—PLDs—at multiple levels. Then, I identify how these distributions have both linear and nonlinear qualities and pinpoint the critical threshold that divides the two. This threshold provides the basis for proposing self-organized criticality as the mechanism that drives PLDs in entrepreneurship, and explicates how this mechanism could work in the context of seminal entrepreneurship constructs. In the method section, the autogenesis hypothesis is tested with data from the Crawford et al. (2015) study, using nonparametric correlations and semi-parametric techniques to compute heretofore unpublished critical threshold estimates.

**Theory Development**

Power laws result from a deep, underlying pattern of emergence (Bar-Yam, 1997). As Brock (2000:29) describes, “Complexity science considers whether these patterns have a property of universality about them.” Such patterns are called scaling laws (Gell-Mann, 1988) because they are expressed as empirical regularities that apply across many orders of magnitude of a phenomenon. A theory to describe these regularities should be abductive, one that provides the inference toward the “best scalable explanation” by examining a mass of facts and suggesting a generative mechanism to explain the patterns in the data (Boisot & McKelvey, 2010).

In all of these systems, activity at the most micro-level aggregates to higher-order activity, thus outcomes at a macro-level are a result of lower-level aggregation (Lewin et al., 1999); the macro-level patterns of outcomes look similar to micro-level patterns. Thus, evidence of a power law at one level is an indication that similar dynamics may be acting at the preceding level (Andriani
& McKelvey, 2009). These patterns are recursive and feedback-driven. Without a significant top-down intervention, like system constraints (e.g., a socialist government) or cross-level negative feedbacks (e.g., suppliers can't deliver product as instructed), expectations for future outcomes drive repeated interactions with other agents in the system, behaviors perpetuate, and micro-level interactions generate emergent order at higher levels that reflect (and are reflected by) the same statistical dynamics across multiple levels (Sornette, 2006; West & Deering, 1995). Emergence, in the classic treatment of complex adaptive systems, occurs when order spontaneously appears at a higher level, without a centralized controller, as a result of the actions and interactions of myriad agents pursuing their own interests at a lower level in a system (Buckley, 1968; Stacey, 1995). Thus, through aggregation, the entire scope of the phenomenon can be driven by a single cognitive process, where an agent recursively interacts with other agents based on their expected outcomes (Beinhocker, 1997; Drazin & Sandelands, 1992).

**Autogenesis: Expectations, Moving Farther from Linearity, and Farther from Equilibrium**

How might this relationship manifest? Drawing from seminal complexity science arguments of autogenesis (where evolution is directed by innate factors independent of the organism and environment): an agent's rules that govern interaction drive a recursive, nonlinear process that engages cross-level agents and more emergent systems from the bottom up, eliciting expectations from other “interactants” to select actions based on available information about the meaning of the action and the preferred response (Goffman, 1967). Habitual action and interaction cause an individual to make choices based on rules that originated in past experiences and socialization; this is the "deep structure" that links individual action to expected reciprocal action (March & Simon, 1958). At intermediate levels of interaction, these rules are inferred, but may not be readily apparent; at higher levels, discernible patterns of “observed structure” emerge (Drazin & Sandelands, 1992). The macro-level patterns are the visually distinct power law distribution of outcomes (Andriani & McKelvey, 2009; Gell-Mann, 1988; Simon, 1955).

The entrepreneur's expectations become schemata—her 'simple rules'—for interacting for others in a system (Anderson, 1999; Drazin & Sandelands, 1992; McKelvey, 2004). These schemata drive how agents think, act, and interact to generate emergence. Anderson (1999) suggested that outcomes of a complex system are caused by the interactions across (a) schemata—decision-rules of agents that are self-regulated according to an agent's expected outcomes and responses to feedback loops; (b) initial conditions—an agent's resources; and (c) environmental dynamics—local competition and resource munificence. Expectations are central to transforming opportunities into new companies, for they drive both present and future interactions with customers, suppliers, investors and other stakeholders (DeKinder & Kohli, 2008). For example, just as schemata determine the perception and creation of opportunities (Alvarez & Barney, 2007), they also direct how an entrepreneur's pattern of interactions leads to legitimacy of the new firm (Delmar and Shane 2004). Further, customers use these interactions to construct an expected value of the firm's offering and its potential outcomes (Gartner, Bird, & Star, 1992; Vargo & Lusch, 2008). Likewise, a founder's expectations about the venture's future growth, and the explicit or implicit expression of that intent to stakeholders, also influences the perceived value and potential of the firm (Wiklund & Shepherd, 2003).

Expectations are a forecast of an outcome that are based on current information that the predicting party has at the time the prediction is made (Muth, 1961). I propose that an entrepreneur's current information is generated from endogenous knowledge and historical context of both current resource endowments (i.e., education, experience) and the availability of exogenous environmental resources. Since human performance is power law distributed (Aguinis...
& O’Boyle 2013; O’Boyle & Aguinis, 2012), it is likely that those in the tail of the distribution will expect consistent levels of high performance. Similarly, environmental resources like city population, innovations within cities, and venture capital availability are also power law distributed (Bettencourt et al. 2010), suggesting that individuals in California’s Silicon Valley or Boston’s VC area may have asymmetrical knowledge of potential resources that other in much smaller cities may not have. Together, expectations for future outcomes is highly skewed because of past experience or environmental context.

Though expectations have been shown to be ephemeral, volatile, and potentially destabilizing (Chiles et al. 2010; Shackle 1979), the combined linear and nonlinear properties inherent in power laws may explain why. A complexity perspective suggests that qualities like these are natural, nonlinear outcomes of human interactions (Cilliers 1998). If entrepreneurs’ expectations about future outcomes are power law distributed, where most aim for “normal” outcomes and a small percentage aspire to extreme outcomes, those at the high end are most likely envisioning something others have not seen. If this is the case, they will have to rely on repeatedly conveying an image of an envisioned future that involves the accumulation and recombination of resources in the environment—what Zott & Quy Nguyen (2007) call “symbolic management”—to potential stakeholders. If business model expectations are too grandiose and well outside the previous experience of the entrepreneur, then there may be some dissonance in personal self-efficacy that could eventually lead to abandoning the idea; as well, if stakeholders don’t perceive that the environment can support such expectations, or if the stakeholders don’t perceive the entrepreneur has sufficient experience to execute the envisioned model, then they may abandon the idea. As an agent interacts with the environment over time, feedback regulates these expectations. When agents receive feedback, it generates more interactions: positive feedback generates more of the same types of interactions, whereas negative feedback generates more of different types of interactions (Forster, et al., 2001; Stacy, 1995).

As Sornette (2006: 223) suggests, “Studying the correlations and their consequences is an essential part of the analysis of a system, since the correlations are the signatures that inform us about the underlying mechanisms.” Though cognitive processes have been shown to be causal explanations for behaviors and outcomes, our view is consistent with a complexity perspective which proposes that the nonlinearity of human interaction renders outcomes as non-deterministic (Rahmandad & Sterman 2008)—where the same action can produce disproportionately positive or negative effects. Additionally, consistent with my complexity argument, I do not suggest the relationship between expectations and outcomes is linear, where higher expectations lead to proportionately higher outcomes. Instead, I hypothesize that since expectations in the tail are nonlinear and have the potential to lead to both superlative achievement and cataclysmic catastrophe, the relationship between expectations and outcomes will be nonlinear.

The constructs to explain PLD phenomena must be scale-free and apply to multiple levels of analysis to account for the generating mechanism. Using the previous complexity argument, where an agent’s initial conditions, rules for interaction based on expectations for future outcomes, and environmental resources influence outcomes, I propose four meta-constructs that align with seminal entrepreneurship research. I first suggest that these initial conditions consist of resource-based constructs analyzed in the Crawford et al. (2015) study: human capital, social capital, intellectual capital, and financial capital—all of which are PLD. These are an individual’s, a team’s, and a venture’s initial endowments. The same study identified that expectations for future outcomes (i.e., revenue, number of employees, relative growth, and absolute growth) are PLD. Drawing from Song et al. (2010), human action is PLD, as are the number of activities undertaken and the time spent by entrepreneurs in pursuit of creating a new venture—this is the recursive
pattern of interaction, this is engagement. Finally, Andriani & McKelvey identify with more than 20 empirical studies which depict organizational environments as PLD; Crawford et al. (2015) find PLDs in all industry segments. Founded on the concept of self-organized criticality, entrepreneurial opportunities, therefore, are most readily discovered and created—and ventures are most likely to emerge—when any of the four meta-constructs exist beyond some critical threshold; conversely, when the constructs are sub-critical, ventures are more likely to experience “exponentially decaying activity, always dying out” (Sornette, 2006:396).

**Method and Results**

For theoretical generalizability and more robust causal influence claims, I test the hypotheses with longitudinal data from the PSED II, CAUSEE, KFS, and Inc. 5000. I analyze data in MATLAB software, using scripts and protocol from Clauset et al. (2009) to conduct semi-parametric maximum-likelihood bootstrap estimates of the entire distribution of all input and outcome variables used in the Crawford et al. (2015) study at the individual, team, and venture level over time. The estimates provide an empirical benchmark for practitioners, policy makers, and teachers to demonstrate the point at which the potential for nonlinearity is the highest. I also conduct nonparametric Spearman’s Rho correlations between nascent expectations for future venture revenue and employees with actual firm outcomes in the fifth year of operation.

Supporting Sornette’s (2006) argument that long-range correlations inform us about the underlying mechanisms, I conduct correlation analyses to test the hypothesis that a founder’s nascent expectations for venture outcomes will be associated with actual venture outcomes. We use two different tests on the same data to demonstrate the differential effects of using Gaussian statistical techniques on Pareto distributed data. First, results for a Pearson correlation (which assumes normal distributions) between a founder’s nascent expectations for a venture’s revenue at the fifth year of operations and actual revenue five years later is insignificant at the 0.05 level (0.122, p=0.077); similarly, the correlation is negative and insignificant between expectations for employees and actual number of employees in the fifth year (-0.005, p=0.477). It is interesting to note that that the linear association is negative for employees, supporting the Aguinis et al. (2013) claim that power laws change substantive conclusions of research findings—including the existence, direction, and size of effect. Since extant research demonstrates that both expectations for future growth and outcomes are power law distributed, however, I test the correlation with Spearman’s rho, a nonparametric rank-order statistic. Instead of relying on actual values, each variable is ranked according to its value in the sample, and then the ranks are correlated. This is more appropriate for our analysis because, like our Pareto analysis suggests, the actual values by themselves are not necessarily important, their importance lies in their value relative to the rest of the population. Expectations for year five employees is significantly correlated with actual employees with a rho of 0.345, and expectations for year five revenue is significantly correlated with actual revenue with a rho correlation of 0.622, with p<0.001 for both.

There are distinct patterns of emergence among the populations of nascent, new, and young ventures. Based on similar scaling exponents within and among the constructs, the dynamics of what entrepreneurs do (engagement) is consistent with what they have (endowments); however, both of these appear to be misaligned with their expectations (i.e., what they envision) for their venture.

**Discussion and Implications**

Based on the analysis above, expectations drive outcomes—and outlier outcomes primarily emerge when expectations for growth begin in the tail of the distribution, beyond the critical threshold of 26. Combining extant research and support for the autogenesis hypothesis, all
constructs are scale-free: they are power law distributed no matter what variables are used to measure them at the individual-, team-, and venture-level of analysis. The framework also suggests that outliers and extreme events drive Schumpeterian change by motivating action from the bottom up, creating competitive imperfections, instigating network feedback effects, while recombining and transforming resources from the environment.

A power law-based framework can more accurately explain the dynamics of influential factors on new venture outcomes. Ecology scholars use an evolutionary perspective to study population dynamics—the combination of resource munificence and the competitive density of existing firms (Aldrich 1990). Population dynamics create opposing tensions on new ventures. As one tension, the abundance of perceived resources (e.g., opportunities like quality or quantity of labor available, gross metropolitan product, total innovations) pulls individuals in to the market. Empirical studies show that environmental resources like these are not only highly skewed, they are also highly correlated, with those in the tail of the distribution exhibit the same underlying dynamics. Bettencourt et al. (2010) find that Gross Metropolitan Product—the value of all goods and services produced within a given region—is power law distributed in the United States. The size of cities and the income of individuals within each GMP are power law distributed. Those environments in the tail of the distribution have a lot more people available in the labor pool, leading to more total creative people in the labor pool, leading to more patents filed in that city. The authors find that the GMP explains 65%-97% of variance in all of the sub-distributions, and all of these variables have power law exponents of ~2.15. Stating the obvious: there are cities in the tail—like New York and Los Angeles—that have underlying interaction dynamics that are qualitatively and quantitatively different than smaller cities. Thus, resources in the environment, no matter how they are measured, are power law distributed, and those areas in the tail have qualities similar to gravity: they pull in additional resources.

Similarly, empirical studies of elite performers support the view that outlier entrepreneurs are like stars: once they emerge beyond a critical threshold in the distribution, their gravity starts to attract resources and produce co-evolutionary effects on the environment (Aguinis and O’Boyle 2013). Here, outliers can create or discover opportunities, depending on their expected outcomes. According to the theory we are proposing, the complex driver of opportunity tension leads to entrepreneurial outcomes that follow a Pareto distribution, where the vast majority of entrepreneurs express a ‘normal’ degree of opportunity tension which leads to linear (normal) outcomes, and a small percent of entrepreneurs express a non-normal, exceptionally high degree of opportunity tension that leads to non-normal outcomes. Describing these differences is essential for understanding the recursive dynamics of the process.

My findings lend insight to research and practicing entrepreneurs, suggesting that ventures with expectations or endowments below critical thresholds of the PLD must engage with potential stakeholders at a quality and quantity in the tail of the distribution to substantively scale the venture. This provides a foundation for future causal inference claims in entrepreneurship, where success—for both scholars and practitioners—can be assessed in relation to: 1) the founder’s subjective expectations for growth, and 2) the venture’s objective outcomes in the power law distribution.

CONTACT: Christopher.Crawford@Ohio.edu; (T) 513-295-8717; Ohio University, 304 Copeland Hall, Athens, OH 45701 USA.

ACKNOWLEDGEMENT

I would like to thank the Ewing Marion Kauffman Foundation for its financial assistance with some portions of this research.