Adjusting the Novelty Thermostat: Courting Creative Success Through Judicious Randomness*

Kyle E. Jennings
jennings@berkeley.edu
Institute of Personality and Social Research
4143 Tolman Hall, University of California, Berkeley, CA 94720

Abstract

Though creativity is usually defined as "novel and appropriate," this is most often understood to mean "as novel as possible, so long as appropriate." While this definition might be suitably applied to finished products, it is less obviously useful as a guiding value during the act of creation. This research tests this and other definitions by using computer simulations based on Campbell's "blind variation and selective retention" theory. Introducing a "temperature" parameter to reduce novelty's importance over time produces results superior to both an even combination of novelty and appropriateness and the prevailing "novel, so long as appropriate" definition. However, choosing the correct temperature adjustment schedule is essential. In this way, the simulations show that human decision processes might plausibly produce the same dynamics as Simulated Annealing, and thus that metaheuristic optimization can be an insightful theoretical guide for creativity researchers. The results also show the feasibility of Campbell's theory. Finally, they show that if creativity is to be a guiding value while creating, then it is best defined as "appropriately novel, and appropriate."

The overall action of creative systems is to advance the status quo. This is difficult, making successful cases rare. Creativity research began by trying to learn how to reproduce these cases. This required a standard of when an idea or artifact is "creative." Because exceptional cases are also different from what came before, the defining characteristic is most often novelty. However, not every novel creation is worth having created: it must also be useful, of value, or broadly, "appropriate." Thus, the most common definition has become "novel and appropriate."

Though appropriateness is necessary, people usually equate the extent of creativity with the extent of novelty. For instance, Runco and Smith (1992) operationalize creativity with a bipolar scale from "common" to "creative" (rather than "rare"). Similarly, Amabile (1996) goes to great length to argue that creativity is different from aesthetics or quality, but not from novelty. This leads to seeing creativity as being "as novel as possible, so long as appropriate." Working definitions of creativity probably play an active role in the creative process (Amabile 1996; Runco 2003). If creators ultimately want results that are *better*, and not simply different, it is important to ask whether "novel, so long as appropriate" is a suitable way to view creativity.

Drawing inspiration from the field of metaheuristic optimization (Blum & Roli 2003) and Campbell's (1960) "blind variation and selective retention" model of the creative process, this paper uses computer simulations to test how well different ways of understanding "novel and appropriate" impact final outcomes. These simulations show that the standard definitions of creativity are incomplete. They also demonstrate that psychological processes can plausibly produce the same dynamics as metaheuristic optimization algorithms, suggesting a new source of theoretical inspiration for creativity researchers.

Background

Broadly, creators are motivated to improve upon the state of the art in their field. Abstractly, this means maximizing some judgment of the solution's quality or value, moving their contribution beyond others.[1] They do this by manipulating the substrate of their domain, be it memetic or physical, hoping to find a better optimum somewhere in the set of all possible configurations.

In this way, creators are like computer optimization algorithms. Both search through the set of possible solutions to a problem with the goal of finding one that optimizes some standard of quality.[2] A perennial problem that both creators

*To appear in *Proceedings of the AAAI Spring Symposium on Creative Intelligent Systems*, Stanford, CA, March 2008

[1] Often, creative people value different kinds of solutions than their contemporaries. If others eventually adopt these values, their work will be recognized. Otherwise, it is likely to be forgotten. In either case, they are still maximizing with respect to their own idiosyncratic values.

[2] There are two common objections to modeling creativity with optimization. First, optimizers are algorithms, which some definitions of creativity exclude (e.g., Amabile 1996). However, "non-algorithmic" should be read as "no guaranteed recipe," not "unrelated to computing." Optimizers are non-deterministic and do not always work well, just as the most genius creator sometimes fails. The second objection is that an optimizer cannot transform the solution space to produce previously inconceivable solutions (Boden 1990). Though true, a person transforming a solution space does so according to some guiding standard, making that an optimization

and optimizers face is knowing when a suitable optimum has been attained. Often, solutions look optimal compared to similar alternatives, but are suboptimal compared to more radical departures. These "local optima" greatly complicate the optimization process. Metaheuristic optimization is a set of general algorithmic techniques for solving this problem. Metaheuristics marshal problem-specific strategies, but are themselves more generic, making them particularly suitable for models of human creativity.

Why Creativity is Hard Creative ideas are rare because tasks that require creativity are more difficult. Perkins (1994) analogizes creativity to finding gold in the Klondike, and notes that the rarity of "regions of payoff" creates a reluctance to change, and a lack of clarity about what direction to embark in when one does. Other authors have noticed this phenomenon in practice. Levitt and March (1988) refer to an organization's unwillingness to try new things as a "competency trap." Dunbar (1993) notes that some participants in a problem solving study waste all of their time trying to improve upon the wrong approach. Getzels and Csikszentmihalyi (1976) find that more creative artists don't solve problems better, but rather find better problems to solve. It is the act of finding the correct region and having the wisdom and courage to leave the wrong one that is difficult. These descriptions evoke the image of an optimizer trapped near a local optimum.

Another source of difficulty is that creative problems have many interdependencies, making progress difficult even in a promising region. Improving one aspect of a solution diminishes another. In engineering, these are called design conflicts (Altshuller 1984), and include things such as wanting something to be durable but light, or portable yet powerful. Even if it's obvious what the conflict is, it's seldom clear how to resolve it. Optimization problems are similarly complex.

Novelty for Novelty's Sake Creativity, via novelty and appropriateness, should help navigate these difficulties. Let appropriateness be equal to how good an idea is. Now suppose that a creative idea is exceptionally novel (dissimilar from other ideas), and consider three possibilities for how appropriate it is. Its appropriateness cannot be far below average, or else it wouldn't be called creative. If it is exceptionally appropriate, it *must* be highly novel; otherwise it would have been easy to find, and the average appropriateness would have adjusted before creativity was needed. This leaves the case where a creative idea is exceptionally novel, but only of average appropriateness. Despite people's reluctance to change and forces that keep existing technologies in place, such ideas are pursued. There must be an adaptive purpose of incurring this expense without appreciable gain. This purpose, it is proposed, is to help move beyond approaches that have exhausted their potential, i.e., local optima.

Without trying a new idea, it isn't possible to know whether it will lead to brilliant discoveries, or be fruitless. Using limited resources to pursue novel ideas is therefore risky, making valuing novelty for its own sake like investment. Just as a new employee should have a more aggressive retirement portfolio than a senior manager, a creator at the beginning of a project can seek significantly more novelty than one whose deadline is fast approaching.

Diversification and Intensification In optimization, this is known as the tradeoff between diversification and intensification (Blum & Roli 2003). Diversification means expanding the search beyond known regions. Intensification is when the search seeks the best solution in just one region. Just as someone can fall in between reckless risk-taking and complete risk aversion, there are gradations between pure diversification and pure intensification. Metaheuristic algorithms work by manipulating this balance.

One of the most popular metaheuristic algorithms is Simulated Annealing (Kirkpatrick, Gelatt, & Vecchi 1983). It works by analogy to the physical process of annealing (Metropolis *et al.* 1953), where a solid is melted, and then slowly cooled. At higher temperatures, chemical bonds are more likely to occasionally move to a less organized state. When cooled slowly enough, the material can settle in and out of different configurations and progress toward a highly organized steady state. If cooled too quickly, the reformed solid will be frozen in place, disorder and all. In *Simulated* Annealing, there is a "temperature" variable that reduces slowly, in so doing moving the balance from diversification toward intensification. At high temperatures, Simulated Annealers are more likely to abandon local optima, while at low temperatures they are more likely to simply improve upon what they've got. If they cool too quickly, they end up pursuing a substandard optimum.

Simulated Annealing has been used an analogy in negotiations (Klein *et al.* 2003) and organizational behavior (Carley & Svoboda 1996). This paper also uses a Simulated Annealing analogy. However, instead of recreating the algorithm exactly, it uses models of psychological processes that are plausibly involved in creation. Because these processes are high-level decisions that could be made by an individual or group, this approach is more general than past theorizing on creativity and Simulated Annealing in neural networks (Martindale 1995).

The primary hypothesis is that like temperature in Simulated Annealing, the extent that novelty is valued for its own sake controls a creator's balance between diversification and intensification. Accordingly, a creator's ability to find exceptionally good solutions should depend upon being able to properly change this balance over time.

Simulation Design

Computer simulations are done to represent the behavior of a creative individual or team working on a problem where "a creative approach" would be needed. Though the terms "problem" and "solution" are used, the model is abstract enough to correspond to both technical and aesthetic pursuits.

The basis of this model is Campbell's "blind variation and selective retention" theory (Campbell 1960; Simonton

problem. While this answer has the obvious potential for infinite regress, most other models also do when prodded.

2003). In this theory, creators make many variations to a starting idea, and then select which to retain. The variation process is "blind," implying that it is subject to random sources of inspiration and bounded rationality, not just reasoned decisions. The "selective retention" phase corrects for this as the creator (consciously or not) sorts through the variations to find those with the most promise.

In the simulation, every possible solution exists in a space, and the creator's current solution is a single point. Each point has an associated "goodness," which the creator seeks to maximize. At each stage of the simulation, the creator generates five (random) variations of the current solution. These and the current solution are each characterized by a rule. The solution rated best by the rule is selected, and then the process repeats itself from that point. The simulation terminates after a fixed period of time (iterations), or when the position stops changing. The simulation is judged on the goodness of its final point.

The simulator compares different decision rules, most corresponding to working definitions of creativity that might be used during the creative process. This allows the comparison of different ways that creativity might be defined, and hence clearer understanding of what "novel and appropriate" ought to mean.

Problem Domain Any problem domain with complex interdependencies and many local maxima suffices. Given their past use in the creativity and innovation literature (e.g., Frenken 2006), Kauffman NK landscapes (Kauffman 1993) were chosen. Originally developed to model genetic evolution, they have two principle parameters: N, the number of genes in a chromosome, and K, the degree of interdependency. Each gene has two possible alleles (forms), and so can be modeled as an N-bit binary string; variations are single bit flips. The fitness (here, goodness) of a genotype (realization of a chromosome) is the average of N fitness components, each dependent upon $K+1$ genes. Which genes a fitness component depends on isn't straightforward. In this simulation it is a random generalized NK map (Altenberg 1996) with $f = N$ and $p_i = K+1$. Fitness values for each possible configuration of $K+1$ genes are drawn from the uniform unit distribution. NK landscapes have many local optima, and finding the global optimum is an NP hard problem (Weinberger 1996).

Novelty and Appropriateness Novelty is the rarity of how a solution configures its constituent parts, judged relative to each prior solution that the creator has accepted. This is done by aggregating the pairwise dissimilarities between a point and each of the prior points. The two predominant theories of human dissimilarity judgment are based on either a distance metric (Shepard 1987), or shared and unshared features (Tversky 1977). The appropriate distance metric on an NK landscape is the Hamming distance, which is equal to the number of bits that differ, making it compatible with both models of human cognition. Additionally, it satisfies Jones' (1995) observation that the shape of a solution space depends solely on the transformations used to go from one point to another.

To accurately model human memory, recently visited and frequently revisited solutions should impact the result more (Higgins 1996). To achieve this, at each step the current position is put onto a queue, even if no change was made. Novelty is the exponentially decaying average (with a coefficient of .9) of the distances between each point on the queue and the current point.

Appropriateness is related to the goodness of the point. Some authors interpret appropriateness dichotomously (e.g., Runco 2003), while others see it as continuous (e.g., Amabile 1996). Thus, the simulator uses solution goodness, dichotomizing it when needed. How people arrive at this sense of goodness isn't clear, but theories do exist (Mangan 1991).

$Iterations \leftarrow 1$, $ConsecutiveRejects \leftarrow 0$
$T \leftarrow T_0$, $AcceptsSinceAdjustment \leftarrow 0$
$X \leftarrow \mathsf{RandomPoint}()$, $\mathbf{Q}_X \leftarrow \langle X \rangle$, $\mathbf{Q}_G \leftarrow \langle G(X) \rangle$
while $Iterations < 1000 \wedge ConsecutiveRejects < 120$ **do**
  $\mathbf{N} \leftarrow \mathsf{RandomNeighbors}(X, 5)$
  $\bar{G} \leftarrow \mathsf{DecayingAverage}(\mathbf{Q}_G)$
  $X' \leftarrow \operatorname{argmax}_{X'' \in X \cup \mathbf{N}} \mathsf{Evaluate}(G(X''), \bar{G}, \mathsf{DecayingDistance}(X'', \mathbf{Q}_X), T)$
  if $X' = X$ **then**
    $ConsecutiveRejects \leftarrow ConsecutiveRejects + 1$
  else
    $X \leftarrow X'$
    $ConsecutiveRejects \leftarrow 0$
    $AcceptsSinceAdjustment \leftarrow AcceptsSinceAdjustment + 1$
  end if
  $\mathbf{Q}_X \leftarrow \mathbf{Q}_X \cdot \langle X \rangle$, $\mathbf{Q}_G \leftarrow \mathbf{Q}_G \cdot \langle G(X) \rangle$
  if $AcceptsSinceAdjustment = 20$ **then**
    $T \leftarrow \alpha \cdot T$, $AcceptsSinceAdjustment \leftarrow 0$
  end if
  $Iterations \leftarrow Iterations + 1$
end while

Figure 1: The simulation algorithm. The Evaluate function, T_0, and α are all that varies between scenarios.

Simulation Algorithm Figure 1 shows the simulation algorithm. As can be seen, it is not specific to NK landscapes. Any domain can be used so long as there is a way to generate random points (RandomPoint) and transformations (RandomNeighbors), as well to measure distance (DecayingDistance) and goodness (G). From here, variations can be produced by changing the decision rule (Evaluate); its arguments are the goodness of the proposed point (G), the decaying average of prior points' goodness ($\bar{G}$), the decaying average of the distances between the proposed point and prior points ($\bar{D}$), and temperature (T). An Evaluate rule can use all or none of these values.

Like Simulated Annealing, the simulation includes a "temperature" parameter, T, that decreases as the simulation progresses. This is meant to model how human creators might adjust their willingness to make large changes as deadlines approach. The "cooling" rate will impact the quality of the final outcome. Cooling slowly means not rushing into limiting decisions too early. In contrast, a creator who cools too quickly will essentially be finding the best solution to the wrong approach. The simulation's cooling schedule is parameterized by the starting temperature (T_0) and a decay constant (α).

Readers familiar with metaheuristic optimization might have noticed that with both temperature and a queue of recently visited points, this algorithm resembles a hybrid of Simulated Annealing and Tabu Search (Glover & Laguna 1997). While true, it is worth bearing in mind that this algorithm is intended as a model of human behavior, not as an optimizer *per se*. Indeed, its heavy reliance on distance calculations makes it rather slow. However, what is hard for computers is often easy for humans, and vice versa. Herein lies the strength of this approach: if this simulation produces similar dynamics to a metaheuristic optimizer using models of psychological processes rather than typical optimization rules, then the computational theory can likely be adapted to human behavior.

Testing Approach This paper reports the results of three experiments. Experiment 1 tests simple combinations of novelty and appropriateness, while Experiment 2 includes temperature in the combination. Experiment 3 pits this temperature-varying combination against the "novel, so long as appropriate" rule. In addition to the algorithm in Figure 1, Experiments 1 and 3 use a hill climbing algorithm that considers the complete neighborhood of a point. This is included to measure the average goodness of the landscape's local maxima, but is not meant to model human capabilities.

Each simulation was run on the same five randomly-generated landscapes. An additional 10 landscapes were used in Experiment 3. Each simulation is run with 100 different randomly-generated points. Reported means and standard deviations are the mean of means and root mean square of standard deviations across landscapes. All statistical tests are run independently for each landscape.[3] Detailed results are available upon request.

Experiment 1

This experiment tests straightforward combinations of novelty and appropriateness. It first shows the effect of considering only novelty or only appropriateness. The novelty-only strategy is compared to a "random walk" strategy that chooses between alternatives randomly. The appropriateness-only strategy is compared to the hill climbing benchmark. It is hypothesized that novelty-only will be indistinguishable from a random walk, and that appropriateness-only will be indistinguishable from hill climbing. As such, the appropriateness-only strategy should outperform the novelty-only strategy.

Additionally, a rule that includes roughly equal parts novelty and appropriateness is tested. This rule is a first attempt at codifying the "novel and appropriate" criterion. For this to be an adequate definition of creativity, it should be able to outperform an appropriateness-only strategy, hence showing the superiority of a creative problem solving approach. However, it is not expected that this simple combination will suffice.

[3]Though statistical tests are less meaningful when sample size is essentially unlimited, finding both significant and insignificant results is illustrative.

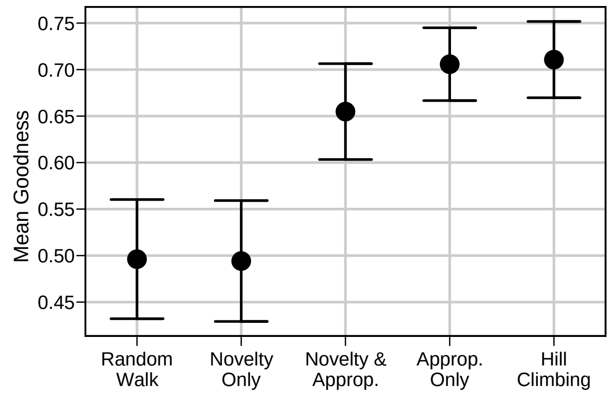

Figure 2: Comparison of approaches in Experiment 1. Error bars show one standard deviation.

Evaluation Rules

This experiment runs the simulator with the following Evaluate rules:

Random walk choose points at random

$$\mathsf{RandomWalk}(G, \bar{G}, \bar{D}, T) \leftarrow \mathsf{Random}()$$

Novelty only choose the point with the highest novelty

$$\mathsf{Novelty}(G, \bar{G}, \bar{D}, T) \leftarrow \bar{D}$$

Novelty and appropriateness choose the point with the highest sum of novelty and appropriateness. Novelty is first divided by 20, which makes the average distance between two points ($N/2 = 10$) equal to the average goodness of a point (0.5).

$$\mathsf{NoveltyApprop}(G, \bar{G}, \bar{D}, T) \leftarrow G + \bar{D}/20$$

Appropriateness only choose the point with the highest appropriateness

$$\mathsf{Approp}(G, \bar{G}, \bar{D}, T) \leftarrow G$$

Results and Analysis

The experiment was conducted on five landscapes with $N = 20$ and $K = 5$. The results are shown in Figure 2. As expected, the novelty-only rule is indistinguishable from the random walk rule—the difference is not statistically significant for any of the landscapes. Additionally, there appears to be no difference between the appropriateness-only rule and hill climbing. Indeed, the difference is not significant for four of the landscapes, though for one it is ($p < .01$). The novelty and appropriateness rule lies in between the novelty-only and appropriateness-only rules (all ps $< .00001$).

These results confirm the hypotheses. The results for the NoveltyApprop rule can be seen from two perspectives. On

M_G (SD_G)	$T_0 = 0.05$	$T_0 = 0.10$	$T_0 = 0.20$	$T_0 = 0.40$
$\alpha = .80$	0.7590 (0.0239)	0.7575 (0.0262)	0.7601 (0.0242)	0.7591 (0.0241)
$\alpha = .85$	0.7626 (0.0241)	0.7620 (0.0242)	0.7623 (0.0229)	0.7628 (0.0243)
$\alpha = .90$	0.7645^1 (0.0234_2)	0.7649^3 (0.0226_2)	0.7626 (0.0232_1)	0.7590 (0.0263)
$\alpha = .95$	0.7588^1 (0.0276)	0.7361 (0.0354)	0.7015 (0.0410)	0.6697_5 (0.0489^5)

Table 1: Mean (standard deviation) goodness of the final points with varying values for initial temperature, T_0, and the cooling coefficient, α, averaged across five landscapes. Superscripts/subscripts indicate number of landscapes where indicated quantity was maximum/minimum.

one hand, they show that considering novelty and appropriateness is a great improvement over considering novelty alone. On the other hand, they show that considering novelty and appropriateness is worse than considering only appropriateness. As such, novelty seems to be unhelpful. If creativity is to be considered beneficial, this rules out a definition where "novel and appropriate" are equal partners.

Each search is judged by the goodness of its final point. This models situations where it is not possible to "set aside" one's best work in case future changes prove disastrous. In fact, much of the finesse in creativity comes in recognizing when you have encountered something noteworthy, even if by accident (Kantorovich & Ne'emant 1989). This translates to knowing when to move away from a local maximum in search of higher peaks, and when to stay put. One possibility is that including novelty was destructive because the NoveltyApprop rule made no provision for this.

To test this, the rules were compared on the best point they encountered in each run. For the Novelty, NoveltyApprop, and Approp rules, these averages were 0.685, 0.777, and 0.706, respectively. Thus, it is clear that considering both novelty and appropriateness directs the search toward more fruitful regions: it simply fails to stay put when need be.

Experiment 2

This experiment corrects the shortcomings of a simple novelty and appropriateness combination by using the "temperature" variable to weight novelty. While temperature's initial value can favor novelty, toward the end it should highly favor appropriateness. In this way, the search can settle upon a promising region of the search space, rather than always abandoning good solutions for something that is simply newer.

The cooling schedule must be appropriate for temperature to function well. In fact, in Simulated Annealing the cooling schedule is the biggest determinant of performance (Romeo & Sangiovanni-Vincentelli 1991). If temperature starts too high and declines too slowly, the search never converges. If the temperature starts too low and declines too quickly, the search degenerates into hill climbing. The same dynamics are expected to apply for this simulation.

Parameterization

This experiment tests the following Evaluate rule with several different temperature parameterizations:

$$\mathsf{Temperature}(G, \bar{G}, \bar{D}, T) \leftarrow G + T \cdot \bar{D}$$

The temperature adjustment schedule is parameterized by T_0, the initial temperature, and α the rate at which temperature is reduced (which happens after 20 variations are accepted at one level). Each combination of $T_0 \in \{.05, .10, .20, .40\}$ and $\alpha \in \{.80, .85, .90, .95\}$ is tested, using the same five landscapes from Experiment 1. Note that $T_0 = .05$ is the same value used to weight novelty in the NoveltyApprop rule. It is predicted that at $T_0 = .05$, $\alpha = .80$, the simulation will behave similarly to Approp, while at $T_0 = .40$, $\alpha = .95$ it will behave similarly to Novelty. Somewhere in between there should be a discernable ideal combination.

Results and Analysis

As seen in Table 1, both initial temperature and cooling rate affect the average outcome. The mean across all five landscapes is highest at $T_0 = .10$ and $\alpha = .90$. Additionally, this is the best combination for three of the five landscapes when considered separately. This combination is also the most consistent: its standard deviation across all five landscapes is the lowest, and it produces the lowest standard deviation in two landscapes when considered separately (no other combination did better). However, adjacent cells with lower cooling coefficients and higher or lower initial temperatures perform similarly, showing that parameter choice does not affect performance very much in this range.

It was predicted that at $T_0 = .05$, $\alpha = .80$, the simulation would behave similarly to Approp. This was not borne out: on all five landscapes, the Temperature rule outperformed the Approp rule (all ps ≈ 0). However, performance does degrade with lower initial temperatures and cooling coefficients. The differences between this cell and the $T_0 = .10$, $\alpha = .90$ cell are significant for three of the five landscapes. Thus, it seems warranted to conclude that results would approach the Approp rule, possibly well before reaching trivially small parameter values.

The worst outcomes, judged both by aggregate and separate results, occur when $T_0 = 0.40$ and $\alpha = .95$. This parameterization was predicted to behave similarly to Novelty. Again, this was not borne out: the rule still outperforms Novelty (all ps ≈ 0). However, the results are quite close to the NoveltyApprop rule ($p < .05$ for three landscapes, $p > .15$ for the other two). At this parameterization, the greater complexity of the Temperature rule is not clearly worthwhile.

In summary, although the range of temperature parameters explored was too narrow to elicit extreme behavior, outcome quality is clearly subject to parameter choice. Indeed, at the highest end of the range, the rule was barely distinguishable from a much simpler (i.e., not time-dependent) approach. For the bulk of the range, however, this rule outperformed any of the others tested in Experiment 1. This is particularly interesting for $T_0 = 0.05$, as this corresponds to the weighting used in the NoveltyApprop rule. Initially including novelty, but gradually reducing its influence, eliminates its destabilizing effect. Varying one's preference for novelty over time appears to be essential.

Experiment 3

Experiment 2's approach is vulnerable since the proper rate at which to discount novelty cannot be known ahead of time. Furthermore, this approach is only indirectly sensitive to how promising the current region is. It would be preferable to have a definition that (a) adjusts the novelty and appropriateness balance in direct response to current conditions, and (b) is less dependent upon parameters that must be chosen in advance.

As noted earlier, the "appropriateness" condition has sometimes been interpreted as all-or-nothing, i.e., "novel, so long as appropriate." This experiment tests a dichotomous definition that uses the decaying average of goodness as a cutoff. That is, the average goodness of recent points is calculated, but points visited more recently carry more weight than those visited a long time ago. If a new point's goodness falls below this average, it is not considered.

This approach is a simple and psychologically plausible way to be sensitive to local conditions. First, it admits downhill moves since the most recent point's influence is counterbalanced by the points before. However, the average rapidly adjusts to improving conditions, thereby reducing the likelihood of downhill moves. It is similarly rapid to adjust to deteriorating conditions, helping the search find a way out of a valley, even if through variations that represent only slight improvements. This promotes transition to new maxima. This approach also smoothes out extremes; if a region has just a few very good or very bad points, they will be balanced out by more typical points. This helps the search find consistently good regions before beginning a pure upward climb. Finally, the search converges naturally when continued rejections stabilize the average at the goodness of the current point.

This approach's biggest strength is that it does not require any explicit adjustment with time. All adjustment is handled by the decaying average. Inasmuch as people's evaluations are much more strongly influenced by recent experience (Fredrickson & Kahneman 1993), this is a plausible model of how people might make dichotomous appropriateness judgments.

Evaluation Rules and Parameterization

This simulation compares the performance of the Temperature rule against the following rule:

$$\mathsf{Threshold}(G, \bar{G}, \bar{D}, T) \leftarrow \textbf{if } G < \bar{G} \textbf{ then } 0 \textbf{ else } \bar{D}$$

Temperature adjustment is parameterized with the best-performing combination from Table 1. To make an even comparison, the Threshold rule was also tuned. The only parameter affecting it is λ_G, the rate of the decaying average of goodness. Simulations were run across all five landscapes for $\lambda_G \in \{.80, .85, .90, .95\}$. In four of the five landscapes, an omnibus ANOVA found no significant differences. In the one landscape with significant differences, $\lambda_G = .90$ was the optimal choice (though it was not statistically different from $\lambda_G = .95$). Because $\lambda_G = .90$ maximized the mean across landscapes, and because it is the same as the distance decay rate, it is used in these comparisons. Still, these results show that the Threshold rule is practically insensitive to parameter choice.

The rules are compared against each other and against hill climbing. In addition to using the five landscapes with $N = 20$ and $K = 5$ from the prior experiments, five landscapes each were created at $N = 20, K = 10$ and $N = 20, K = 15$. The same rule parameters were used with these landscapes. (The results for the Temperature rule with $N = 20, K = 5$ are taken from Experiment 2.)

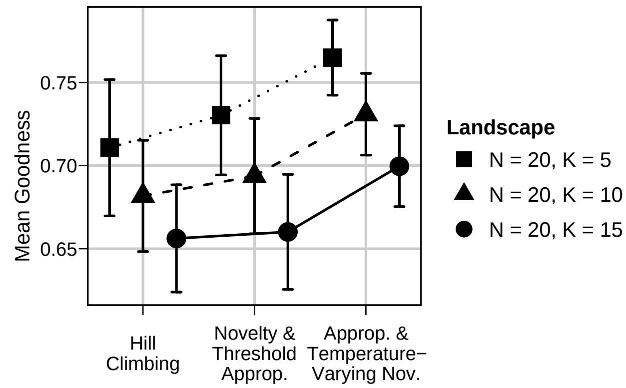

Figure 3: Mean goodness obtained in Experiment 3 across landscapes of increasing complexity. Error bars show one standard deviation.

Results and Analysis

Results are shown in Figure 3. As can be seen, the Threshold rule outperforms hill climbing in simpler landscapes. This difference was significant for all landscapes with $K = 5$, and for four of the five landscapes with $K = 10$. However, for $K = 15$ it only outperformed hill climbing on one landscape. The Temperature rule, on the other hand, always handily outperformed both hill climbing and the Threshold rule.

Each rule's performance drops as the landscape gets more complex. However, it appears that the Threshold rule suffers a greater drop than the Temperature rule. Indeed, with $K = 5$, the Threshold rule exceeds hill-climbing by 2.74%, while Temperature exceeds it by 7.63%, making the Temperature rule 178% better. For $K = 10$, these values are 1.76% and 7.21%, making the Temperature rule 311% better. Finally, for $K = 15$, these values are 0.603% and 6.62%, making the Temperature rule almost 1000% better.

This indeed confirms that while both rules drop in performance, the Threshold rule is much more fragile.

Altogether, while these results do show that a dichotomous definition of appropriateness can outperform hill-climbing, they also show that deliberately reducing the role of novelty over time produces vastly superior results, particularly as the problem grows more complex. This suggests that the "novel, so long as appropriate" definition of creativity has limited value in the creative process.

Discussion

If the goal of creativity research is to help reproduce the conditions of creative success, then there must be a suitable definition of creativity. This research has shown that if "novel and appropriate" is the definition guiding creators during the creative process, then appropriateness must be continuous and novelty's influence must decrease with time. For this reason, "novel and appropriate" might better be "appropriately novel, and appropriate." Indeed, the key to creative success might not be originality as much as it is knowing how to adjust this "novelty thermostat."

This research has also shown that Campbell's blind variation and selective retention theory is a plausible model of the creative process. This adds to the mounting evidence (e.g., Simonton 2003) of the important role that chance plays in creative success. However, this must not be misinterpreted as suggesting that randomness is important and reason is not. Instead, randomness must be used judiciously, and it must become less important as time progresses.

Finally, this research has shown that Simulated Annealing dynamics can be reproduced using models of psychological processes. This adds to cases where Simulated Annealing has been a productive model of social phenomena (Carley & Svoboda 1996; Klein *et al.* 2003). Creativity research has often proceeded from observation to hypothesis, i.e., in a bottom-up manner (Jennings 2007), which must be balanced out by top-down theorizing. Metaheuristic optimization, as well as a broader complex systems perspective, should be useful bases.

Adjusting the Novelty Thermostat This research has shown that adjusting the novelty thermostat is essential to achieving better outcomes. However, only monotonic, geometrically-decreasing cooling schedules were tested. More sophisticated approaches exist (e.g., Ingber 1996), and could be even better models of human creativity.

This simulation assumed that creativity is the operative judgment throughout an idea's gestation. One could argue that creativity is only needed to find an interesting problem, after which normal problem solving suffices. However, these simulations could easily represent problem finding alone, and thus still show that "novel, so long as appropriate" is a substandard definition. Furthermore, methodologies with enforced boundaries (such as IDEO's) are probably better modeled with a steep drop in temperature than with a switch to pure hill climbing.

Different creators have different cooling schedules, which could make one creator more successful than otherwise equivalent competitors. However, the best cooling schedule is obvious only in retrospect. Thus, success does not necessarily imply that this person "knew" the right schedule to use, illustrating that success is not sufficient evidence of creative skill (Jennings 2007).

Blindness and Randomness In this model, "blind" variation was modeled as a throw of the dice. Though it did not explicitly test more deliberate ways of producing variations, it did illustrate that simple randomness yields better results than exhaustively considering all options at each stage (hill climbing). However, there are more intelligent heuristics than pure expediency, and so it is worth wondering whether producing variations less randomly would improve performance. This would be essential with realistically sized neighborhoods (rather than ones with a mere 20 possibilities, as in these simulations).

Still, more deliberate sampling might have its limits. Expertise hurts creativity if it prevents the exploration of promising new paths that happen to violate sacred but outdated assumptions (Frensch & Sternberg 1989). This can happen at the evaluation stage (via biased appropriateness judgments), and it can also happen at the variation stage (via a biased sample of the neighborhood). Having spent their entire careers optimizing in one part of the search space, experts might be unlikely to generate variations that will move toward better optima. An advantage relative novices have is that their "blinder" variations sometimes lead in useful directions. However, they are also more likely to produce useless variations, which is one of the many reasons that there are more Ph.D. students than full professors.

Selection Pressures Misunderstandings about novelty's role probably stem from imprecision about whether the set of ideas has been subject to selection pressures (e.g., peer criticism, funding decisions, market response). Prior to selection pressures, novel ideas that are lacking in appropriateness might still survive. Here, the ideas ultimately identified for their creativity will be novel, but more importantly they will be exceedingly appropriate. After selection pressures, novel ideas lacking in appropriateness will have been eliminated. The ideas seen as creative will still be exceedingly appropriate, but unlike before they will also be exceedingly novel. Only in such cases does "novel, so long as appropriate" suffice.

This illustrates why it is important not to take post-selection insights and translate them unquestioningly into pre-selection situations. When we see extraordinary results, we have a natural desire to assume that they had extraordinary antecedents (Campbell 1960; Schaffer 1994). However, this needn't be the case. Accordingly, highly novel results do not imply highly novel origins, at least at all stages of the creative process. Calls to increase creative success via unquestioning openness and fetishized originality should be regarded with skepticism, which top-down theorizing makes clear.

This work has illustrated that rather than viewing novelty as the defining characteristic of creativity, it should be viewed as an enabler of creative progress that must be properly managed. This suggests that creative skill lies not just in doing something novel, but also in knowing how much novelty to pursue at what time. Clearer understanding of these issues will depend on pursuing creativity research with equal attention to high-level theory and ground-level behavioral evidence. Metaheuristic optimization suggests a route for formalizing ideas such as blind variation and selective retention, ultimately leading the way toward increased scientific understanding of creativity. Only this will make it possible to more reliably reproduce the successes of those who have advanced the status quo.

Acknowledgments

The author is supported by a National Science Foundation Graduate Research Fellowship. Tom Griffths, Jono Hey, Caneel Joyce, and Rachel Beth Egenhoefer provided valuable help with this paper and related work. Finally, I wish to thank Don Thomas for first introducing me to simulated annealing, and for first suggesting that I pursue a Ph.D.

References

Altenberg, L. 1996. NK fitness landscapes. In Back, T.; Fogel, D.; and Michalewicz, Z., eds., *The Handbook of Evolutionary Computation*. Institue of Physics Press.

Altshuller, G. S. 1984. *Creativity as an Exact Science: The Theory of the Solution of Inventive Problems*. Gordon and Breach Publishers.

Amabile, T. 1996. *Creativity in Context*. Westview Press.

Blum, C., and Roli, A. 2003. Metaheuristics in combinatorial optimization: Overview and conceptual comparison. *ACM Computing Surveys* 35(3):268–308.

Boden, M. A. 1990. *The Creative Mind: Myths and Mechanisms*. BasicBooks.

Campbell, D. T. 1960. Blind variation and selective retention in creative thought as in other knowledge processes. *Psychological Review* 67:380–400.

Carley, K. M., and Svoboda, D. M. 1996. Modeling organizational adaptation as a simulated annealing process. *Sociological Methods and Reseach* 25(1):138–168.

Dunbar, K. 1993. Concept discovery in a scientific domain. *Cognitive Science* 17:397–434.

Fredrickson, B. L., and Kahneman, D. 1993. Duration neglect in retrospective evaluations of affective episodes. *Journal of Personality and Social Psychology* 65(1):45–55.

Frenken, K. 2006. *Innovation, Evolution, and Complexity Theory*. Edward Elgar.

Frensch, P. A., and Sternberg, R. J. 1989. Expertise and intelligent thinking: When is it worse to know better? In *Advances in the psychology of human intelligence*, volume 5. Lawrence Erlbaum Associates. 157–188.

Getzels, J. W., and Csikszentmihalyi, M. 1976. *The Creative Vision: A Longitudinal Study of Problem Finding in Art*. John Wiley & Sons.

Glover, F., and Laguna, M. 1997. *Tabu Search*. Kluwer Academic Publishers.

Higgins, E. T. 1996. Knowledge activation: Accessibility, applicability, and salience. In Kruglanski, A. W., and Higgins, E. T., eds., *Social psychology: Handbook of basic principles*. The Guilford Press.

Ingber, L. 1996. Adaptive simulated annealing (ASA): Lessons learned. *Control and Cybernetics* 25(1):33–54.

Jennings, K. E. 2007. Statistical errors in creativity research: Survivorship bias and alpha inflation in attributing creativity and its causes. In *Proceedings of the 10th European Conference on Creativity and Innovation*. [URL: http://www.eccix.org/pdf/59_Kyle_E_Jennings.pdf].

Jones, T. 1995. One operator, one landscape. Working Papers 95-02-025, Santa Fe Institute.

Kantorovich, A., and Ne'emant, Y. 1989. Seredipity as a source of evolutionary progress in science. *Studies in History and Philosophy of Science* 20(4):505–529.

Kauffman, S. A. 1993. *The Origins of Order: Self-Organization and Selection in Evolution*. Oxford University Press.

Kirkpatrick, S.; Gelatt, C. D.; and Vecchi, M. P. 1983. Optimization by simulated annealing. *Science* 220(4598):671–680.

Klein, M.; Faratin, P.; Sayama, H.; and Bar-Yam, Y. 2003. Negotiating complex contracts. *Group Decision and Negotiation* 12:111–125.

Levitt, B., and March, J. G. 1988. Organizational learning. *Annual Review of Sociology* 14:319–338.

Mangan, B. 1991. *Meaning and the Structure of Consciousness: An Essay in Psycho-aesthetics*. Ph.D. Dissertation, University of California, Berkeley.

Martindale, C. 1995. Creativity and connectionism. In Smith, S. M.; Ward, T. B.; and Finke, R. A., eds., *The Creative Cognition Approach*. MIT Press.

Metropolis, N.; Rosenbluth, A.; Rosenbluth, M.; Teller, A.; and Teller, E. 1953. Equations of state calculations by fast computing machines. *Journal of Chemical Physics* 21(6):1087–1092.

Perkins, D. N. 1994. Creativity: Beyond the Darwinian paradigm. In Boden, M. A., ed., *Dimensions of Creativity*. MIT Press.

Romeo, F., and Sangiovanni-Vincentelli, A. 1991. A theoretical framework for simulated annealing. *Algorithmica* 6:302–345.

Runco, M. A., and Smith, W. R. 1992. Interpersonal and intrapersonal evaluations of creative ideas. *Personality and Individual Differences* 13(3):295–302.

Runco, M. A. 2003. Idea evaluation, divergent thinking, and creativity. In Runco, M. A., ed., *Critical Creative Processes*. Hampton Press.

Schaffer, S. 1994. Making up discovery. In Boden, M. A., ed., *Dimensions of Creativity*. MIT Press.

Shepard, R. N. 1987. Toward a universal law of generalization for psychological science. *Science* 237(4820):1317–1323.

Simonton, D. K. 2003. Scientific creativity as constrained stochastic behavior: The integration of the product, person, and process perspectives. *Psychological Bulletin* 129(4):475–494.

Tversky, A. 1977. Features of similarity. *Psychological Review* 84(4):327–352.

Weinberger, E. D. 1996. NP completeness of Kauffman's N-k model, a tuneable rugged fitness landscape. Technical Report 96-02-003, Santa Fe Institute.