CHAPTER FIFTEEN

EVOLUTIONARY ROBOTICS AND CREATIVES CONSTRAINTS

DUSTIN R. STOKES[1], JON BIRD[2]
Centre for Research in Cognitive Science, University of Sussex (UK)[1],
Centre for computational Neuroscience and Robotics, University of Sussex (UK)[2]

Analyzing creativity is no easy task. The difficulty of the task is often compounded by a problematic analytical strategy: start with paradigm instances of artistic genius and work down. Using the basic techniques of evolutionary robotics, a team of researchers based primarily at the University of Sussex[1] are attempting to artificially evolve robots with the capacity to draw, or at least make marks in a way that resembles drawing. One project goal is purely theoretical, namely, using artificial life and evolutionary robotics to study creative behaviour. This implies a more immediate goal, which is not to evolve art or artists, but to evolve minimally creative agents. We thus take the traditional tack on creativity and stand it on its head: rather than starting with high level creativity or genius and then working down, we start with a minimal notion of creativity and then work up. We are interested in building the simplest creative systems possible, that is, systems with little complexity that act in ways that at least satisfy certain conditions of creativity.

This paper highlights some of the preliminary stages of research, clarifying the underlying conceptual assumptions and the experimental methodology. The project is still quite young, and so conclusions are few.

---

[1] The project Creativity, Cognition, Computational Intelligence, and Aesthetics: A Multidisciplinary Investigation, is funded by the AHRC. The research team comprises cognitive scientists, artificial life researchers, philosophers, roboticists, and artists. The goals are many: develop an autonomous artwork (Bird et. al, forthcoming); study creative behaviour (Bird and Stokes, 2006; forthcoming); study evaluation (Bird and Stokes, forthcoming); study issues of agency and autonomy (Boden, forthcoming); and study computer and interactive art.

Nevertheless, early results already shed light on various problematic issues that come with analyzing or modelling creativity.

One such issue, discussed at length in the final section, is as follows. A popular folk conception of creativity is that creative thinking or acting requires a constraint-free set of circumstances. So with an increase in constraints on thought and action comes a decrease in the potential for creativity. In its strongest form, the claim is that one cannot act creatively if one is working under constraints.

With respect to artistic creativity, Kant seems to endorse this general claim. He writes,

> When the imagination is used for cognition, then it is under the constraint of the understanding and is subject to the restriction of adequacy to the understanding's concept. But when the aim is aesthetic, then the imagination is free, so that, over and above that harmony with the concept, it may supply, in an unstudied way, a wealth of undeveloped material for the understanding which the latter disregarded in its concept (Kant 1987: 185).

This reasoning derives, for Kant, from his view that rules may not be invoked to make aesthetic judgments. In creating art, the artist thus must avoid the constraints of rule-bound thought, opting instead for the free play of the imagination.[2]

This general conception is echoed across much of traditional literature on creativity, most especially in romantic thought. Creative persons, the thought goes, work best when they are free from constraints of various sorts: social, conceptual, intellectual. This understanding of creativity is further encouraged by famous reports of creative geniuses, where a flash of insight comes during a contemplative moment of free imagery.[3]

---

[2] See Kant 1987: §46-50. For criticism of this position of Kant's, see Cohen 2003.

[3] Examples from the chemist Friedrich von Kekulé are perhaps most famous. With respect to his insight on a new model of molecular structure, he writes:

> One fine summer evening, I was returning by the last omnibus… 'outside' as usual, through the deserted streets of the metropolis, which are at other times so full of life. I fell into a reverie, and lo! The atoms were gambolling before my eyes. Whenever, hitherto, these diminutive beings had appeared to me, they had always been in motion; but up to that time, I had never been able to discern the nature of their motion. Now, however, I saw how, frequently, two smaller atoms united to form a pair; how a larger one embraced two smaller ones; how still larger ones kept hold of three or even four of the smaller; whilst the whole kept whirling in a giddy dance. I

For simplicity, call this the *romantic thesis*.[4] Broadly, the thesis is that a person (or more basically, an agent or system) may act creatively only if the person is free from constraints. This is a thin thesis indeed: many questions remain. Constraints of what sort? Must a person be entirely constraint-free? If not, how much or how many constraint(s) is too much? Answers to these questions underwrite romantic creativity theses of differing strengths.

One can, just from the armchair, think of examples from the history of art and science that counter romantic theses of varying strengths. One can also think of principled reasons—after reflection upon concepts of creative behaviour—that the romantic thesis is conceptually flawed.[5] Alternatively, one might take the romantic thesis into the laboratory. Is there an empirical method for dispelling the romantic thesis? Indeed there is. And as it turns out, it does not much matter how one fills out the thesis. Empirical considerations suffice to show the falsity of the romantic thesis, whatever its strength. We must first provide some conceptual basis for the empirical studies.

1 - Minimal Creativity

To what, at minimum, do we attribute creativity? Linguistic intuitions are inconclusive on the matter, but they do suggest two necessary conditions. If there is one obvious condition on creativity, it is *novelty*. But that doesn't take us far. It only invites another challenge, namely, understanding novelty. The first question is *how* new novel properties need be to be novel (or at least to be novelty of interest). We might think that novelty should be understood as literally as possible: some *F* is novel if and only if *F* has never occurred before. This is novelty *simpliciter*. In a related spirit, we might think of novelty in terms of what Margaret Boden

---

saw how the larger ones formed a chain…(Findlay 1965: 38-9). See also Boden 2004: 25-28.

[4] Two notes. It is really a class of theses, and in Section 5, different strengths of the thesis are considered. Second, there are many theories of creativity which could suitably be called 'romantic', some which incorporate this point about constraints plus some other features, and some that are different entirely. The choice of the term "romantic" is thus just one of convenience and not meant to denote all the possible theories that answer to this term. For examples, see Bergson 1998, 1992; Collingwood 1938; Dewey 1958; Schelling 1936, among others. More recently, see Sawyer 2000. For criticism, see Boden 2004; Koestler 1964.

[5] See, for example, Boden 1995, 2004; P. Stokes 2005

calls *historical novelty*, an *F* is historically novel if and only if it is new to the history of ideas (Boden 2004). Boden contrasts this with *psychological novelty*, which is relative to some particular mind. Here the idea is that in spite of the historical fact that a thought has been tokened or a discovery made or an artefact created before, there are circumstances where a second tokening of the thought or act, would reasonably count as novel in some interesting way.[6] Psychological novelty, as contrasted with the other types mentioned, is a *relative* novelty; in this case the comparison class is a psychological profile or mind.[7]

Historical novelty is no doubt an interesting issue, but it isn't the obvious business of cognitive science. What makes something historically novel and not merely psychologically novel are socio-historical facts, not cognitive or behavioural ones. And so an interest in cognition and behaviour of creative processes is really an interest in some kind of relative novelty. This choice for relative novelty implies another decision point: what is the appropriate comparison class for novelty? A research project attempting to model creative processes, using a bottom-up, synthetic approach, does well not to prejudice the issue towards psychological or cognitive agents from the start. Boden's psychological novelty can thus be extended beyond the cognitive to the behavioural. Behavioural novelty may be understood in, at least, two ways.

Analogous to Boden's understanding of psychological novelty, a behaviour may be novel relative to the behavioural history of the acting agent. Thus a behaviour Φ is *individual-relative novel* for some agent *A* if and only if Φ is novel relative to the behavioural history of *A*. More broadly, some behaviour Φ is *population-relative novel* for some agent *A* if and only if Φ is novel relative to a population of which *A* is a member. There is some flexibility in both definitions. With the first, one must determine how individuals are individuated. With the second, one must

[6] For example, imagine two research teams, A and B, working—independently of one another—on a cure for some disease. If team A discovers the cure a day before team B, but team B discovers the cure entirely independently of team A, there is an obvious sense in which team B has done something novel, and in spite of the fact that it is not historically novel. One can think of countless similar examples in theoretical, scientific, and artistic domains.

[7] Of course, one may think of historical novelty as relative novelty of the broadest scope. That is, an F is historically novel relative to all of history (rather than some subset of history, e.g. the behaviour of some agent, group, or population). Our interest is nonetheless in a relative novelty where the comparison class is more restricted.

determine how the relevant populations are individuated. And with both senses, one needs to determine a level of description for the behaviours.

Population relative novelty might just be a form of historical novelty, where the behavioural history relative to which a behaviour is novel is just that of the specified population. This reveals just how flexible the notion of population relative novelty is: depending upon the population of comparison, one gets very different answers to the question whether something is novel. Thus at least some features of historical novelty are the business of cognitive science, namely the behavioural and cognitive ones. So from the perspective of cognitive science—no matter the scope of the comparison class (e.g. the behavioural histories of individuals or entire populations)—an interest in novelty is an interest in the behavioural and possibly cognitive changes necessary for the agent in question to act in ways novel relative to that comparison class. Other features, the purely historical and sociological ones, are better analyzed by history and sociology, among other disciplines. It is not the job of the cognitive scientist to explain how cubism was startling to the artworld or why the conceptual shift to quantum mechanics was groundbreaking.

Intuitions secure us a second condition for creativity. Creativity is not properly attributed to events or artefacts that result from accidents. We attribute creativity only to things that result, in some non-trivial sense, from *agency*. Some *F* is creative only if *F* is the product of agency. Agency can be understood in radically different ways. Philosophers, for example, analyze what constitutes agency and tend to focus on intentional (and thus cognitive) states and deliberate actions as requisite for agency. Researchers in artificial life tend to focus on the origins of agency, and less on deliberation, reasoning, or intentionality. These differences in approach can result in confusing differences in concept use. However, the second, bottom-up approach can be understood as continuous with the first, that is, as an alternative way to clarify the constituents of agency, or at least its evolutionary and developmental foundations.

Agency, at least for a start, may be understood broadly. A system is an agent if that system is self-moving, that is, not entirely controlled by an *external* system or programmer. This requires neither self-generation nor cognition or deliberation. Thus many of the simulated robots found in artificial life research qualify while remote controlled robots do not. Some behaviour, artefact, or event *F* is the product of the agency of A only if *F* would not have occurred had A not acted in some autonomous way.

Here then are the two necessary conditions identified:

- *Agency*: *x* is creative only if *x* is the product of agency;

- *Novelty*: *x* is creative only if *x* is behaviourally novel

The agency condition depends upon a broad, *no strings attached* notion of autonomy and a counterfactual dependence between the behaviour and that autonomy. The novelty condition separates into two: individual-relative and population-relative. One may specify other comparison classes for behavioural novelty. Is this enough? Not likely. Necessity aside, satisfaction of these two conditions does not look sufficient for creativity. I can right now, token the following thought "Worms are not the only wormy thing in this one-pony town." This is surely a thought that depends upon my agency in the right way and is a thought which is novel relative to me. In fact, it is likely historically novel. But it is not obviously creative. So a third condition (or more) is needed for creativity. What additional conditions should be appended to the above two is unclear, and largely because intuitions on creativity after agency and novelty vary and in ways according to the context of inquiry. Call it an open question, then, what additional necessary conditions are conjointly sufficient for creativity (in some context or other), assuming that the concept is even cleanly definable. In any case, the two identified conditions are necessary and must be satisfied irrespective of any additional conditions. Thus to model creative behaviour, one must model something that, at minimum, involves the right kind of agency and novelty.

The choice for the broad construal of both the agency and novelty conditions is motivated by our underlying research project. A central theoretical goal of the project is to see what can be learned about creativity and cognition using synthetic, bottom-up modelling techniques. The immediate task is thus to evolve the most minimal systems that satisfy these minimal conditions. The supposition that autonomous agents can be produced using a bottom-up, synthetic approach, and examined for individual and population-relative novelty, allows for fruitful methods of modelling and opportunities for hypothesis generation. Even if the project fails to evolve agents that meet more demanding conditions of creativity, the mere attempt should illuminate features of the explanandum. This is a basic methodological assumption shared with much of cognitive science.

2 - Evolutionary Robotics

Evolutionary robotics (ER) is a biologically-inspired approach to creating autonomous robots (Husbands et al. 1997; Nolfi and Floreano 2000). Using evolutionary search algorithms (Holland 1975), different aspects of a robot—the software control system and sometimes hardware properties such as body shape and sensor and actuator characteristics design—are evolved.

Typically, a population of agents is tested for their ability to perform some desired behaviour and the fittest individuals tend to get selected to produce the next generation of robots. This process continues until either the robots perform at a satisfactory level or an experiment has been carried out for a large number of generations (typically thousands). Agents may be individuated by their *genotype*, the data structure that encodes the controller and other properties of the robot design, or their *phenotype*, the software and hardware instantiation of the genotype. In the initial population the genotypes (generally a string of numbers) are usually randomly generated. An experimenter defines a *fitness function* for automatically measuring the fitness of each agent. The genotype of each agent in turn is instantiated as a robot, placed in a test environment, and its fitness tested. The testing process is carried out in simulation, in the real world or in a combination of both. For example, as a robot moves around an arena it might gain fitness for avoiding obstacles or collecting objects or performing phototaxis.

In the initial generations the performance is generally very poor, which is not surprising as the designs result from random genotypes. However, some individuals will be fitter than others and these tend to get selected to form the next generation of agents. This new population is created by applying operations to the genotypes of selected agents. Mutation involves randomly changing some of the numbers in the genotype. Crossover, inspired by sexual reproduction, consists of combining parts of two individual's genotypes to create a new genotype different from the two 'parents'.

ER has been applied successfully to the creation of a wide range of simulated and real autonomous robots that demonstrate increasingly complex behaviours. Examples include discriminating and moving objects around to achieve multi-stage tasks such as garbage collection (Nolfi and Floreano 2000); visually navigating through complex environments (Floreano et al. 2007); and competitive (Floreano and Nolfi 1997) and co-operative group behaviour (Quinn et al. 2003).

One of the advantages of ER pertinent to our project is that it facilitates experimenters overcoming their *inductive bias* (Pollock et al. 2001). For example, an experimenter might have strong assumptions about how a robot should interact with its environment and the type of software and hardware required to successfully perform a desired behaviour. Whether an experimenter's prejudices are explicit or implicit, they limit exploration of the possible space of designs, potentially biasing experiments in unproductive ways. ER, however, can potentially exploit any constraints arising from the interaction between components in the robot and between the robot and the environment, even ones of which an experimenter may be unaware. Keeping conceptual assumptions to a minimum can sometimes lead to simpler and more robust robots than would be generated by conventional design (Pfeifer and Scheier, 1999). Using an ER approach to model creativity can thus potentially reduce the influence of various preconceptions, theories and case studies of creativity.

3 - Evolving a Drawing Agent: Experiments and Methodology

Drawing is an embodied activity. Thus, our project will ultimately carry out experiments on a physically embodied robot that makes marks in response to feedback from its environment, including the marks that it has left previously. This choice is motivated by a number of factors. Marks on a page provide a visual record of the robot's movements; embodiment potentially affords unpredicted environmental influences on the robot's behaviour; and, very practically, a physical robot and its behaviours are more compelling to both the specialized and unspecialized eye, whether modelling creativity or whatever.

That said, we carry out the development of robot controllers in simulation. This choice is largely pragmatic. Artificial evolution, though usually quicker than natural evolution, takes time, generally involving hundreds or thousands of generations. Simulating can often expedite the process.[8] Simulations also afford experimenters the opportunity to go back and observe developmental features of agents and behaviours. Over the course of an experiment, one can rewind the clock as it were to see how behaviour varies given changes in certain parameters. Finally, physical robots are neither cheap nor easy to build. And in early evolutionary stages, the agents are dumb, tending to destroy themselves by crashing into walls. Experimenting in simulation enables working out these knots, before the control software is transferred to physically embodied robots.

There are many examples where experimental results with physically embodied robots are interesting and unpredicted, and just in virtue of the robots *being situated*. Thus, one worry goes, a simulated version of the real thing may not faithfully track the way a similar embodied agent behaves. And this is largely because the simulated environmental features

[8] All of this is contingent on the complexity of the task being evolved, and the size of the population of agents.

may be inaccurate with respect to the actual environmental features. A situated agent may exploit or use these features in its behaviour.[9] Our methodology reflects these concerns. First, the sensorimotor activity (more specifically, of the infrared sensors) of the simulated agents in the early stages of experimentation is partly governed by actual readings of a physical, Khepera robot (Figure 0-1). This provides an initial set of simulated experiments, still informed by details of embodied robots. By carefully pairing these readings with noise, it has been repeatedly shown that simulated robot controllers can be successfully transferred to physically embodied robots without a deterioration in behaviour (Jakobi 1998). At a second stage of experimentation, simulated experiments based on measurements from the sensorimotor morphology of a custom built robot platform will be conducted (Figure 0-2). Once fit individuals have been evolved in simulation, the software controllers will be transferred to the physically embodied robot. This final step may prove the crucial one for getting interesting or radically unpredictable results.

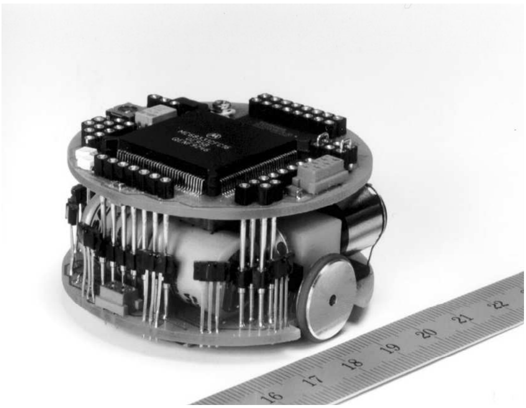

Figure 0-1 Khepera robot

[9] For example, see Scutt 1994; Bird and Layzell 2002.

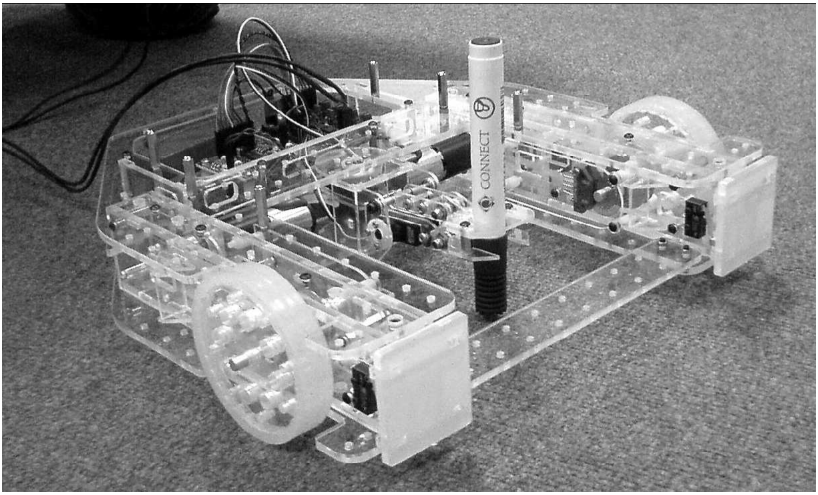

Figure 0-2 Early prototype robot at Sussex

Our robots use artificial neural networks controllers (ANNs)(Nolfi 1997). In such systems, the low level primitives consist of "neurons" and connections between them. In our khepera simulations, each robot controller consists of seven sensors (six infrared and one line detector) and six motor neurons (a pair of left motor neurons, a pair of right motor neurons, a pair of pen motor neurons). Each sensor connects to every motor neuron, making 42 total connections in the network. At each sensory-motor cycle, the most strongly activated neuron out of each pair of motor neurons is selected to control the appropriate motor. The genotypes encode properties of the neurons and strengths of the connections between them in the ANNs, which are then decoded into robot controllers.[10]

A fitness trial involves placing an individual in a random position and orientation (with the pen always in the down state initially) in a walled arena and testing them for 200 (±20) sensory-motor cycles. Each individual in the population is tested over 10 trials. The individuals are all tested on the same series of initial positions and orientations each generation, and these change every generation. The population consists of 100 individuals and the experiments run for 600 generations.

We would ultimately like novel behaviour from our robotic agents. And it would be nice to get some drawings: some marks on the page that are in some way aesthetically pleasing or interesting. However, it would

[10] For more technical details, see Bird and Stokes 2006; Bird et al. forthcoming.

be a mistake, for reasons given above, to incorporate these desires in a heavy handed way into the fitness functions. A primary motivation for the use of ER is overcoming inductive biases. The initial fitness functions thus reward simple changes in sensory-motor correlations. An agent acquires fitness for correlation between changes in its line sensor state and changes in its pen state. The line detector has two states (on or off), as does the pen (up or down). An agent gains fitness for raising its pen when a line is detected and it gains fitness for lowering its pen when the line detector changes from on to off.[11] If agents crash into the arena boundaries the testing ends, limiting the opportunity to gain fitness. These functions are independent of our underlying theoretical motivations and biases about creativity.

4 - Results and Interpretations

Figure 0-3 and Figure 0-4 provide examples of low to mid fitness individuals in the first stage of experiments, where only the pen state/line sensor correlations and wall avoidance determine fitness.

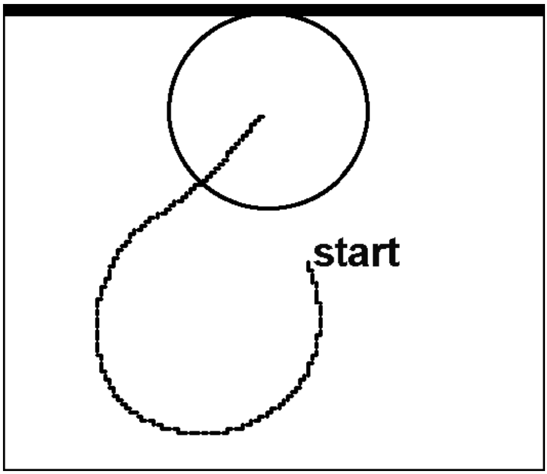

Figure 0-3 An example of low-fitness behaviour typical of a robot in the first generation. It makes a continuous line up to the point that it crashes into the arena wall. The robot does not change its pen state during the whole trial.

[11] In both cases, fitness is rewarded if the pen state changes within a short period of time after the change in line detection. So there is, as it were, a brief window of opportunity for acquisition of fitness.

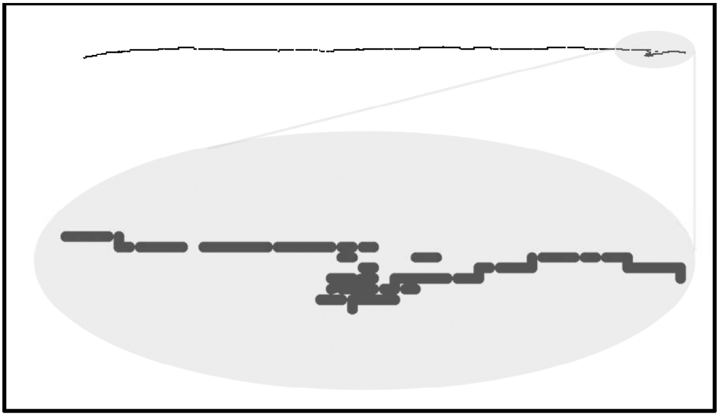

Figure 0-4 Performance of a mid fitness robot in an early generation. The image shows the line made in the arena - the robot started near the top left hand corner and moved across to the right top hand corner, raising and lowering the pen during this movement. The grey region is magnified to show the more complex marks the robot made at the end of the trial period: it made short line segments and moved forwards and backwards over them. The robot thereby gains fitness for this "squiggling" behaviour, correlating activity between changes in its pen state ("up to down" or "down to up") and changes in its line sensor state ("on to off" or "off to on").

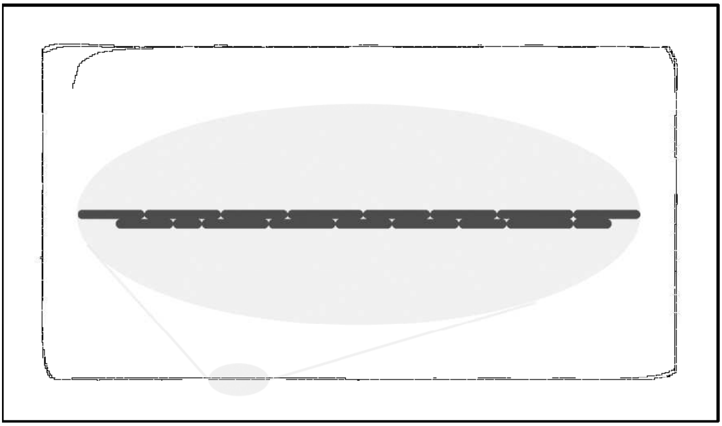

Figure 0-5 Performance of a high fitness individual after 500 generations. The robot completes one circuit of the arena with its pen down by following the walls. On its second circuit it sweeps left and right over the line, marking line segments parallel to the line it had previously left.

These individuals are *agents* on the broad construal suggested. The individual robot's activity is not imposed or controlled by the programmers. Rather, the robots are set free in the environment and their fitness assessed over a series of trials. Behaviour is, at least in this early stage of experimentation, reactive: it derives from the sensorimotor activity of the system, itself responsive to the environment. However, the interaction between agent and environment involves reciprocal feedback since the sensorimotor activity and thus the behaviour of the robot depend upon the environment which is itself partly constituted by marks left on the page by that very agent. When we hold the starting conditions and testing arena constant, different agents respond differently (where agents are individuated by their artificial neural network controllers). Indeed some of the agents act in behaviourally novel ways. Compare, for example, the results displayed in Figure 0-3 to Figure 0-5.

Across generations, moving from individual behaviour like that depicted in Figure 0-3 to Figure 0-4 to Figure 0-3, we see the emergence of behavioural novelty. Low fitness agents in early generations (Figure 0-3) do not go far, generally just running into the arena boundary straightaway. Mid-fitness agents in early generations (Figure 0-4) begin to act in ways that acquire fitness for changes in pen state/line sensor state correlations; some perform a kind of "squiggling" behaviour, sweeping back and forth over their own lines. And high-fitness agents in later generations (Figure 0-5) do the same, some of them exploiting reliable features of their environment, namely following the arena walls so as to score maximum fitness. Other agents in the same starting and testing conditions did not perform these behaviours. The behaviours were thus novel relative to the relevant populations. Indeed, some of these behaviours—e.g. the wall following behaviour—are novel relative to all agents and all starting and testing conditions. Novel behaviours thus emerge as a successful means for acquiring fitness.

These are preliminary results, and the interpretations thin. There are a number questions unanswered and technical issues unexplained. And the applicability of the conceptual framework depends importantly upon some of these questions and issues. Let all that stand. Even if these results fail to conclusively satisfy the two necessary conditions specified, they nonetheless afford fruitful comparisons and lessons.

5 - Robots and Romance

Recall the romantic thesis, which motivates a worry like the following. "What is all this talk about constraints and creativity? Isn't a creative process, by definition, one that is unconstrained? Surely the more constrained an artist or problem solver is, the less likely she is to secure some creative result or solution. So the project is wrong-headed from the start, involving highly constrained artificial agents."

There is a grain of truth to the thesis and the worry it motivates. We do sometimes need to overcome, insofar as we can, various environmental and conceptual constraints in order to have a creative thought or perform a creative act. A creative thought may be outside the conceptual space in which we are currently operating and constrained by. Kekulé would not have discovered the ring-like structure for the benzene molecule had he not thought outside the concepts of the organic chemistry of his day. Impressionism would not have been born, had artists like Monet not decided to pack up their studios and head outside.

But all of this is consistent with rejecting the supposition that creative processes must be unconstrained ones. The latter supposition is motivated by a romantic notion of creativity that rests on the following argument. (1) If an *F* is unpredictable, then *F* must result from processes that are unconstrained. (2) Creative *F*s are unpredictable. (C) Therefore, creative *F*s must result from processes that are unconstrained.

One can secure counterexamples to this romantic argument. In composing *The Well-Tempered Clavier*, Bach was constrained by the 12 tones scale, tempered tuning, and clavier instrumentation. His composition is, however, clearly a creative masterpiece. Poets famously constrain their compositions by working within certain poetic forms. A photographer might choose to limit herself by using an old camera or an outdated development technique. And so on.

There are also questions about unpredictability. First, must all creative *F*s be unpredictable? If we accept psychologically or behaviourally relative notions of creativity—in accordance with the novelty conditions considered above—then the answer seems to be no. With enough knowledge about the information and skills possessed by a child working on a problem, I might predict how she will solve the problem. And the solution may, as a matter of fact, be novel relative to her mind. Assuming that the other conditions on creativity (whatever they should turn out to be) are satisfied, this is an instance of creativity in spite of its being predictable. So (2) at least looks contentious.

A more fundamental question concerns the very notion of predictability, and thus, unpredictability. Is this an epistemic or a modal issue?[12] Is a particular failure to predict an event sufficient to reasonably

[12] To keep these options distinct, the relevant modality would presumably be nomological.

call the event "unpredictable"? Or, are unpredictable events only those that, in principle, given all of the relevant information, *cannot* be predicted? The romantic generally has an epistemic notion in mind, couching his descriptions in talk of surprise and unexpectedness. Nonetheless, to be charitable to the romantic, one shouldn't think of predictability in a purely subjective way. To do so would result in (appropriate) predictability attributions—to the same event—which vary with the epistemic agent. A relativity this trivial should be avoided. Instead, one can understand unpredictability just in terms of the normative expectations that one would—from a privileged perspective—have on any normal epistemic agent given her evidentiary relations with the world. If from such a stance, one would not reasonably expect the agent to foresee some forthcoming event *E* (or to have foreseen some past event), then *E* is unpredictable from the agent's perspective. If one would expect that agent to foresee *E*, then *E* is predictable from the agent's perspective.[13] This normative notion of predictability captures our ordinary understanding of the concept, while the modal notion is too far removed to be useful in a discussion of creativity.[14]

These problems to one side, our experiments provide a novel angle for rejecting the romantic argument. Our robots are highly constrained: by their sensorimotor morphologies, by their immediate environment, and by our fitness functions. However, they do act in unpredictable ways. The "squiggling" behaviour in Figure 0-4 and the additional wall following behaviour in Figure 0-5 are behaviours that the programmers did not, and indeed could not be reasonably expected to, predict. Irrespective of whether we call such behaviours creative, they are unpredictable in spite of their resulting from highly (and identifiably) constrained processes. We thus deny the romantic argument by denying the conditional in (1): unpredictability does not imply lack of constraint. The romantic can keep unpredictability, but not without constraints.

One could weaken the romantic argument in the following way. (W1) If an *F* is unpredictable, then *F* must result from processes that are minimally unconstrained. (W2) Creative *F*s are unpredictable. (WC)

[13] This begs many questions about epistemic justification and norms, meta-knowledge, and human rationality. And so the clarification of predictability is only as good as one's assumptions about these issues. However, the clarification should suffice to give one an intuitive, working understanding of a normative, inter-subjective notion of predictability, sufficient at least for assessing the romantic argument.

[14] For a discussion of unpredictability and creativity, see Boden 1995, and 2004, p. 233-55.

Therefore, creative *F*s must result from processes that are minimally unconstrained. Interpret 'minimally unconstrained' functionally: a process is minimally unconstrained if at least one component or event in the process can produce varying outputs given a particular input. Any process with this input-output flexibility, let's say, is minimally unconstrained.

Intermediate between the weak and strong argument, the romantic might argue thus. (I1) If an *F* is unpredictable, then *F* must result from processes that are mostly unconstrained. (I2) Creative *F*s are unpredictable. (IC) Therefore, creative *F*s must result from processes that are mostly unconstrained.

The intermediate argument fails for the same reasons the strong argument fails. Our robotic agents are very constrained—by our fitness functions, by their sensorimotor morphology, by their (artificial) evolutionary history, and by their immediate environment. They are *not* mostly unconstrained. They nonetheless act in unpredictable ways. We thus have a counterexample to (I1).

One might worry that the romantic is too easy a target. But taking the romantic position seriously reveals the theoretical promise in our evolutionary robotics approach. One can, using minimal, bottom up methods of experimentation, test notions of and assumptions about creativity. The romantic thesis provides a simple case in point. Supported by arguments of varying strengths, it is shown false by empirical example. Such methods should enable testing of other theses, distinctions, and concepts of creativity, providing novel angles on an age old set of questions. So even if an easy target, consideration of the romantic theses—supported by the strong and intermediate arguments—proves instructive of the wider applicability of our ER approach. This is also true for the weak argument for the romantic thesis.

The weak argument is valid and, at least on the liberal interpretation suggested, its premises plausibly true. This is a challenge, however, neither to the suggested conceptualization of minimal creativity nor to the corresponding robotics experimentation. On the contrary: we accommodate the intuition underlying the weak argument with the agency condition. The input-output flexibility in (W1) and the corresponding conclusion is akin to the no strings attached autonomy required for agency and thus, as argued, for minimal creativity. The weak romantic conclusion re-stated is simply that creative *F*s result only from autonomous processes. In critiquing the romantics, we have demonstrated the empirical possibility of constrained minimally creative processes. We have also come full circle, revealing the grain of truth in the romantic position: creativity requires autonomous agency. We have accomplished both by way of bottom up, synthetic modelling and its interpretation. This implies that either artificial life researchers are romantics after all *or* that the romantic can have his romantic argument only at the cost of his romanticism. The second seems the safer bet.

References

Beer, R. and J., Gallagher. 1992. Evolving dynamical neural networks for adaptive behaviour. *Adaptive Behaviour* 1: 91-122.

Bergson,H. 1998[1911]. *Creative Evolution*, trans. A. Mitchell. New York: Dover.

—. 1992[1942]. *The Creative Mind*, trans. M. Andison. New York: Citadel Press.

Bird, J. and Layzell, P. 2002. The Evolved Radio and its Implications for Modelling the Evolution of Novel Sensors. *Proceedings of Congress on Evolutionary Computation*, 1836-41.

Bird, J. and D. Stokes. Forthcoming. Minimal Creativity, Evaluation, and Fractal Discrimination. *Proceedings of the 4th International Joint Workshop on Computational Creativity*, edited by A. Cardoso and G. Wiggins. London: Goldsmith's College.

—. 2006. Evolving Minimally Creative Robots. *Proceedings of The Third Joint Workshop on Computational Creativity*, European Conference on Artificial Intelligence, edited by S. Colton and A. Pease 1-5.

Bird, J., D., Stokes, P., Husbands, P., Brown, and B., Bigge. Forthcoming. Towards Autonomous Artworks. *Leonardo Electronic Almanac*

Boden, M. Forthcoming. Stillness as Autonomy. *Proceedings of the Compters in Art and Design Education-Conference 2007*.

—. 2004. *The Creative Mind*, 2nd Edn. London: Routledge.

—. 1995. Creativity and Unpredictability. *SEHR*, Vol. 4. Stanford Humanities Review.

Clark, A. 1997. *Being There: Putting Brain, Body, and World Back Together Again*. Cambridge, MA: MIT Press.

Cliff, D., I. Harvey, and P. Husbands. 1993. Explorations in evolutionary robotics. *Adaptive Behavior* 2: 73–110.

Cohen, T. 2003. The Inexplicable: Some Thoughts After Kant. In *The Creation of Art: New Essays in Philosophical Aesthetics*, edited by B. Gaut and P. Livingston. Cambridge: Cambridge University Press.

Collingwood, R.G. 1938. *The Principles of Art*. London: Oxford University Press.

Dewey, J. 1958. *Art as Experience*. New York: Capricorn.

Findlay, A. 1965. *A Hundred Years of Chemistry*, 3rd Edn., edited by T.I. Williams. London: Duckworth.

Floreana, D., P. Husbands, and S. Nolfi. 2007. Evolutionary Robotics. *The Handbook of Robotics*, edited by B. Siciliano and O. Khatib, Springer.

Harvey, I., P., Husbands, and D., Cliff. 1994. Seeing the light: Artificial evolution, real vision. *From Animats to Animals 3*: *Proceedings of the Third International Conference on Simulation of Adaptive Behavior*, edited by D. Cliff, P. Husbands, J.-A. Meyer, and S.W. Wilson. Cambridge, MA: MIT Press.

Holland, J. 1975. *Adaptation in Natural and Artificial Systems*. Cambridge, MA: MIT Press.

Husbands, P., I. Harvey, D. Cliff, and G. Miller. 1997. Artificial evolution: A new path for artificial intelligence. *Brain and cognition* 34:130-159.

Jakobi, N. 1998. Running across the reality gap: Octopod locomotion evolved in a minimal simulation. *Evolutionary Robotics: First European Workshop, EvoRobot98*, edited by P. Husbands and J.-A. Meyer, Springer, 39–58.

Kant, I. 1987[1790]. *Critique of Judgment*, translated by W.S. Pluhar. Indianapolis: Hackett.

Koestler, A. 1964. *The Act of Creation*. New York: Macmillan.

Nolfi, S. 1997. Using emergent modularity to develop control system for mobile robots. *Adaptive Behavior* 5: 343–363.

Nolfi, S. and D., Floreano. 2000. *Evolutionary Robotics: The Biology, Intelligence and Technology of Self-Organizing Machines*. Cambridge, MA: MIT Press.

Pollack, J.B., H., Lipson, G., Hornby, and P., Funes. 2001. Three generations of automatically designed robots. *Artificial Life* 7: 215 – 223.

Quinn, M., L. Smith, G. Mayley, and P. Husbands. 2003. Evolving controllers for a homogeneous system of physical robots: Structured cooperation with minimal sensors. *Philosophical Transactions of the Royal Society of London, Series A: Mathematical, Physical and Engineering Sciences*. 361: 2321–2344.

Ray, T.S. 1995. An Evolutionary Approach to Synthetic Biology: Zen and the Art of Creating Life. In *Artificial Life*, edited by C. Langton. Cambridge, MA: MIT Press.

Sawyer, D.K. 2000. Improvisation and the Creative Process: Dewey, Collingwood, and the Aesthetics of Spontaneity. *The Journal of Aesthetics and Art Criticism* 58: 149-61.

Schelling, F.W.J. 1936[1809]. *Of Human Freedom*, translated by J. Guntman. Chicago: Open Court.

Scutt, T. 1994. The five neuron trick: Using classical conditioning to learn how to seek light. *From Animals to Animats 3: Proceedings of the Third International Conference on Simulation of Adaptive Behavior*, edited by D. Cliff, P. Husbands, J.-A. Meyer, and S.W. Wilson, 364–370. Cambridge, MA: MIT Press,

Stokes, P. 2005. Creativity from Constraints: The Psychology of Breakthrough. New York: Springer