Computational Creativity as Dynamic, Multiobjective, Multiagent Optimization

Sean Luke
Department of Computer Science
George Mason University
Washington, DC, USA
sean@cs.gmu.edu

Abstract

I propose a unified model of computational creativity which treats it as a dynamic, multiobjective, multiagent optimization process. I present an informal model, discuss its attributes and features, and argue for why these elements (optimization, dynamic change, multiple objectives, and multiple agents) are important to a framework in which computationally creative algorithms may be discussed, analyzed, and compared.

Introduction

In this position paper I propose a model of computational creativity which attempts to unify four features I feel are critical to consider in the development of creative algorithms. Some of these features have been discussed to one degree or another in other literature, but largely individually, and I think they have been given short shrift when considered together. The goal is not to propose an algorithm, nor construct a model for simulation of actual human creativity, but to build a framework or common language in which one can describe and compare methods, and to argue for the importance of certain areas which have not been adequately considered. I am fully aware of the armchair philosophy involved here, but hope it might serve to stir discussion.

These four features are as follows:

- Computational creativity is an optimization process.
- It is dynamic optimization, that is, it occurs in an environment which changes over time and strongly impacts on assessment functions and optimization biases.
- It is (commonly) multiagent optimization: there are multiple optimization processes running in parallel. These processes may use entirely different algorithms and internal representations of artifacts, yet impact on one another.
- It is multiobjective optimization: it is optimizing not only for both value and novelty, but for multiple and different aspects of each at the same time.

A Holistic System In the proposed model, one or more optimization processes are encapsulated in an *agent*, and the environment may hold one more or agents which impact on one another. A creative optimization process operates over a *creative space* of *internal representations* of *artifacts*. The process iteratively selects an internal representation, generates an artifact from it, and then assesses the artifact with *multiple* objective functions of value and novelty in several dynamically changing *contexts* of the current environment, potentially including other agents. The agents may *influence* one another, and so too may processes within a single agent. The creative space, the number and type of objective functions, the contexts, the number and kinds of agents in the environment, their creative processes, and even the parameters of the optimization processes an agent uses are all subject to dynamic change over time. The model also supports interaction with humans in a co-creative context.

This is clearly a systems view of creativity similar to the DIFI framework (Feldman, Csikszentmihalyi, and Gardner 1994): an agent or process is approximately a DIFI individual, and a context in some sense encompasses DIFI's notions of Field and Domain. A distinguishing feature of the model, however, is that it commits strongly to distribution. There can be multiple agents. Each agent can have multiple creative processes and multiple representations of artifacts. Each process can have multiple contexts for assessment with multiple value and novelty objective functions. Each agent can have multiple different peer groups of other agents for influence.

I do not propose a specific algorithm for the model's creative optimization process, though I discuss Evolutionary Algorithms as one example. There are many algorithms which could reasonably produce "creative" artifacts using different optimization procedures, all of which could reasonably fit under this framework.

Novelty and Value as Optimization Objective Functions

Literally on page 1, Boden (1992) defined creativity as:

> Creativity is the ability to come up with ideas or artifacts that are *new*, *surprising*, and *valuable*. [Emphasis hers]

Two of these features (*novelty* and *value*) have come to be the hallmarks of artifacts in the computational creativity literature, at least among computational models. My proposed framework likewise deals with optimizing and producing artifacts with respect to novelty and value as objective functions.

Boden suggests that an artifact is *novel* or *new* in one of two ways: first it may be *P-creative*, or *psychologically creative*, meaning that it is new to you the creator. Or the artifact may be *H-creative* or *historically creative*, meaning that it is new in all of creation history (Boden 1992). I argue that these are degenerate cases: artifacts may have different degrees of novelty depending on the *context* in which they are assessed and the *level* of their dissimilarity with historical artifacts in that context. For example, a blues song may be very novel to critics, but not to other musicians in your circle.

Creative artifacts must also provide *value* to some audience. Value can be either *objective* or *subjective*. It might be beyond question that you have constructed a faster car, but debatable as to whether that car might be prettier. But even if a value assessment is objective, its interpretation is not: the car's speed is objective, but the desired goal (faster, slower), is up to you. Thus both objective and subjective assessments may be viewed as part of an *aesthetic*, and as you are seeking both a pretty *and* fast car, there may be more than one aesthetic.

Dynamic Optimization

If I may start with a nitpick. The notion of *creativity as search* is common, and tends to be formulated in a fashion strongly reminiscent of state-space search (Wiggins 2006b; Ventura 2011; Ritchie 2012; Linkola and Kantosalo 2019). However Wiggins warned (2006a, §2.4.11) that the creative process was somehow different from "the familiar state-space search in the AI literature" for several valid reasons. I think many of these reasons boil down to creativity being more properly described as optimization.

The term *search* is historically muddled in AI. It classically refers to methods like *state-space search,* where there is something to search *for*, that is, something which satisfies a *goal predicate*; though *local search* is a misnomer for optimization methods such as hill-climbing, and the term "search" leaks into certain other optimization methods. But in optimization there is no goal predicate: the objective is usually to find as good a result as possible given the resources (time, memory, etc.) available. This is normally done by iteratively producing results from a space of candidate solutions, assessing them, and then producing more results influenced by the assessment of the earlier results. There is no guarantee that there is an optimal result at all, nor that there is only one.

Unlike state-space search in particular, optimization's version of Wiggins's $\mathscr{T}$ traversal function would not necessarily be a local state transition or reachability function. Rather it would be a more general function which simply takes past candidate solutions and produces new ones. In most global optimization algorithms this function would draw from a probability distribution over the *entire space* of candidate solutions. Thus everything would be theoretically reachable in one step, though local traversal would be more prevalent.

When applied to computational creativity, optimization can but would not necessarily require a hard constraint of *validity* as in $\mathscr{R}$ from Wiggins (2006a). Rather than draw a sharp boundary around all valid artifacts, we could define artifacts distant from the "expected form" as being of lesser value. In an optimization algorithm there *would* exist a special subset of Wiggins's $\mathscr{U}$ consisting of *all artifacts that the algorithm can represent internally* (as *genotypes*), as discussed later.

Computational Creativity is an Optimization Process A creator is wandering through the space of artifacts, seeking artifacts refined or improved with respect to a combination of various novelty and value assessments. These artifacts are generated based on some potentially stochastic function applied to his current history of artifacts, mental state, and feedback received from various sources.

Even if the creator is indifferent to external assessment, he may still be driven by his personal assessment. And even given purposeless creator, a DIFI-style Field would still act as a rejection sampler. At any rate, it seems likely that as an engineering pursuit, the aim of an artificial creation *algorithm* would be to produce elements optimizing some criterion.

Dynamic Change over Time Computational creativity is an optimization process in a *dynamically changing* environment. That is, the expected trajectory of the optimization process may deviate over time due to external factors. For example, as critics, audiences, and society evolve in taste or style or needs, the notion of value would likewise change. To quote Tower of Power, what's hip today might become passé.

A multiagent context (discussed next) introduces more opportunities for dynamic change over time. As other agents (or other creative processes in the same agent) produce artifacts, these will impact on the notion of novelty over time. Other agents (or humans) could also *influence* an agent through their work, either deliberately or inadvertently.

It is possible that changes in value, or perhaps radical new discoveries by other agents, will require obsoleting the optimization process itself and adopting an improved one.

Multiagent Optimization

Creativity is often done in the context of other creative agents. Of course it does not *require* more than one agent: but even then an agent may have multiple creative processes — irons in the fire — which might influence each other, and so a single agent may be usefully thought of as a multiagent optimization process. For example, an agent may sometimes be designing cars, and other times drawing pictures of plants, and have one process draw unexpected inspiration from the other.

At any rate, multiagent systems can impact on computational creativity in several ways, the first two of which are modeled in Saunders (2019). First, other agents (in addition to fans, critics, etc.) might directly assess an agent's creative work. Second, other agents' output might *influence* (or *inspire*, or even *appall*) an agent, biasing his optimization trajectory. Third, the creative output of other agents may change the *zeitgeist* in which an agent's creative work is assessed, and thus the assessment functions themselves. Fourth, agents may *cooperate* to produce creative work by trading off discoveries, or *compete* to moot one another's efforts.

The second and third cases are interesting algorithmically, in that a creative agent is biased in ways other than raw assessment. This bias could be in the form of seeding: an agent adds another agent's artifact to its current distribution from which to resample artifacts. Or the discoveries of another agent may act as an attractive target, bending the trajectory of an agent as he wanders through the optimization space.

In the fourth case, external agents might affect the dynamic change of value functions. For example, in *competitive optimization* one agent is seeking a better mousetrap, while another agent is building a better mouse, and thus they are changing the goalposts on one another in real time. Or consider *cooperative optimization*: if Agent A is working on part 1 of a two-part problem, new approaches by Agent B working on part 2 may force Agent A to reevaluate the value of her solutions. Or Agent A's work might be meant to complement Agent B's work, and while Agent B is disinterested in A, his output changes how A's work is assessed.

Multiagent systems provide opportunities for multiple contexts and audiences, and for multiple peer groups of agents. An agent may be aware of both local bands and non-local but genre-related groups, and he may produce songs appreciated differently by local audiences, internet fans, or online critics.

We must also consider the possibility that different agents will employ different optimization algorithms with different *internal representations* for their candidate artifacts. For example, a genetic programming (GP) system would represent artifacts internally as tree structures, whereas a neural network (NN) would represent them as fixed-length arrays of numbers. To assess the final artifact (a car say), we must first map the internal representation to car form. But similarity among internal representations of artifacts (genotypes) may not be well correlated to similarity in car form (*phenotype*).

Furthermore the mapping is not a bijection. There may be valuable and novel cars for which only GP has a genotype (the NN simply can't construct it), or for which the probability of producing a genotype in NN is low due to its very different optimization approach compared to GP. Indeed, in response to a car produced by GP — or a human! — the neural network agent might say, "I would never have *thought* of that". Thus it is possible for agents following one particular approach to influence other agents by making artifacts that the other agents are simply incapable of producing (but will now try).

Computational Co-creative Systems Software may collaborate with a human in the creative process (Davis et al. 2015). Karimi et al. (2018) define such systems as the "interaction between at least one AI agent and at least one human where they take action based on the response of their partner and their own conceptualization of creativity during the co-creative task." In our model this is simply an extension of a cooperative multiagent creative system, where at least one creator is a human, as a stand-in for an agent. A human in the agent's peer group may influence, cooperate with, or reveal creative artifacts to the agent; or influence its value function. I do not here model the impact of the agent and its feedback *on the human,* as that would enter the realm of psychology!

Who Determines What is a Creative Work? The social creativity models of Saunders (2019) and Linkola and Kantosalo (2019) both place emphasis on agents serving both as creators and as the DIFI Field, that is, as the gatekeepers of value or novelty. In Saunders, agents produce creative works, which are then handed off to other agents for assessment and feedback. In Linkola and Kantosalo, the validity, transition, and value functions $\mathscr{R}, \mathscr{T}, \mathscr{E}$ from Wiggins (2006b) are extended to produce the agent-wide collective sets $\mathcal{R}_S, T_S, E_S$ of "societal-wide valid" artifacts, artifacts reachable by the society as a whole, and artifacts with "society-wide value".

In the model proposed here, this is not the case. While agents *could* serve as the Field or as part of it, non-personal value assessment would more often be primarily made up of external entities such as audiences and critics. Agents would impact on novelty functions of course, and influence other agents, through the dissemination of their artifacts.

Multiobjective Optimization

Assessment of creative artifacts has always been multiobjective: at the very least it has been commonly assumed that artifacts are assessed based on both *novelty* and *value*. But even these may be further broken into multiple subobjectives. The creative output of a given agent may be valuable (or not) in different ways and in different degrees to the agent himself and to different audiences, be they fans, critics, or other agents; and it may assessed via different objective measures. It is also possible that creative output may be novel to a different degree in different contexts, giving rise (for example) P- versus H-creativity; or the work may be considered *more novel* by one audience than by another. The number and type of objective functions used by an agent's optimization system may vary dynamically as he comes in contact with different audiences and groups, and likewise the multiple assessments of a given artifact may change over time.

One common way to optimize multiple objectives is to attempt find solutions approaching the *Pareto Front*. A solution is in the Pareto Front if no other solution is superior to it in all objectives. While I do not suggest a particular multiobjective optimization approach, I note that classic approaches based on pure Pareto Front methods may not work well, and approaches which emphasize or encourage a subregion in the Front may be more effective. This is because the corners of the Pareto Front (such as "not at all novel but highly valuable") are not likely to be considered creative: there clearly must be some sort of inclusion of both novelty and value.

Likewise, it has been argued that *extremely novel* artifacts (random noise, say) might be considered undesirable due to a non-monotonic novelty function ("it's too different") (Boden 1992; Paese, Winterstein, and Colton 2001; Saunders 2019). I argue that such solutions would instead be downgraded not because they are unusual but because the critic cannot fathom how they could be of *value*: that is, they would fall in the "highly novel but not valuable" corner of the Pareto Front.

The various objectives of novelty and value may be at odds with one another. For example, some art critics might value works based on stylistic similarity to exemplars ("the classics"), thus setting up a tension between novelty and value. But this is not necessarily the case in general. Engineers would be more than willing to accept extraordinarily novel, indeed alien, solutions if they were shown to work well.

Example Optimization Realization

Optimization approaches drawn from neural networks, reinforcement learning, state-space methods, and others could fit under this model. But to demonstrate model feasibility, I offer one prominent technique: Evolutionary Computation (EC) (Luke 2013), a broad family of stochastic optimization algorithms including the *Genetic Algorithm*. EC has

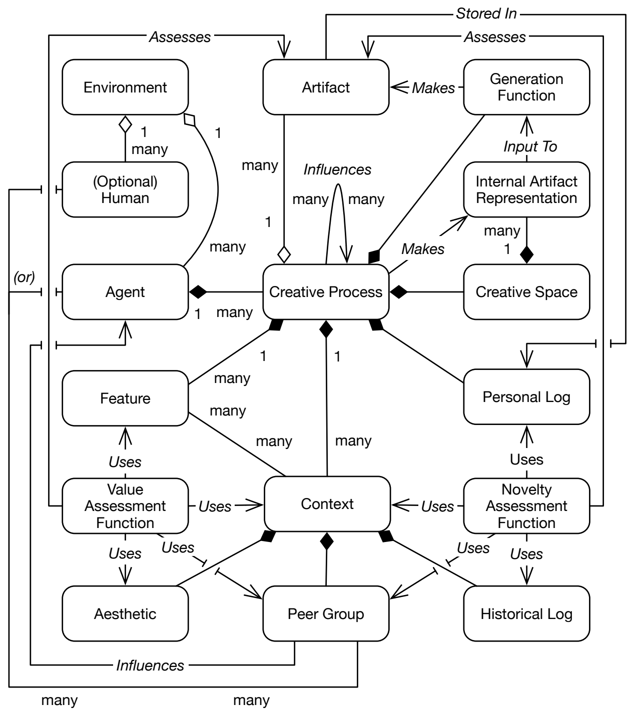

Figure 1: UML diagram of model classes and relationships.

standard methods covering every facet of the model. This includes techniques for parallel optimization processes in which agents communicate artifacts (*island models, Particle Swarm Optimization*) or influence other agents' value functions through the introduction of artifacts (e.g. *competitive* or *cooperative coevolution*). EC has a robust set of methods for multiobjective optimization, optimizing in the face of dynamically changing objective functions, guaranteeing diversity, and interacting with humans (*interactive evolution*); and is readily adapted to heterogenous optimization algorithms.

Model Overview

The elements of the model are shown in Figure 1. We begin with the *environment*, which holds one or more *agents* and possibly *humans* for co-creative systems. An agent is a computational entity engaged in creative output, and may have one or more *creative processes* active at a time. Each creative process is an optimization procedure which produces *artifacts* over time, by drawing *internal representations* of them from a *creative space*, then converting them into artifacts via a *generation* function. The generation function makes possible heterogeneous, parallel optimization approaches.

The artifacts are then *assessed* for novelty and value. Novelty is assessed with regard to a *context* in which the artifact has been produced. Several contexts may be associated with and special to a given process. Agents maintain *personal logs* of past artifacts, and a context holds a *historical log* of past artifacts generated by agents no longer existing, and a *peer group* of other agents, or humans, whose personal logs may be consulted in order to determine how novel the new artifact is. The novelty of an artifact may be assessed in different contexts, such as a personal context ("it's new to me"), or a context of a small peer group, or a wider historical context, and so the assessment of novelty may be multiobjective. A context's peer groups and historical log may also be used to *influence* the optimization (such as through inspiration).

Value is similarly assessed in terms of one or more contexts associated with the process. For purposes of value assessment, a context holds an *aesthetic*, that is, all the information, objective data, audiences, critics, norms, rules, guidelines, personal beliefs, etc., by which an artifact's value may be assessed. Humans and agents in the context's peer group may optionally provide input. Value is assessed with regard to a *feature* of the artifact, such as how pretty it is, or how fast it is. Features may be both objective (speed) or subjective (beauty). Like novelty, the value of an artifact may be assessed in different contexts (different audiences with different opinions, say). Further, it may be assessed with respect to different features. As there can be more than one context and more than one feature, value assessment is also multiobjective.

This model is a multiagent system. There can be many agents whose products inform the contexts in which an agent's artifacts are assessed for novelty; and these agents might also be part of the audience which assesses value. Even without other agents, the agent itself may have multiple creative processes which could influence one another even though they're operating over different creative spaces.

Nearly everything is dynamically changing. Agents in the environment may come and go over time. So too can an agent's creative processes, and the creative space of a process may evolve and change as well. Artifacts are produced over a timeline. Assessments are done with regard to changing contexts and features. Contexts can change in their makeup and effect with the current zeitgeist and style. Artifacts may have their novelty and value reassessed in light of new discoveries.

Model Details

Agents and Creative Processes An *environment* E is a set of one or more *agents* $\mathbf{A} = \{A_1, ..., A_a, ...\} \subseteq \mathbb{A}$. The number of agents is not fixed and agents may be introduced, removed, or changed in state over time, and so the state of E at time t may be described as E^t, its current set as $\mathbf{A}^t$, and the state of a given agent A_a as A_a^t. We will continue to use the t convention for other elements throughout the model. E may also contain a set of humans $\mathbf{M}^t = \{M_1^t, ..., M_m^t, ...\} \subseteq \mathbb{M}$, whose composition may change over time.

At time t an agent A_a^t employs a set of one or more *creative processes* $\mathbf{P}^{a,t} = \{P_1^{a,t}, ..., P_p^{a,t}, ...\} \subseteq \mathbb{P}$. Agents may vary in the number and type of creative processes they employ over time (hence *t*). A creative process $P_p^{a,t}$ is running an optimization algorithm, and so has an internal state which changes over time t as well.

Artifacts, Creative Spaces, and Logs An artifact is a product output from a creative process. For our purposes, it a sample drawn from a large (and possibly infinite) set of possible artifacts called a *creative space*. Each creative process $P_p^{a,t}$ is associated with a single creative space $S^{a,p,t} \in \mathbb{S}$, which can change in its membership over time (hence t). The creative space holds artifacts in their *internal representation* $r^{a,p,t}$ appropriate to the optimization process.

In order to assess the artifact, or make it understood by other agents, we must *generate* it from the representation. An artifact $x^{a,p,t}$ is produced at time t by $P_p^{a,t}$ by first drawing $r^{a,p,t}$ from $S^{a,p,t}$ and then converting it to $x^{a,p,t}$ via a generation function $x^{a,p,t} \leftarrow \textit{Generate}(P_p^{a,t}, r^{a,p,t})$. To keep things simple, we may assume $P_p^{a,t}$ produces only zero or one artifacts at any given time t, and so $x^{a,p,t}$ and $r^{a,p,t}$ are uniquely defined. The creative process $P_p^{a,t}$ maintains a *personal log* $L^{a,p,t}$ of artifacts it has produced up until time t.

Contexts and Influence Each creative process $P_p^{a,t}$ holds one or more *contexts* $\mathbf{C}^{a,p,t} = \{C_1^{a,p,t}, ..., C_c^{a,p,t}, ...\} \subseteq \mathbb{C}$ which together affect the objective functions used in the process's optimization, and so bias its production of artifacts. Contexts can come and go, and will change in state over time.

Each context has two aspects. First, the context has an *aesthetic* $Z^{a,p,c,t}$, which is all collective information used to assess the value of a creative artifact and so influences the creative process. Second, the context has a *memory* of past artifacts to be used for novelty assessment. Part of this memory is drawn from the personal logs of a *peer group* $G^{a,p,c,t} \subseteq (\mathbf{A}^t \vee \mathbf{M}^t)$ of other agents $\mathbf{A}^t$ and humans $\mathbf{M}^t$ (to the degree a human's "log" is available). Another part is a *historical log* $H^{a,p,c,t}$ of artifacts of agents and humans known to the agent but no longer present at time t. Aesthetics, peer groups, and logs change over time.

Agents and humans in peer groups can *influence* an agent via their artifacts, biasing the creative process in ways external to assessment as appropriate to the process's algorithm. An agent's creative processes may also influence one another.

Novelty Assessment A creative process may contain multiple novelty assessments, each an application of the *novelty function* in a given *context*. The novelty function $n \in \mathbb{R} \leftarrow \textit{Novelty}(C_c^{a,p,t'}, H^{a,p,c,t'}, L^{a,p,t'}, G^{a,p,c,t'}, x^{a,p,t})$ assesses the novelty of an artifact $x^{a,p,t}$ with respect to context $C_c^{a,p,t'}$ at time $t' \geq t$ compared to artifacts generated by agents in its peer group $G^{a,p,c,t'}$ and held in their respective personal logs, or artifacts in the context's historical log $H^{a,p,c,t'}$, or in the process's own personal log $L^{a,p,t'}$. We say t' rather than t because an artifact $x^{a,p,t}$ may be reassessed differently in the future, though it can only be compared fairly for novelty against artifacts $x^{a',p',t''} : \forall a' \in \mathbb{A}, \forall p' \in \mathbb{P}, \forall t'' < t \leq t'$ found in the logs at time t'.

Value Assessment and Features A creative process may have multiple value assessments, each applying the *value function* in a given *context* and with respect to a given *feature*. For each process $P_p^{a,t}$ at time t there is a set of one or more such features $\mathbf{F}^{a,p,t} = \{F_1^{a,p,t}, ..., F_f^{a,p,t}, ...\} \subseteq \mathbb{F}$. A feature is immutable, but *which* features are held by a given process, and the number of them, may change over time, hence the t. Contexts may use some or all of the features available in a process as appropriate. The value function $v \in \mathbb{R} \leftarrow \textit{Value}(C_c^{a,p,t'}, Z^{a,p,c,t'}, F_f^{a,p,t'}, G^{a,p,c,t'}, x^{a,p,t})$ assesses the value of $x^{a,p,t}$ in context $C_c^{a,p,t'}$ at time $t' \geq t$ with respect to its aesthetic $Z^{a,p,c,t'}$ and a given feature $F_f^{a,p,t'}$. This function may optionally take into consideration feedback from agents and humans in the context's peer group $G^{a,p,c,t'}$. As features, contexts, peer groups, and aesthetics can change over time, valuations can do so as well.

Conclusion and Future Work

I provide a unifying model and argue that existing models of computational creativity have not adequately considered it as a dynamic optimization process, responding to objectives in different contexts, and in an environment with other processes offering competition, collaboration, and inspiration.

The present model has shortcomings which may be addressed in future versions. It does not yet consider Boden's *transformational creativity* (1992). It does not consider artifacts which are incomplete or improved over time. Finally, it does not consider *combinatorial creativity*, whereby artifacts are the synthesis of other artifacts combined in novel ways.

My thanks to Victoria Hoyle and Habeen Chang for their considerable help in the development of the paper.

References

Boden, M. 1992. *The Creative Mind.* Abacus.

Davis, N.; Hsiao, C.-P.; Popova, Y.; and Magerko, B. 2015. An enactive model of creativity for computational collaboration and co-creation. In Zagalo, N., and Branco, P., eds., *Creativity in the Digital Age*. Springer. chapter 7.

Feldman, D. H.; Csikszentmihalyi, M.; and Gardner, H. 1994. *Changing the World: A Framework for the Study of Creativity*. Praeger.

Karimi, P.; Grace, K.; Maher, M. L.; and Davis, N. M. 2018. Evaluating creativity in computational co-creative systems. In *International Conference on Innovative Computing and Cloud Computing*.

Linkola, S., and Kantosalo, A. 2019. Extending the creative systems framework for the analysis of creative agent societies. In *International Conference on Computational Creativity*, 204–2111.

Luke, S. 2013. *Essentials of Metaheuristics*. Lulu, 2nd edition. http://cs.gmu.edu/∼sean/book/metaheuristics/.

Paese, A.; Winterstein, D.; and Colton, S. 2001. Evaluating machine creativity. In *4th International Conference on Case Based Reasoning*, 129–137.

Ritchie, G. 2012. A closer look at creativity as search. In *International Conference on Computational Creativity*.

Saunders, R. 2019. Multi-agent-based models of social creativity. In Veale, T., and Cardoso, F. A., eds., *Computational Creativity: The Philosophy and Engineering of Autonomously Creative Systems*. Springer. 305–326.

Ventura, D. 2011. No free lunch in the search for creativity. In *International Conference on Computational Creativity*.

Wiggins, G. A. 2006a. A preliminary framework for description, analysis and comparison of creative systems. *Knowledge-Based Systems* 19:449–458.

Wiggins, G. A. 2006b. Searching for computational creativity. *New Generation Computing* 24:209–222.