Searching for Surprise

Georgios N. Yannakakis and Antonios Liapis
Institute of Digital Games, University of Malta, Msida, Malta
{georgios.yannakakis; antonios.liapis}@um.edu.mt

Abstract

Inspired by the notion of *surprise* for unconventional discovery in computational creativity, we introduce a general search algorithm we name *surprise search*. Surprise search is grounded in the divergent search paradigm and is fabricated within the principles of metaheuristic (evolutionary) search. The algorithm mimics the self-surprise cognitive process of creativity and equips computational creators with the ability to search for outcomes that deviate from the algorithm's expected behavior. The predictive model of expected outcomes is based on historical trails of where the search has been and some local information about the search space. We showcase the basic steps of the algorithm via a problem solving (maze navigation) and a generative art task. What distinguishes surprise search from other forms of divergent search, such as the search for novelty, is its ability to diverge not from earlier and seen outcomes but rather from predicted and unseen points in the creative domain considered.

Introduction

The search for unconventional computational outcomes has traditionally been the core challenge of computational creativity (Boden 2004). As a response to this challenge, several notions or dimensions of creativity have been investigated, either as search heuristics or as criteria for the assessment of the creative process and its outcomes. *Value* and *novelty* have arguably been the most popular of those notions (Boden 1995; Ritchie 2007; Wiggins 2006). According to Ritchie (2007), value is the degree to which an outcome is of high quality whereas novelty is the degree to which an outcome is dissimilar to existing examples within a domain. Boden (2004) and Ritchie (2007) claim that novelty and value are the essential criteria for assessing creativity and Wiggins (2006) provides definitions for novelty and value as different features that are relevant to creativity.

According to alternative views within computational creativity, however, novelty and value are not sufficient for the discovery of unconventional outcomes (Grace et al. 2014; Kulkarni and Simon 1988). Boden (1995) argued for the distinction between novelty, unexpectedness and value. This distinction is derived from the observation that creative outputs of high value are not merely novel but also unexpected. Both outcomes from creative artwork and outcomes from creative problem solving are often attributed creativity due to the unexpectedness they elicit to an audience of evaluators. The notion of surprise appears to be an underlying aspect of the creative process which eventually is manifested on the final creative outcome (Macedo et al. 2009; Macedo and Cardoso 2002). As novelty does not cater for the temporal aspects of discovery, it is suggested that *surprise* is included as a core assessment dimension of a generated outcome (Grace et al. 2014; Maher 2010). Further, computational processes that realize *transformational creativity* (Boden 2004) in which the creator breaks the domain's rules and leads to unconventional problem solving and highly novel yet important artifact creation are far from being achieved. Recent views on aspects of surprise, however, have been proposed as potentiators of transformational creativity (Grace and Maher 2015).

Following the perspective of a large volume of work within computational creativity (Grace and Maher 2015; Grace et al. 2014; Maher, Brady, and Fisher 2013; Macedo and Cardoso 2002; Macedo et al. 2009; Macedo and Cardoso 2001) we argue that surprise is distinct from novelty and value: an outcome can be both novel and valuable, but not necessarily surprising. While surprise is naturally geared and driven by novelty, it stems from a violation of an expectation (Maher, Brady, and Fisher 2013) rather than from a new unseen outcome. Expectation does not necessarily imply novelty; consequently, surprise can be seen as novelty in a temporal space of unseen or expected outcomes (temporal novelty), rather than in a space of existing and already seen outcomes. Studies in cognitive science suggest that humans are not only capable of self-surprise but, most importantly, that surprise is a core internal driver of creativity (Grace and Maher 2015) and a crucial component of general intelligence (Ortony and Partridge 1987). Thus, in our view, surprise constitutes a powerful drive for computational discovery as it incorporates predictions of an expected outcome that it attempts to deviate from. These predictions may be based on relationships in the solution space as well as historical trends derived from the algorithm's sampling of the domain. By modeling surprise, not only do we attempt to advance our knowledge in understanding the phenomenon but — most importantly for the purpose of this paper — we equip artificial creators with capacities to search for surprising outcomes (Macedo et al. 2009).

When it comes to computational search the dominant approach towards obtaining outcomes of high value is to ad-hoc design a function that will reward outcomes with respect to a particular *objective*. An objective function characterizes the value (or quality) of the outcome, and is used in the majority of evolutionary computation studies. However, *divergent* search beyond objectives, such as *novelty*, has proven far more efficient in a number of tasks such as robot navigation (Lehman and Stanley 2011a) and locomotion (Lehman and Stanley 2011b). Similarly, in open-ended evolution studies within artificial life (Channon 2001) it is typical to consider open-ended search for e.g. survival (Yaeger 1994; Adami, Ofria, and Collier 2000) instead of particular objectives. Given the subjective and human-centric nature of creativity, studies within computational creativity and generative systems (Boden 2004; Ritchie 2007; Wiggins 2006) have also focused on the creative capacity of search rather than accomplishing specific objectives.

In this paper we draw inspirations from the above perspectives in computational creativity, divergent search and open ended evolution and we propose the use of *surprise* as a new form of evolutionary divergent search for computational creativity. Our hypothesis is that the search for surprise (i.e. *surprise search*) is beneficial to computational creativity as it complements our search capacities with highly efficient and robust algorithms beyond the search for objectives or mere novelty. As a first step towards testing our hypothesis, we herein introduce the idea of surprise search and propose a general evolutionary algorithm that realizes it. We also provide examples of the surprise search algorithm within the domains of problem solving (for maze navigation) and generative art.

Novelty of this Paper

It is important to note that several studies have already used the notion of surprise for computational modeling (Grace et al. 2014; Grace and Maher 2015; Maher, Brady, and Fisher 2013; Macedo and Cardoso 2002; Macedo et al. 2009; Macedo and Cardoso 2001; Saunders and Gero 2004). These formulations of surprise are similar to ours as they measure a degree of deviation from expected outcomes which are predicted by a model. Macedo and Cardoso (2002; 2009; 2001) performed extensive experiments to test whether the surprise scores derived from their model of surprise match the ones rated by humans under similar circumstances. In other relevant studies, a surprise score has been used to assess the creative capacity of design outcomes (Grace et al. 2014; Maher, Brady, and Fisher 2013). To the best of our knowledge, however, no study utilizes surprise directly as a heuristic within the generative or the creative search process. The model of surprise in our proposed algorithm both drives the computational search for unexpected outcomes and can also be used to evaluate the degree of unexpectedness of an obtained solution, artwork or computational product.

Other aspects of unexpectedness such as intrinsic motivation (Oudeyer, Kaplan, and Hafner 2007; Merrick and Maher 2009) and artificial curiosity (Schmidhuber 2010; Saunders and Gero 2004) have also been modeled in the literature. The concepts of novelty within reinforcement learning research are also interlinked to the idea of surprise search (Kaplan and Hafner 2006; Oudeyer, Kaplan, and Hafner 2007). As a high-level concept, surprise (as described in this paper) matches the notion of Schmidhuber (2010) which rewards new patterns of a growing world model that a curious agent attempts to learn. As an algorithm, however, the search for surprise does not resemble artificial curiosity and intrinsic motivation as it is builds upon evolutionary divergent search and is motivated by open-ended evolution rather than improving a world model.

Surprise, Novelty and Value

In this section we discuss the notion of surprise as a potential form of divergent search for computational creativity which is manifested as both unconventional problem solving and computational art. For that purpose we first draw inspiration from the literature and attempt to define surprise; we then compare it against the notions of novelty and value (Ritchie 2007) which arguably define the most popular criteria of creativity assessment for computational outcomes.

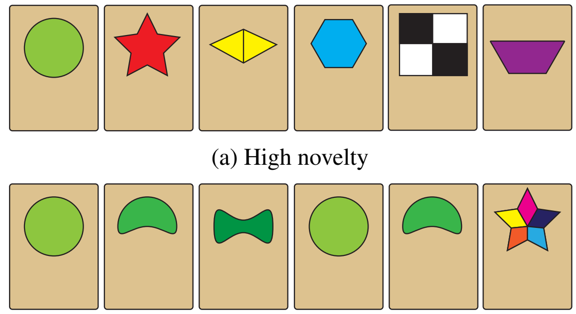
Surprise

The study of surprise has been central in neuroscience (Donchin 1981), psychology (Ekman 1992), and cognitive science (Ortony and Partridge 1987; Kulkarni and Simon 1988). In neurophysiology there has been evidence for the existence of particular event-related brain potentials that can be attributed to unexpected events and, thus, used as predictors of unexpectedness and event memorability (Donchin 1981). In psychology and emotive modeling studies, surprise defines one of Ekman's six basic emotions (Ekman 1992) that can be derived from generic and culture-independent facial expressions. While the facial expression of surprise is generic and manifested similarly across people, surprise is characterized by startle physiological responses. As a result, the classification of surprise as an emotive state has been questioned by several cognitive science studies and instead defined as a temporal-based cognitive process of the unexpected (Meyer, Reisenzein, and Schützwohl 1997; Lorini and Castelfranchi 2007), a violation of a belief (Ortony and Partridge 1987; Kulkarni and Simon 1988), or a reaction to a mismatch (Lorini and Castelfranchi 2007).

Beyond value and novelty, surprise has defined a core criterion for assessing both the creative outcomes and the creative process of a computational creator. Within studies in computational creativity, surprise (along with novelty and value) has been attributed to the creative output of a computational process in creative design and beyond (Grace et al. 2014; Maher 2010; Maher, Fisher, and others 2012; Maher, Brady, and Fisher 2013), defined as a response to novelty (Wiggins 2006) or used to model agent behavior (Macedo and Cardoso 2002; Macedo et al. 2009; Macedo and Cardoso 2001). Computational models of surprise have also been used for traffic control (Horvitz et al. 2005), for detecting surprising features in images (Itti and Baldi 2005), and for detecting interesting experiments for computational scientific discovery (Kulkarni and Simon 1988).

Variant types and taxonomies of surprise have been suggested in the literature. An important distinction is between *active* versus *passive* surprise (Ortony and Partridge 1987; Grace et al. 2014): the first being the explicit expectation that was formed actively prior to a stimulus, the latter being a mere assumption arising from earlier experience. The main overarching element of surprise across any of its taxonomies, however, is the degree to which an observation is expected. Thus, independently of the variant definitions across the disciplines that study surprise as a phenomenon, we can safely derive a general definition of surprise that satisfies the key characteristics of that notion. For the purposes of this paper, we define surprise as the *deviation from the expected* and we use the notions *surprise* and *unexpectedness* interchangeably due to their highly interwoven nature (Reisenzein 2000): unexpectedness being the approximate cognitive appraisal cause of surprise.

Inspired by the relevant literature on surprise, we view surprise for computational search as the degree to which expectations about a solution are violated through observation (Grace et al. 2014). Surprise search acts as a variant divergent search algorithm, similar to novelty search described below. While novelty search diverges from previously and currently seen outcomes, surprise search attempts to deviate from expected but unseen outcomes. Our hypothesis is that if modeled appropriately, surprise may enhance divergent search and complement or even surpass the creative capacity of traditional forms of divergent search such as novelty.



(a) High novelty

(b) High surprise

Figure 1: Illustrating the difference between novelty and surprise in a card game example. The cards are drawn in sequence, with the last one placed rightmost. The rightmost card in Fig. 1a does not share the colors or shapes of the other cards, and is thus novel; however, it is not a surprise that a completely different card is revealed (although admittedly, predicting the next revealed card would be quite difficult). The fourth and the rightmost cards of Fig. 1b are surprising, however. In the first case the player expects a new unseen card based on the three cards already revealed but, instead, the first card reappears. In the latter case, due to the repetition of previous card patterns, the player expects another green, curved shape rather than the depiction in the last card.

Novelty

Novelty and surprise are different notions by definition, as it is possible for a solution to be both novel and/or expected to variant degrees. Following the core principles of Lehman and Stanley (2011a) and Grace et al. (2014), novelty is defined as the degree to which an outcome is *different from prior* outcomes within a particular domain. On the other hand, surprise is the degree to which an outcome is *different from the expected* outcomes in a particular domain.

Expectations are naturally based on inference from past experiences; analogously surprise is built on the temporal model of past outcomes. Surprise is a temporal notion as expectations are temporal by nature. Prior information is required to predict what is expected; hence a *prediction of the expected* (Maher 2010; Macedo and Cardoso 2002) is a necessary component for modeling surprise computationally. By that logic, surprise can be viewed as a *temporal novelty* process. Another interesting temporal metaphor of the relationship between surprise and novelty is that the first can be viewed as the *time derivative* of the latter — e.g. position (novelty) and velocity (surprise). While novelty deviates from positions in the search space, surprise deviates from positions as predicted by a model of earlier positions; the model resembles the trajectory of search.

To exemplify the difference between the notions of novelty and surprise, we will use a simple card memory game. In this game each player is given a stack of unseen cards. Players take turns revealing one card at a time, placing them in a sequence. Players have to predict which card will be revealed next. The winner of the game is the one that correctly predicts the next card. In this example the novelty of the game outcome (i.e. next card) is the highest possible if all cards revealed in the past are different. The surprise value of the game outcome in that case is low as the player has gradually internalized a model of expectedness of a new, unseen, card every time. On the other hand, the novelty of the game outcome decreases if seen cards are revealed after a while. In that scenario surprise is increased as the game deviates from the expected outcome which calls for a new card every time. Clearly both surprise and novelty depend on the amount of cards revealed (i.e. history of instances) and the amount of cards available (i.e. how large is the domain). Figure 1 illustrates the difference between the notions of novelty and surprise in the card game example discussed above.

As a guide for evolutionary search, the concept of novelty has primarily been integrated in *novelty search*, which explicitly ignores the objective (or value) of the problem it attempts to solve. Novelty search performs divergent evolutionary search in order to handle deceptive fitness landscapes (Whitley 1991) and premature convergence to local optima. Earlier divergent search methods (e.g. (Angeline and Pollack 1994; Wessing, Preuss, and Rudolph 2013)) provide control mechanisms, modifiers or alternate objectives which complement the gradient search towards better solutions. In contrast, novelty search motivates exploration of the search space by rewarding individuals which are different without considering whether they are objectively 'better' than others. Novelty search is different than a random walk, however, as it explicitly provides higher rewards to more diverse solutions and also because it maintains a memory of the areas of the search space that it has previously explored. The latter is achieved with a *novel archive* of past novel individuals, with individuals with a high novelty score being constantly added to this archive. Each individual's novelty score is the average distance from a number of closest neighbors in the problem space; neighbors can be members of the current population or the novel archive. The distance measure is problem-dependent: examples include the distance between agents' final positions in a two-dimensional maze, or the distance in the position of a robot's center of mass (Lehman and Stanley 2011a). Novelty search has also been integrated for adjusting properties of images such as brightness and symmetry (Lehman and Stanley 2012).

Value

Value has been defined as the degree to which a generated outcome is of high quality within its domain (Ritchie 2007). While in computational art and aesthetics value is largely a subjective notion — which is often measured via the valued ratings of domain experts — in creative problem solving the notion of value is clearly objective. In particular, the quality of any solution or output is determined by its distance to a predetermined goal within a set of constrains imposed by the domain per se. The notion of value can be directly linked to the notion of *objective* in optimization. More specifically, within metaheuristic search the value of an evolved solution is naturally assessed by its fitness value to a given problem. While it is natural to think that measuring progress in terms of fitness (Goldberg and Holland 1988; Michalski, Carbonell, and Mitchell 2013) is the most appropriate approach towards finding a high-fit solution, recent findings from evolutionary divergent search (Lehman and Stanley 2011a; Lehman, Stanley, and Miikkulainen 2013) suggest that explicit objective (fitness) design can be detrimental to evolutionary search, e.g. when the problem is deceptive (Whitley 1991) or open-ended (e.g. in the case of autotelic creative tasks).

While as concepts surprise and novelty have common characteristics (see earlier discussion), value can be seen as an orthogonal concept in the search for good quality outcomes. Value clearly distinguishes from novelty and surprise as it is the degree of outcome quality rather than the degree to which an outcome differs from other outcomes (novelty) or the degree to which an outcome differs from expected outcomes (surprise) in its class. Value, if used as a direction for search, points to a direct assessment of the outcome's quality whereas both novelty and surprise imply an indirect and divergent way of traversing the search space for obtaining an outcome of high quality.

Since value is orthogonal to novelty or surprise, there are ways of integrating it in divergent search e.g. via constraints that accepted artifacts should have a minimum value (i.e. fitness score) while individuals satisfying these constraints can evolve towards divergence. Examples of constrained novelty search, in particular, have been proposed by Liapis, Yannakakis, and Togelius (2015) and Lehman and Stanley (2010) for problem solving tasks (level generation and maze navigation respectively), as well as Vinhas et al. (2016) for evolutionary art and Liapis et al. (2013) for novel game object generation. Surprise search can similarly be combined with minimal value constraints, since feasible and infeasible individuals can be evolved in separate populations (Liapis, Yannakakis, and Togelius 2015) towards different goals.

The Surprise Search Algorithm

Based on the above discussion, *surprise* as a driver of evolutionary search can be summarized as a mechanism for rewarding individuals which exhibit behaviors which diverge from the *expected behaviors* of the current population based on *prior observed behaviors*. Like novelty search, surprise search operates exclusively in the behavioral (or phenotypic) space[1]: both predictions and prior evolutionary trends refer to the phenotypic space (e.g. behaviors of an evolving agent or output of an artificial painter).

Two components are therefore necessary for surprise search: a *predictive model* which creates the expected behaviors based on past and current outputs, and a *deviation formula* which assesses whether (and to what degree) the actual behaviors deviate from the predicted behaviors. To a certain extent, both of these components are domain-specific and problem-dependent; this section presents certain core properties of each component, while the specific parameters can be tweaked depending on the problem at hand.

Predictive Model

There are multiple ways to predict future outcomes, from simple extrapolation to machine learning. At its core, the set of predicted outcomes $\mathbf{p}$ is derived from the formula in Eq. (1), where m is the predictive model, h is the history (i.e. how far in the past the model has to look to estimate the future) and k is the locality (i.e. how many data points the model has to consider per generation and, as a result, how many predictions it must make).

$$\mathbf{p} = m(h, k) \tag{1}$$

History (h) refers to how far into the past the predictive model observes when making predictions into the future. At the absolute minimum, the two previous generations must be considered, in order to assess a degree of behavioral (outcome) change which can be expected in the current generation. Earlier information can also be used, by looking at previous generations further in the past, or by considering an archive of important past predictions. The latter concept is similar to the rationale of the novel archive (Lehman and Stanley 2011a) in novelty search, where the most novel individuals from past generations are stored. Deviation from behaviors expected currently can be viewed as a form of *passive* surprise (Ortony and Partridge 1987; Grace et al. 2014) as they are assumptions which have not been actively considered. Deviation from a surprise archive can be viewed as a form of *active* surprise (Ortony and Partridge 1987) in that predictions in the archive have been "entertained" (to use the term of Ortony and Partridge) in the past. We consider h to be a problem-dependent parameter for the algorithm.

[1]Behavior and phenotype of e.g. an artificial evolutionary process are terms used interchangeably in this paper.

Locality (k) refers to the granularity in which the trends of past populations are observed. Locality can stretch from global (i.e. each generation predicts a single descriptive feature of the population of size P in the current generation) to individual (i.e. each individual traces its own lineage of parents, grandparents etc. and attempts to surprise itself). A parameter k determines the level of *prediction locality* which can vary from 1 (individual) to P (global) or anything in-between. The level of prediction locality (k) in the outcome space is a problem-dependent parameter that can be derived empirically.

Predictive model (m) refers to a model which can calculate a future outcome from current and past data as collected based on k and h. As noted above, any modeling approach can be used for such purposes: from a simple linear regression of points in the outcome space, to non-linear extrapolations, or machine learned models (e.g. artificial neural networks or support vector machines). Depending on the locality of the prediction (k), the model may derive a vector of expected outcomes to deviate from. Again, we consider the employed predictive model, m, to be problem-dependent.

Deviation Formula

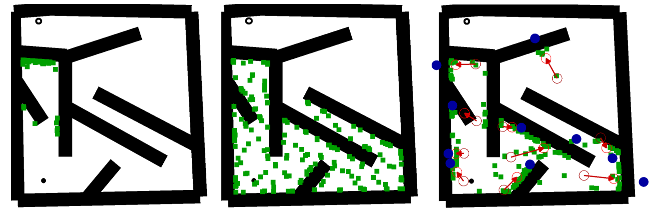
We are primarily inspired by the calculation of novelty (Lehman and Stanley 2011a) in the design of a deviation formula for surprise. The formula in Eq. (2) calculates surprise as the average distance of the n closest predictions made using the predictive model p. The formula assumes that the more divergent an observed behavior is from predicted behaviors, the more surprising it is. Just like with novelty search, the distance metric (d_s) is domain-dependent and can affect what is considered surprising (and therefore the evolutionary search itself). The examples in the next section demonstrate optimization for divergence from the expected in maze navigation and generative art tasks; the same examples show different ways of calculating dissimilarity.

$$s(i) = \frac{1}{n} \sum_{j=0}^{n} d_s(i, p_{i,j}) \qquad (2)$$

It should be noted that the deviation formula is purposefully simple, as it is an intuitive way in which humans consider divergence. However, there is potential in exploring different formulas, e.g. so that results which are not too similar but yet not too dissimilar are prioritized versus too dissimilar outputs which can be perceived as atypical, alien or random (Grace et al. 2014). This can be achieved in the distance function itself, or by applying a normalizing function on $d_s(i, p_{i,j})$ (e.g. a Gaussian function). Following the surprise model of Itti and Baldi (2005), d_s can alternatively be formulated as the difference between posterior and prior beliefs of an observer.

Examples of Surprise Search

To illustrate how surprise search can work (and ways in which it differs from search for value or novelty), this paper uses two test beds: a maze navigation task, and a generative art activity. These two exemplar tasks are tackled by



(a) Objective (b) Novelty (c) Surprise

Figure 2: The process of different types of search for maze navigation on the "hard" maze of Lehman and Stanley (2011a). The solid black circle (bottom left) is the starting position, and the empty black circle (top left) is the goal; green dots represent the agents. For surprise search (Fig. 2c), red circles represent the population's centroids of the two prior generations; the red arrow is the direction of the centroid, from the earliest generation to the latest generation, while the blue circles are the predictions for the clusters' centroids in the current generation. Surprise search rewards the individuals in the current population (green) which deviate from the closest prediction point (blue).

evolution and constitute computational creativity domains. On one hand maze navigation focuses on *problem solving*: there is an end-state and a clear performance measure, which is whether the maze has been solved. For that purpose, we use the maze example of Lehman and Stanley (2011a) for its appropriateness in testing novelty due to the deceptiveness of the problem. On the other hand, a generative art task represents *autotelic creativity*, where there is no stopping condition and arguably no measure of success (Compton and Mateas 2015). Indeed, defining an objective for generative art would be subjective and ultimately ad-hoc; therefore we do not focus on objective-driven search for this task. We briefly present the problems, the representations used, and how the different strategies (objective-driven, novelty search, and surprise search) explore the space of each problem's possible solutions.

Maze Navigation

In the maze navigation task, an agent controlled by an artificial neural network (ANN) enters a maze enclosed by walls at the start position of the maze, and has a specific timeframe (i.e. simulation steps) to find the goal position of the maze (see Fig. 2). The agent has 6 line trace sensors along its perimeter, measuring the distance to the nearest wall, and 4 "radars" which inform it on which side of the agent the goal is (if within range). These 10 inputs, along with a bias, are used as input to an ANN, which outputs the change in speed and the change in direction of the agent (2 outputs). The ANN is evolved using neuroevolution of augmenting topologies (NEAT) which adds complexity to initially simple networks during the course of evolution (Stanley and Miikkulainen 2002). A population of 500 ANN-controlled agents is tested in every generation: their final position at the end of the allocated timeframe in one generation can be seen in Fig. 2. This paper discusses the general principles of the search process, highlighting the differences between surprise, novelty and objective search; Gravina et al. (2016) provide an in-depth analysis of the differences in efficiency and robustness between the three algorithms in the maze navigation task.

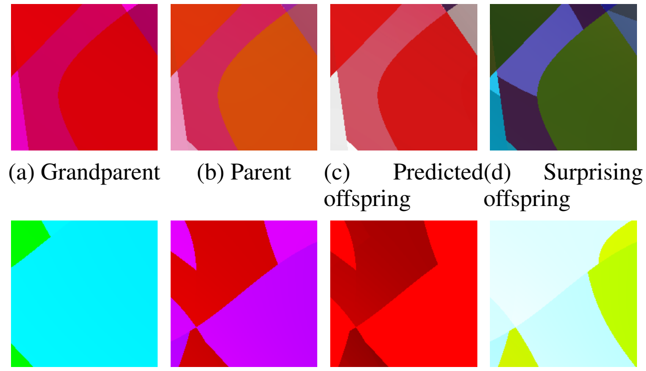
When using objective-driven search, Lehman and Stanley (2011a) calculate the fitness of each individual based on how close it is to the goal. This is a "reasonable" performance metric, which however falls short as it does not consider walls and can cause individuals to get stuck in the dead-end at the top left corner of the maze in Fig. 2a, which acts as a local optimum. In order to find the global optimum and solve the maze, the agents must explore areas of the maze with the lowest fitness (i.e. the bottom right corner); therein lies the deceptiveness of the problem under the current objective function.

When using novelty search, Lehman and Stanley (2011a) calculate the fitness (or *novelty score*) of each individual based on its final position, and its distance from the closest final positions of other individuals in the population or in a novel archive. This drives individuals to explore more of the space, and separate themselves from current and past discovered locations. This process eventually pushes some of the most novel individuals to the goal (see Fig. 2b).

When using surprise search, instead, we suggest grouping individuals into k clusters; each predicted point is calculated based on the linear interpolation of a cluster in the population of the two previous generations (see Fig. 2c). In other words, the prediction locality of surprise search in this problem is determined by the number of clusters (k) chosen (k is 10 in this example), h involves two subsequent generations of the population and m is a linear regression function. The fitness (or *surprise score*; s in Eq. (2)) of individuals in the current population is calculated based on the deviation (d_s is Euclidean distance) of each individual from the closest predicted point. This rewards agents who diverge from an expected behavior. Note that while novelty search deviates from points of the maze that have been previously explored, surprise search deviates from predicted points which may have not been reached yet by the agents, or may never will (such as points outside of the maze in Fig. 2c).

Generative Art

In the generative art example, colorful images are generated via evolving compositional pattern-producing neural networks (CPPNs). These neural networks can have different activation functions (e.g. sigmoid or Gaussian curves) which produce symmetries and repetitions in the output (Stanley 2006). In that regard, the CPPN-based artwork is similar to the output of *PicBreeder* (Secretan et al. 2011), although this example uses a simplified representation and evolutionary strategy. Each pixel in the colored image is represented as a HSB (hue, saturation, brightness) triplet, and the CPPN produces these three output values using the x, y coordinates of the pixel as input. Fig. 3 shows how an outcome (Fig. 3a) produces a mutated offspring in the next generation (Fig. 3b). In this example, we predict one expected outcome per individual in the population, taking into account their own parent (i.e. we predict based on genotypic history, rather than phenotypic similarity as in the maze example). The predicted outcome is based on the differences in HSB values between the parent and grandparent ($h = 2$) of the evaluated individual, applied on the parent. Similarly to the maze navigation surprise search setup, m is a simple linear regression model and $h = 2$; however, the k value considered here is P (where P is the size of the population), as each image has an individual prediction. Fig 3c provides an example prediction: as the bottom left area becomes lighter (Fig. 3b) compared to the earlier image (Fig. 3a), it is predicted that the same area will become even brighter in the current generation. Another example is in Fig 3g: as the entire canvas from the grandparent (Fig. 3e) to the parent (Fig. 3f) shifts the hue towards warmer colors, it is predicted that the offspring canvas will consist entirely of red colors; moreover, the left-most darker region in the parent's canvas is expected to become even darker in the offspring's canvas.


(a) Grandparent (b) Parent (c) Predicted offspring (d) Surprising offspring

(e) Grandparent (f) Parent (g) Predicted offspring (h) Surprising offspring

Figure 3: Two examples of surprise search in generative art, illustrating how surprise and predictions can be computed. Images depict potential outcomes of the process.

A surprising outcome, in these examples, can be measured based on the per-pixel difference (e.g. d_s can be the Euclidean distance) in hue, saturation and brightness between the predicted outcome and the actual offspring. For instance, Fig. 3d has a high surprise score since the hues of the entire image have shifted to greens and blues, and the image is overall darker. Similarly, Fig. 3h is surprising since the image is overall lighter and shifted hues towards colder colors. It should be noted that Fig. 3h is similar to Fig. 3e; as with the card game example, the novelty score of such an image would not necessarily be high as the previous outcome (which likely is stored in the novel archive) is visually close to the evaluated outcome. The surprise score, on the other hand, takes into account the "trajectory" of evolution and predicts the expected outcome assuming a direction from previous to current outcomes.

Discussion and Conclusions

This paper introduced the notion of surprise for computational search, provided a general algorithm that follows the principles of searching for surprise and presented two examples implementing the core idea of *deviation from expected*: a maze navigation problem and a generative visual art task. We argue that surprise search may show advantages over other forms of evolutionary divergent search such as novelty search. Based on the advantages of novelty over objective search, we can safely assume that a divergent search-based algorithm like surprise will manage to outperform traditional fitness-based evolution (i.e. objective search) in highly deceptive problems. Our hypothesis is that, similarly to novelty search, deviation from expected outcomes in the search space may result in higher exploratory capacity and diversity; both of which are beneficial properties for computational (evolutionary) search.

Surprise search operates similarly to novelty search with respect to evolutionary dynamics. As surprise search makes predictions for the current generation based on a set of observed behaviors in prior generations, it maintains a temporal window of where search has been. However, surprise search operates differently to novelty search with respect to the goal: surprise maximizes deviation from the expected outcomes whereas novelty maximizes deviation from previous and current outcomes. This evidently creates a new form of divergent search that considers prior behaviors *indirectly* to make predictions to deviate from. The comparative envisaged advantages of surprise search over other forms of divergent search are inherent to the way the algorithm searches, attempting to deviate from predicted *unseen* behaviors instead of prior *seen* behaviors.

As surprise search ignores objectives, a concern could be whether it is merely a version of random walk. Surprise search is not a random walk as it explicitly maximizes unexpectedness. Surprise search allows for a temporal archive of outcomes that accumulates a record of earlier positions in the problem space. Gravina, Liapis, and Yannakakis (2016) compared the behavior of surprise search versus random search (random fitness values) in the maze navigation experiment, demonstrating the differences in both performance and behavior between the two.

This position paper introduced a general form of the surprise search algorithm and examples of its implementation; extensive empirical studies need to be performed to provide evidence for the advantages of surprise as a form of divergent search. The two examples used in this paper are indicative types of test beds for surprise search. For problem solving tasks (such as maze navigation), the algorithm's effectiveness needs to be tested through tasks of varying degrees of deception and complexity. Initial experiments with surprise search in the maze navigation domain indicate that it is as efficient as novelty search, and tends to find solutions faster and more often than both traditional objective and novelty search (Gravina, Liapis, and Yannakakis 2016). For computational art, the algorithm's expressivity and creative capacity can be assessed, based on its ability to deviate from expected outcomes or based on creativity assessment models such as FACE or IDEA (Pease and Colton 2011). Finally, the surprise score introduced can be used to complement any computational creativity assessment method considered.

Acknowledgments

This work has been supported in part by the FP7 Marie Curie CIG project AutoGameDesign (project no: 630665).

References

Adami, C.; Ofria, C.; and Collier, T. C. 2000. Evolution of biological complexity. *Proceedings of the National Academy of Sciences* 97(9).

Angeline, P. J., and Pollack, J. B. 1994. Competitive environments evolve better solutions for complex tasks. In *Proceedings of the International Conference on Genetic Algorithms*.

Boden, M. 1995. Creativity and unpredictability. *Constructions of the Mind: Artificial Intelligence and the Humanities. Stanford Electronic Humanities Review* 4(2).

Boden, M. A. 2004. *The Creative Mind: Myths and Mechanisms*. Routledge.

Channon, A. 2001. Passing the alife test: Activity statistics classify evolution in geb as unbounded. In *Advances in Artificial Life*. Springer.

Compton, K., and Mateas, M. 2015. Casual creators. In *Proceedings of the International Conference on Computational Creativity*.

Donchin, E. 1981. Surprise! surprise? *Psychophysiology* 18(5):493–513.

Ekman, P. 1992. An argument for basic emotions. *Cognition & emotion* 6(3-4).

Goldberg, D. E., and Holland, J. H. 1988. Genetic algorithms and machine learning. *Machine learning* 3(2).

Grace, K., and Maher, M. L. 2015. Specific curiosity as a cause and consequence of transformational creativity. *Proceedings of the International Conference on Computational Creativity June*.

Grace, K.; Maher, M. L.; Fisher, D.; and Brady, K. 2014. Modeling expectation for evaluating surprise in design creativity. In *Design Computing and Cognition*.

Gravina, D.; Liapis, A.; and Yannakakis, G. N. 2016. Surprise search: Beyond objectives and novelty. In *Proceedings of the Genetic and Evolutionary Computation Conference*. ACM.

Horvitz, E. J.; Apacible, J.; Sarin, R.; and Liao, L. 2005. Prediction, expectation, and surprise: Methods, designs, and study of a deployed traffic forecasting service. In *Proceedings of the 2005 Conference on Uncertainty and Artificial Intelligence. AUAI Press*.

Itti, L., and Baldi, P. F. 2005. Bayesian surprise attracts human attention. In *Advances in neural information processing systems*, 547–554.

Kaplan, F., and Hafner, V. V. 2006. Information-theoretic framework for unsupervised activity classification. *Advanced Robotics* 20(10).

Kulkarni, D., and Simon, H. A. 1988. The processes of scientific discovery: The strategy of experimentation. *Cognitive science* 12(2):139–175.

Lehman, J., and Stanley, K. O. 2010. Revising the evolutionary computation abstraction: Minimal criteria novelty search. In *Proceedings of the Genetic and Evolutionary Computation Conference*.

Lehman, J., and Stanley, K. O. 2011a. Abandoning objectives: Evolution through the search for novelty alone. *Evolutionary computation* 19(2).

Lehman, J., and Stanley, K. O. 2011b. Evolving a diversity of virtual creatures through novelty search and local competition. In *Proceedings of the Genetic and Evolutionary Computation Conference*.

Lehman, J., and Stanley, K. O. 2012. Beyond open-endedness: Quantifying impressiveness. In *Proceedings of the International Conference on Artificial Life*.

Lehman, J.; Stanley, K. O.; and Miikkulainen, R. 2013. Effective diversity maintenance in deceptive domains. In *Proceedings of the Genetic and Evolutionary Computation Conference*.

Liapis, A.; Martínez, H. P.; Togelius, J.; and Yannakakis, G. N. 2013. Transforming exploratory creativity with DeLeNoX. In *Proceedings of the International Conference on Computational Creativity*.

Liapis, A.; Yannakakis, G. N.; and Togelius, J. 2015. Constrained novelty search: A study on game content generation. *Evolutionary Computation* 23(1):101–129.

Lorini, E., and Castelfranchi, C. 2007. The cognitive structure of surprise: looking for basic principles. *Topoi* 26(1).

Macedo, L., and Cardoso, A. 2001. Modeling forms of surprise in an artificial agent. In *Proceedings of the nnual Conference of the Cognitive Science Society*.

Macedo, L., and Cardoso, A. 2002. Assessing creativity: the importance of unexpected novelty. *Structure* 1(C2):C3.

Macedo, L.; Cardoso, A.; Reisenzein, R.; Lorini, E.; and Castelfranchi, C. 2009. Artificial surprise. *Handbook of research on synthetic emotions and sociable robotics: New applications in affective computing and artificial intelligence* 267–291.

Maher, M. L.; Brady, K.; and Fisher, D. H. 2013. Computational models of surprise in evaluating creative design. In *Proceedings of the fourth international conference on computational creativity*, volume 147. Citeseer.

Maher, M. L.; Fisher, D. H.; et al. 2012. Using ai to evaluate creative designs. In *2nd international conference on design creativity, Glasgow, UK*, 45–54.

Maher, M. L. 2010. Evaluating creativity in humans, computers, and collectively intelligent systems. In *Proceedings of the 1st DESIRE Network Conference on Creativity and Innovation in Design*.

Merrick, K. E., and Maher, M. L. 2009. *Motivated reinforcement learning: curious characters for multiuser games*. Springer Science & Business Media.

Meyer, W.-U.; Reisenzein, R.; and Schützwohl, A. 1997. Toward a process analysis of emotions: The case of surprise. *Motivation and Emotion* 21(3).

Michalski, R. S.; Carbonell, J. G.; and Mitchell, T. M. 2013. *Machine learning: An artificial intelligence approach*. Springer Science & Business Media.

Ortony, A., and Partridge, D. 1987. Surprisingness and expectation failure: what's the difference? In *Proceedings of the 10th international joint conference on Artificial intelligence-Volume 1*, 106–108. Morgan Kaufmann Publishers Inc.

Oudeyer, P.-Y.; Kaplan, F.; and Hafner, V. V. 2007. Intrinsic motivation systems for autonomous mental development. *IEEE Transactions on Evolutionary Computation* 11(2).

Pease, A., and Colton, S. 2011. Computational creativity theory: Inspirations behind the face and the idea models. In *Proceedings of the Second International Conference on Computational Creativity*.

Reisenzein, R. 2000. The subjective experience of surprise. *The message within: The role of subjective experience in social cognition and behavior* 262–279.

Ritchie, G. 2007. Some empirical criteria for attributing creativity to a computer program. *Minds and Machines* 17(1).

Saunders, R., and Gero, J. S. 2004. Curious agents and situated design evaluations. *AI EDAM: Artificial Intelligence for Engineering Design, Analysis and Manufacturing* 18(02):153–161.

Schmidhuber, J. 2010. Formal theory of creativity, fun, and intrinsic motivation (1990–2010). *IEEE Transactions on Autonomous Mental Development* 2(3).

Secretan, J.; Beato, N.; D'Ambrosio, D. B.; Rodriguez, A.; Campbell, A.; Folsom-Kovarik, J. T.; and Stanley, K. O. 2011. Picbreeder: A case study in collaborative evolutionary exploration of design space. *Evolutionary Computation* 19(3):373–403.

Stanley, K. O., and Miikkulainen, R. 2002. Evolving neural networks through augmenting topologies. *Evolutionary Computation* 10(2).

Stanley, K. O. 2006. Exploiting regularity without development. In *Proceedings of the 2006 AAAI Fall Symposium on Developmental Systems*.

Vinhas, A.; Assuncao, F.; Correia, J.; Machado, P.; and Ekárt, A. 2016. Fitness and novelty in evolutionary art. In *Proceedings of Evolutionary and Biologically Inspired Music, Sound, Art and Design (EvoMusArt)*. Springer.

Wessing, S.; Preuss, M.; and Rudolph, G. 2013. Niching by multiobjectivization with neighbor information: Trade-offs and benefits. In *Proceedings of the Evolutionary Computation Congress*.

Whitley, L. D. 1991. Fundamental principles of deception in genetic search. In *Foundations of Genetic Algorithms*. Morgan Kaufmann.

Wiggins, G. A. 2006. A preliminary framework for description, analysis and comparison of creative systems. *Knowledge-Based Systems* 19(7).

Yaeger, L. 1994. Poly world: Life in a new context. *Proc. Artificial Life* 3.