

What to expect when you're expecting: The role of unexpectedness in computationally evaluating creativity

Kazjon Grace and Mary Lou Maher

{k.grace,m.maher}@uncc.edu

The University of North Carolina at Charlotte

Abstract

Novelty, surprise and transformation of the domain have each been raised – alone or in combination – as accompaniments to value in the determination of creativity. Spirited debate has surrounded the role of each factor and their relationships to each other. This paper suggests a way by which these three notions can be compared and contrasted within a single conceptual framework, by describing each as a kind of *unexpectedness*. Using this framing we argue that current computational models of novelty, concerned primarily with the *originality* of an artefact, are insufficiently broad to capture creativity, and that other kinds of expectation – whatever the terminology used to refer to them – should also be considered. We develop a typology of expectations relevant to computational creativity evaluation and, through it describe a series of situations where expectations would be essential to the characterisation of creativity.

Introduction

The field of computational creativity, perhaps like all emergent disciplines, has been characterised throughout its existence by divergent, competing theoretical frameworks. The core contention – unsurprisingly – surrounds the nature of creativity itself. A spirited debate has coloured the last several years' conferences concerning the role of *surprise* in computational models of creativity evaluation. [Feyerabend \(1963\)](#) argued that scientific disciplines will by their nature develop incompatible theories, and that this theoretical pluralism beneficially encourages introspection, competition and defensibility. We do not go so far as to suggest epistemological anarchy as the answer, but in that pluralistic mindset this paper seeks to reframe the debate, not quell it.

We present a way by which three divergent perspectives on the creativity of artefacts can be placed into a unifying context¹. The three perspectives on evaluating creativity are that, in addition to being valuable, 1) creative artefacts are novel, 2) creative artefacts are surprising, or 3) creative artefacts transform the domain in which they reside. We propose that these approaches can be reconceptualised to all derive from the notion of *expectation*, and thus be situated within a framework illustrating their commonalities and differences.

Creativity has often been referred to as the union of novelty and value, an operationalisation first articulated (at least to the authors' knowledge) in [Newell, Shaw, and Simon \(1959\)](#). Computational models of novelty (eg. [Berlyne, 1966, 1970](#);

[Bishop, 1994](#); [Saunders and Gero, 2001b](#)) have been developed to measure the *originality* of an artefact relative to what has come before. Newell and others (eg. [Abra, 1988](#)) describe novelty as necessary but insufficient for creativity, forming one half of the novelty/value dyad.

Two additional criteria have been offered as an extension of that dyad: surprisingness and transformational creativity. Surprise has been suggested as a critical part of computational creativity evaluation because computational models of novelty do not capture the interdependency and temporality of experiencing creativity ([Macedo and Cardoso, 2001](#); [Maher, 2010](#); [Maher and Fisher, 2012](#)), but has also been considered unnecessary in creativity evaluation because it is merely an observer's response to experiencing novelty ([Wiggins, 2006b](#)). Boden's transformational creativity ([Boden, 2003](#)) (operationalised in [Wiggins, 2006a](#)) has been offered as an alternative by which creativity may be recognised. In both cases the addition is motivated by the insufficiency of originality – the comparison of an artefact to other artefacts within the same domain – as the sole accompaniment to value in the judgement of creativity.

Thus far these three notions – novelty, surprise and transformativity – have been considered largely incomparable, describing different parts of what makes up creativity. There has been some abstract exploration of connections between the two – such as Boden's (2003) connection of “fundamental” novelty to transformative creativity – but no concrete unifying framework. This paper seeks to establish that there is a common thread amongst these opposing camps: *expectations* play a role in not just surprise but novelty and transformativity as well.

The foundation of our conceptual reframing is that the notions can be reframed thusly:

- *Novelty* can be reconceptualised as occurring when an observer's expectations about the continuity of a domain are violated.
- *Surprise* occurs in response to the violation of a confident expectation.
- *Transformational creativity* occurs as a collective reaction to an observation that was unexpected to participants in a domain.

We will expand on these definitions through this paper. Through this reframing we argue that *unexpectedness is involved in novelty, surprise and domain transformation, and is thus a vital component of computational creativity evaluation*. The matter of where in our field's pluralistic and still-emerging theoretical underpinnings the notion of unexpectedness should reside is – for now – one of terminology

¹Creative processes are another matter entirely, one beyond the scope of this paper.

alone. This paper sidesteps the issue of whether expectation should primarily be considered the stimulus for surprise, a component of novelty, or a catalyst for transformative creativity. We discuss the connections between the three notions, describe the role of expectation in each, and present an exploratory typology of the ways unexpectedness can be involved in creativity evaluation.

We do not seek to state that novelty and transformativity should be subsumed within the notion of surprise due to their nature as expectation-based processes. Instead we argue that the notions of novelty, surprise and transformativity are all related by another process – expectation – the role of which we yet know little. We as a field have been grasping at the trunk and tail of the proverbial poorly-lit pachyderm, and we suggest that expectation might let us better face the beast.

The eye of the beholder

Placing expectation at the centre of computational creativity evaluation involves a fundamental shift away from comparing artefacts to artefacts. Modelling unexpectedness involves comparing the reactions of observers of those artefacts to the reactions of other observers. This reimagines what makes a creative artefact different, focussing not on objective comparisons but on subjective perceptions. This “eye of the beholder” approach framing is compatible with formulations of creativity that focus not on artefacts but on their artificers and the society and cultures they inhabit (Csikszentmihalyi, 1988). It should be noted that no assumptions are made about the nature of the observing agent – it may be the artefact’s creator or not, it may be a participant in the domain or not, and it may be human or artificial.

The observer-centric view of creativity permits a much richer notion of what makes an artefact different: it might relate to the subversion of established power structures (Florida, 2012), the destruction of established processes (Schumpeter, 1942), or the transgression of established rules (Dudek, 1993; Strzalecki, 2000). These kinds of cultural impacts are as much part of an artefact’s creativity as its literal originality, and we focus on expectation as an early step towards their computationally realisation.

The notion of transformational creativity (Boden, 2003) partially addresses this need by the assumption that cultural knowledge is embedded in the definition of the conceptual space, but to begin computationally capturing these notions in our models of evaluation we must be aware of how narrowly we define our conceptual spaces. The notion common to each of subversion, destruction and transgression is that *expectations about the artefact are socio-culturally grounded*. In other words, we must consider not just how an artefact is described, but its place in the complex network of past experiences that have shaped the observing agent’s perception of the creative domain. A creative artefact is *unexpected* relative to the rules of the creative domain in which it resides. To unravel these notions and permit their operationalisation in computational creativity evaluation we focus not on novelty, surprise or transformativity alone but on the element common to them all: *the violation of an observer’s expectations*.

Novelty as expectation

Runco (2010) documents multiple definitions of creativity that give novelty a central focus, and notes that it is one of the only aspects used to define creativity that has been widely adopted. Models of novelty, unlike models of surprise, are not typically conceived of as requiring expectation. We

argue that novelty can be described using the mechanism of expectation, and that doing so is illuminative when comparing novelty to other proposed factors.

Novelty can be considered to be expectation-based if the knowledge structures acquired to evaluate novelty are thought of as a model with which the system attempts to predict the world. While these structures (typically acquired via some kind of online unsupervised learning system) are not being built for the purpose of prediction, they represent assumptions about how the underlying domain can be organised. Applying those models to future observations within the domain is akin to expecting that those assumptions about domain organisation will continue to hold, and that observations in the future can be described using knowledge gained from observations in the past. The *expectation of continuity* is the theoretical underpinning of computational novelty evaluation, and can be considered the simplest possible creativity-relevant expectation.

Within the literature the lines between novelty and surprise are not always clear-cut, a conflation we see as evidence of the underlying role of expectation in both. Novelty in the Creative Product Semantic Scale (O’Quin and Besemer, 1989), a creativity measurement index developed in cognitive psychology, is defined as the union of “originality” and “unexpectedness”. The model of interestingness in Silberschatz and Tuzhilin (1995) is based on improbability with respect to confidently held beliefs. The model of novelty in Schmidhuber (2010) is based on the impact of observations on a predictive model, which some computational creativity researchers would label a model of transformativity, while others would label a model of surprise. Each of these definitions suggests a complex relationship that goes beyond the notion of originality as captured by simple artefact-to-artefact comparisons.

Surprise as expectation

Many models of surprise involve the observation of unexpected events (Ortony and Partridge, 1987). In our previous work we give a definition of surprise as *the violation of a confidently-held expectation* (Maher and Fisher, 2012; Grace et al., 2014a), a definition derived from earlier computational models both within the domain of creativity (Macedo and Cardoso, 2001) and elsewhere (Ortony and Partridge, 1987; Peters, 1998; Horvitz et al., 2012; Itti and Baldi, 2005).

Models of surprise have previously looked at a variety of different kinds of expectation: predicting trends within a domain (Maher and Fisher, 2012), predicting the class of an artefact from its features (Macedo and Cardoso, 2001) or the effect on the data structures of a system when exposed to a new piece of information (Baldi and Itti, 2010). The first case concerns predicting attributes over time, and involves an expectation of continuity of trends within data, the second case concerns predicting attributes relative to a classification, and is an expectation of continuity of the relationships within data, and the third case concerns the size of the change in a predictive mechanism, and is based on an expectation of continuity, but measured by the post-observation change rather than the prediction error. In each of these cases it is clear that a related but distinct expectation is central to the judgement of surprisingness, but as of yet no comprehensive typology of the kinds of expectation relevant to creativity evaluation exists. The *expectations of continuity* that typically make up novelty evaluation can be extended to cover the above cases. This paper investigates the kinds of expectation that are relevant to creativity evaluation independent of whether they

are an operationalisation of surprise or some other notion.

Transformativity as expectation

Boden’s *transformational creativity* can be reconceptualised as unexpectedness. We develop a notion of transformativity grounded in *an observer’s expectations that their predictive model of a creative domain is accurate*. This requires a reformulation of transformation to be *subjective* to an observer – Boden wrote of the transformation of a domain, but we are concerned with the transformation of an observer’s knowledge about a domain. To demonstrate the role of expectation in this subjective transformativity, we consider the operationalisation of Boden’s transformative creativity proposed by Wiggins (2006b,a), and extend it to the context of two creative systems rather than one.

One system, the *creator*, produces an artefact and chooses to share it with the second creative system, the *critic*. For the purposes of this discussion we investigate how the critic evaluates the object and judges it transformative. In Wiggins’ formalisation the conceptual space is defined by two sets of rules: \mathcal{R} , the set of rules that define the boundaries of the conceptual space, and \mathcal{T} , the set of rules that define the traversal strategy for that space. Wiggins uses this distinction to separate Boden’s notion of transformational creativity into \mathcal{R} -*transformational*, occurring when a creative system’s rules for bounding a creative domain’s conceptual space are changed, and \mathcal{T} -*transformational*, occurring when a creative system’s rules for searching a creative domain’s conceptual space are changed. In the case of our critic it is the set \mathcal{R} that we are concerned with – the critic does not traverse the conceptual space to generate new designs, it evaluates the designs of the creator.

Once we assume the presence of more than one creative agent then \mathcal{R} , the set of rules bounding the conceptual space, cannot be ontological in nature – it cannot be immediately and psychically shared between all creative systems present whenever changes occur. \mathcal{R} must be mutable to permit transformation and individual to permit situations where *critic* and *creator* have divergent notions of the domain. Divergence is not an unusual case: If a transformational artefact is produced by *creator* and judged \mathcal{R} -*transformational* by it, and then shared with *critic*, there must by necessity be a period between the two evaluations where the two systems have divergent \mathcal{R} – even with only two systems that share all designs. With more systems present, or when creative systems only share selectively, divergence will be greater. To whom, then, is such creativity transformational?

To reflect the differing sets belonging to the two agents we refer to \mathcal{R} as it applies to the two agents as *critic $_{\mathcal{R}}$* and *creator $_{\mathcal{R}}$* . If a new artefact causes a change in *critic $_{\mathcal{R}}$* , then we refer to it as *critic $_{\mathcal{R}}$ -transformational*. This extends Boden’s distinction between P- and H-creativity: A creative system observing a new artefact (whether or not it was that artefact’s creator) can change only its own \mathcal{R} , and thus can exhibit only P-transformativity. We distinguish “P-transformativity” from “P-creativity” to permit the inclusion of other necessary qualities in the judgement of the latter: novelty, value, etc.

We can now examine the events that lead *critic* to judge a new artefact to be *critic $_{\mathcal{R}}$ -transformational*. The rules that make up *critic $_{\mathcal{R}}$* cannot have been prescribed, they must have developed over time, changing in response to the perception of P-transformational objects. The rules that make up Wiggins’ set \mathcal{R} must be inferred from the creative system’s past

experiences. The rules in *critic $_{\mathcal{R}}$* cannot be descriptions of the domain as it exists independently of the critic system, they are merely *critic*’s current best guess at the state of the domain. The rules in \mathcal{R} are learned estimates that make up a *predictive model* of the domain – they can only be what the creative system *critic* *expects* the domain to be.

A kind of expectation, therefore, lies at the heart of both the transformational and the surprise criteria for creativity. The two approaches both concern the *un*-expectedness of an artefact. They differ, however, in how creativity is measured with respect to that unexpectedness. Transformational creativity occurs when a creative system’s expectations about the boundaries of the domain’s conceptual space – Wiggins’ \mathcal{R} – are updated in response to observing an artefact that broke those boundaries. Surprisingness occurs when a creative system’s expectations are violated in response to observing an artefact. Transformation, then, occurs in response to surprisingness, but both can occur in the same situations. This is not to say that all expectations are alike: “surprise” as construed by various authors as a creativity measure has involved a variety of kinds of expectation. The purpose of this comparison is to demonstrate that there is a common process between the two approaches, and we suggest that this commonality offers a pathway for future research.

From individual to societal transformativity

A remaining question concerns the nature of H-transformativity in a framework that considers all conceptual spaces to be personal predictive models. This must be addressed for an expectation-based approach to model transformation at the domain level – that which Boden originally proposed. If all \mathcal{R} and transformations thereof occur within a single creative system, then where does the “domain” as a shared entity reside? Modelling creativity as a social system (Csikszentmihalyi, 1988) is one way to answer that question, with the notion that creativity resides in the interactions of a society – between the creators, their creations and the culture of that society. This approach argues that the shared domain arises emergently out of the interactions of the society (Saunders and Gero, 2001b; Sosa and Gero, 2005; Saunders, 2012), and that it is communicated through the language and culture of that society. The effect of this is that overall “historical” creativity can be computationally measured, but only if some bounds are placed on history. Specifically, the transformativity of an artefact can be investigated with respect to the history of a defined society, not all of humanity.

One approach to operationalising this socially-derived H-creativity would be through a multi-agent systems metaphor: for an artefact to be judged H-creative it would need to receive a P-creative judgement from a majority of the pool of influence within the society, assuming that each agent possesses personal processes for judging the creativity of artefacts and the influentialness of other creative agents. This very simple formalisation does not model any of the influences discussed in Jennings (2010), but is intended to demonstrate how it would be possible to arrive at H-transformativity within a society given only P-transformativity within individual agents.

A framework for creative unexpectedness

The notion of expectation needs to be made more concrete if it is to be the basis of models of creativity evaluation. We develop a framework for the kinds of expectation that are

relevant to creativity evaluation, and situate some prior creativity evaluation models within that framework. The framework is designed to describe *what to expect* when modelling expectation for creativity. The framework is based on six dichotomies, an answer to each of which categorises the subject of an expectation relevant to the creativity of an artefact. These six questions are not intended to be exhaustive, but they serve as a starting point for exploration of the issue.

First we standardise a terminology for describing expectations:

- The *predicted* property is what is being expected, the dependent variable(s) of the artefact’s description. For example, in the expectation “it will fit in the palm of your hand” the size of artefact is the predicted property.
- The *prediction* property is the information about the predicted, such as a range of values or distribution over values that is expected to be taken by artefacts. For example, in the expectation “the height will be between two and five metres” the prediction is the range of expected length values.
- The *scope* property defines the set of possible artefacts to which the expectations apply. This may be the whole domain or some subset, for example “luxury cars will be comfortable”.
- The *condition* property is used to construct expectations that predict a relationship between attributes, rather than predict an attribute directly. These expectations are contingent on a relationship between the predicted property and some other property of the object – the condition. For example, the expectation “width will be approximately twice length” predicts a relationship between those two attributes in which the independent variable length affects the dependent variable width. In other expectations the prediction is unconditional and applies to artefacts regardless of their other properties.
- The *congruence* property is the measure of fit between an expectation and an observation about which it makes a prediction – a low congruence with the expectation creates a high unexpectedness and indicates a potentially creative artefact. Examples of congruence measures include proximity (in attribute space) and likelihood.

Using this terminology an expectation makes a *prediction* about the *predicted* given a *condition* that applies within a *scope*. An observation that falls within that scope is then measured for *congruence* with respect to that expectation. The six dichotomies of the framework categorise creativity-relevant expectations based on these five properties.

1. Holistic vs. reductionist

Expectations can be described as either *holistic*, where what is being predicted is the whole artefact, or *reductionist*, where the expectation only concerns some subset of features within the artefact. Holistic expectations make predictions in aggregate, while reductionist expectations make predictions about one or more attributes of an artefact, but less than the whole.

An example of a holistic expectation is “I expect that new mobile phones will be similar to the ones I’ve seen before”. This kind of expectation makes a prediction about the properties of an artefact belonging to the creative domain in which the creative system applies. The attribute(s) of all artefacts released within that domain will be constrained by that prediction. In this case what is being *predicted* is the whole

artefact and the *prediction* is that it will occupy a region of conceptual space. The *scope* is all possible artefacts within the creative domain of the system. The *congruence* measure calculates distance in the conceptual space.

This kind of expectation is typically at the heart of many computational novelty detectors – previously experienced artefacts cause a system to expect future artefacts to be similar within a conceptual space. One example is the Self-Organising Map based novelty detector of (Saunders and Gero, 2001a), where what is being predicted is the whole artefact, the scope is the complete domain, the prediction is a hyperplane mapped to the space of possible designs, and the congruence is the distance between a newly observed design and that hyperplane.

An example of a reductionist expectation is “I expect that new mobile phones will not be thinner than ones I’ve seen before”. This is a prediction about a single attribute of an artefact, but otherwise identical to the holistic originality prediction above: it is an expectation about all members of a creative domain, but about only one of their attributes. What is being *predicted* is the “depth” attribute, the form of that *prediction* is an inequality over that attribute, and the *scope* is membership in the domain of mobile phones.

Macedo and Cardoso (2001) use reductionist expectations in a model of surprise. An agent perceives some attributes of an artefact and uses these in a predictive classification. Specifically the agent observes the façades of buildings and constructs an expectation about the kind of building it is observing. The agent then approaches the building and discovers its true function, generating surprise if the expectation is violated. In this case the *predicted* property is the category to which the building belongs and the prediction is the value that property is expected to take.

2. Scope-complete vs. scope-restricted

Expectations can also be categorised according to whether they are *scope complete*, in which case the scope of the expectation is the entire creative domain (the universe of possibilities within the domain the creative system is working), or *scope-restricted*, where the expectation applies only to a subset of possible artefacts. The subset may be defined by a categorisation that is exclusive or non-exclusive, hierarchical or flat, deterministic or stochastic, or any other way of specifying which designs are to be excluded.

The mobile phone examples in the previous section are scope-complete expectations. An example of a scope-restricted expectation would be “I expect smartphones to be relatively tall, for a phone”. In this case the *predicted* property is device height (making this a reductionist expectation) and the *prediction* is a region of the height attribute bounded by the average for the domain of phones. The scope of this expectation, however, is artefacts in the category “smartphones”, a strict subset of the domain of mobile phones in which this creative system operates. This kind of expectation could be used to construct hierarchical models of novelty.

Peters (1998) uses this kind of hierarchy of expectations in a model of surprise – each level of their neural network architecture predicts temporal patterns of movement among the features identified by the layers below it, and surprise is measured as the predictive error. At the highest level the expectations concern the complete domain, while at lower levels the predictions are spatially localised.

3. Conditional vs. unconditional

Conditional expectations predict something about an artefact contingent on another attribute of that artefact. Unconditional expectations require no such contingency, and predict something about the artefacts directly. This is expressed in our framework via the condition property, which contains an expectation’s independent variables, while the predicted property contains an expectation’s dependent variable(s). A conditional expectation predicts some attribute(s) of an artefact conditionally upon some other attribute(s) of an artefact, while an unconditional expectation predicts attribute(s) directly. In a *conditional* expectation the prediction is that there will be a relationship between the independent attributes (the condition) and the dependent attributes (the predicted). When an artefact is observed this can then be evaluated for accuracy.

Grace et al. (2014a) details a system which constructs conditional expectations of the form “I expect smartphones with faster processors to be thinner”. When a phone is observed with greater than average processing power and greater than average thickness this expectation would be violated. In this case the *predicted* property is the thickness (making this a reductionist expectation), the *prediction* is a distribution over device thicknesses, and the *scope* is all smartphones (making this a scope-restricted expectation given that the domain is all mobile devices). The difference from previous examples is that this prediction is *conditional* on another attribute of the device, its CPU speed. Without first observing that attribute of the artefact the expectation cannot be evaluated. In Grace et al. (2014a) the *congruence measure* is the unlikelihood of an observation: the chance, according to the prior probability distribution calculated from the prediction, of observing a device at least as unexpected as the actual observation.

4. Temporal condition vs. atemporal condition

A special case of conditional expectations occurs when the conditional property concerns time: the age of the device, its release date, or the time it was first observed. While all expectations are influenced by time in that they are constructed about observations in the present from experiences that occurred in the past, temporally conditional expectations are expectations where time is the contingent factor. Temporal conditions are used to construct expectations about trends within domains, showing how artefacts have changed over time and predicting that those trends will continue.

Maher, Brady, and Fisher (2013) detail a system which constructs temporally conditional expectations of the form “I expect the weight of more newly released cars to be lower”. Regression models are constructed of the how the attributes of personal automobiles have tended to fluctuate over time. In this case the *predicted* property is the car’s weight, the *prediction* is a weight value (the median expected value), and the *scope* is all automobiles in the dataset. The *conditional* is the release year of the new vehicle: a weight prediction can only be made once the release year is known. The *congruence measure* in this model is the distance of the new observation from the expected median.

5. Within-artefact temporality vs. within-domain temporality

The question of temporally conditional expectations requires further delineation. There are two kinds of temporally contingent expectation: those where the time axis concerns the

whole domain, and those where the time axis concerns the experience of an individual artefact. The above example of car weights is the former kind – the temporality exists within the domain, and individual cars are not experienced in a strict temporal sequence. Within-artefact temporality is critically important to the creativity of artefacts that are perceived sequentially, such as music and narrative. In this case what is being predicted is a component of the artefact yet to be experienced (an upcoming note in a melody, or an upcoming twist in a plot), and that prediction is conditional on components of the artefact that have been experienced (previous notes and phrases, and previous plot events).

Pearce et al. (2010) describes a computational model of melodic expectation which probabilistically expects upcoming notes. In this case the *predicted* property is the pitch of the next note (an attribute of the overall melody), the *prediction* is a probability distribution over pitches. While the *scope* of the predictive model is all melodies within the domain (in that it can be applied to any melody), the *conditional* is the previous notes in the current melody. Only once some notes early in the sequence have been observed can the pitch of the next notes be estimated.

6. Accuracy-measured vs. impact-measured

The first five categorisations in this framework concern the expectation itself, while the last one concerns how unexpectedness is measured when those expectations are violated. Expectations make predictions about artefacts. When a confident expectation proves to be incorrect there are two strategies for measuring unexpectedness: how incorrect was the prediction, and how much did the predictive model have to adjust to account for its failure? The first strategy is *accuracy-measured* incongruence, and aligns with the probabilistic definition of unexpectedness in Ortony and Partridge (1987). The second strategy is *impact-measured* incongruence, and aligns with the information theoretic definition of unexpectedness in Baldi and Itti (2010). In the domain of creativity evaluation the accuracy strategy has been most often invoked in models of surprise, while the impact strategy has been most associated with measures of transformativity.

Grace et al. (2014b) proposes a computational model of surprise that incorporates impact-measured expectations. Artefacts are hierarchically categorised as they are observed by the system, with artefacts that fit the hierarchy well being neatly placed and artefacts that fit the hierarchy poorly causing large-scale restructuring at multiple levels. The system maintains a stability measure of its categorisation of the creative domain, and its expectation is that observations will affect the conceptual structure proportional to the current categorisation stability (which can be considered the system’s confidence in its understanding of the domain). Measuring the effect of observing a mobile device on this predictive model of the domain is a measure of impact. These expectations could be converted to a measure of accuracy by instead calculating the classification error for each observation, not the restructuring that results from it. The system would then resemble a computational novelty detector.

Experiments in expectability

To further illustrate our framework for categorising expectation we apply it to several examples from our recent work modelling surprise in the domain of mobile devices (Grace et al., 2014b,a). This system measures surprise by constructing expectations about how the attributes of a creative artefact

relate to each other, and the date which a particular artefact was released is considered as one of those attributes. Surprise is then measured as the unlikelihood of observing a particular device according to the predictions about relationships between its attributes. For example, mobile devices over the course of the two decades between 1985 and 2005 tended, on average, to become smaller. This trend abruptly reversed around 2005-6 as a result of the introduction of touch screens and phone sizes have been increasing since. The system observes devices in chronological order, updating its expectations about their attributes as it does so. When this trend reversed the system expressed surprise of the form “The height of device A is surprising given expectations based on its release date”. Details of the computational model can be found in earlier publications.

Figure 1 shows a plot of the system’s predictions about device CPU speed the system made based on year of release. At each date of release the system predicts a distribution over expected CPU clock speeds based on previous experiences. The blue contours represent the expected distribution, with the thickest line indicating the median. The white dots indicate mobile devices. The gradient background indicates hypothetical surprise were a device to be observed at that point, with black being maximally surprising. The vertical bands on the background indicate the effect of the model’s confidence measure – when predictions have significant error the overall surprise is reduced as the model is insufficiently certain in its predictions, and may encounter unexpected observations because of inaccurate predictions rather than truly unusual artefacts. An arrow indicates the most surprising device in the image, the LG KC-1, released in 2007 with a CPU speed of 806Mhz, considered by the predictive model to be less than 1% likely given the distribution of phone speeds before that observation. Note that after soon after 2007 the gradient of the trend increases sharply as mobile devices started to become general-purpose computing platforms. The KC-1 was clearly ahead of its time, but without the applications and touch interface to leverage its CPU speed it was never commercially successful.

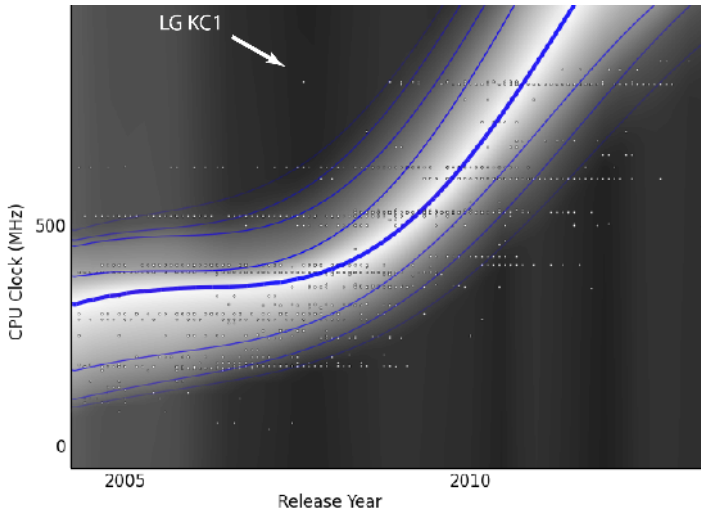


Figure 1: Expectations about the relationship between release year and CPU speed within the domain of mobile devices. The LG KC-1, a particularly unexpected mobile device, is marked.

This is a *reductionist, scope-complete, within-domain tem-*

porally conditional expectation, with congruence measured by *accuracy*. It is reductionist as the predicted attribute is only CPU speed. It is scope-complete because CPU speeds are being predicted for all mobile devices, the scope of this creative system. It is conditional because it predicts a relationship between release year and CPU speed, rather than predicting the latter directly, and that condition is temporal as it is based on the date of release. It is within-domain temporal, as the time dimension is defined with respect to the creative domain, rather than within the observation of the artefact (mobile phones are typically not experienced in a strict temporal order, unlike music or narrative). It is accuracy-measured as incongruence is calculated based on the likelihood of the prediction, not the impact of the observation on the predictive model.

Figure 2 shows another expectation of the same kind as in Figure 1, this time plotting a relationship between device width and release year. The notation is the same as in Figure 1 although without the background gradient. The contours represent the expected distribution of device masses for any given device volume. Here, however, the limits of the scope-complete approach to expectation are visible. Up until 2010 the domain of mobile devices was relatively unimodal with respect to expected width over time. The distribution is approximately a Poisson, tightly clustered around the 40-80mm range with a tail of rare wider devices. Around 2010, however, the underlying distribution changes as a much wider range of devices running on mobile operating systems are released. The four distinct clusters of device widths that emerge – phones, “phablets” (phone/tablet hybrids), tablets and large tablets – are not well captured by the scope-complete expectation. If a new device were observed located midway between two clusters it could reasonably be considered unexpected, but under the unimodality assumption of the existing system this would not occur. A set of *scope-restricted temporally conditional* expectations could address this by predicting the relationship between width and time for each cluster individually. Additionally a measure of the *impact* of the devices released in 2010 on this predictive model could detect the transformational creativity that occurred here.

Figure 3 shows a plot of the system’s predictions about device mass based on device volume. Note that – unsurprisingly – there is a strong positive correlation between mass and volume, and that the distribution of expected values is broader for higher volumes. Two groups of highly unexpected devices emerge: those around 50-100cm³ in volume but greater than 250gr in mass, and those in the 250-500cm³ range of volumes but less than 250gr mass. Investigations of the former suggest they are mostly ruggedised mobile phones or those with heavy batteries, and investigations of the latter suggest they are mostly dashboard-mounted GPS systems (included in our dataset as they run mobile operating systems).

This is a *reductionist, scope-complete, atemporal condition*, with congruence measured by *accuracy*. By our framework, the difference between the expectations modelled in Figure 1 and Figure 3 are that the former’s conditional prediction is contingent on time, while the latter’s is contingent on an attribute of the artefacts.

Figure 4 shows the results of a different model of surprise, contrasted with our earlier work in Grace et al. (2014b). An online hierarchical conceptual clustering algorithm (Fisher, 1987) is used to place each device, again observed chronologically, within a hierarchical classification tree that evolves and restructures itself as new and different devices are observed.

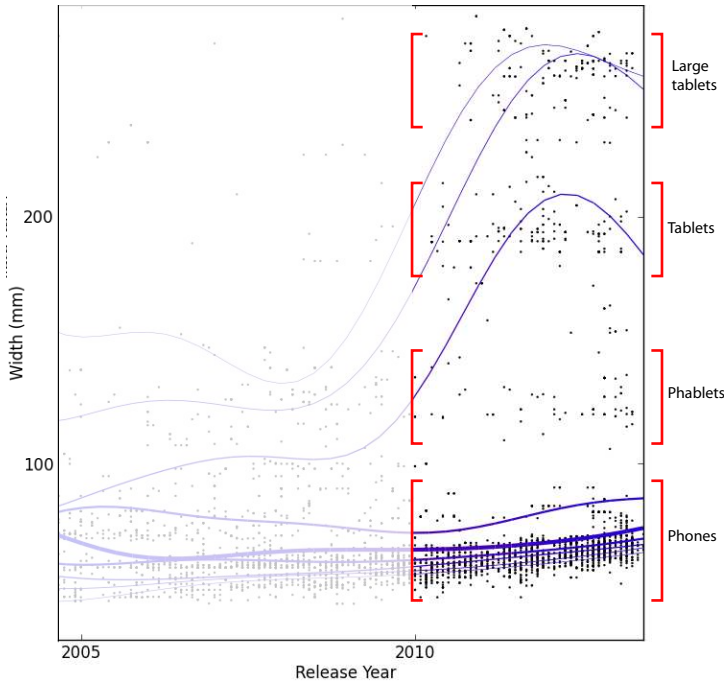


Figure 2: Expectations about the relationship between the release year and width of mobile devices. Note that the distribution of widths was roughly unimodal until approximately 2010, when four distinct clusters emerged.

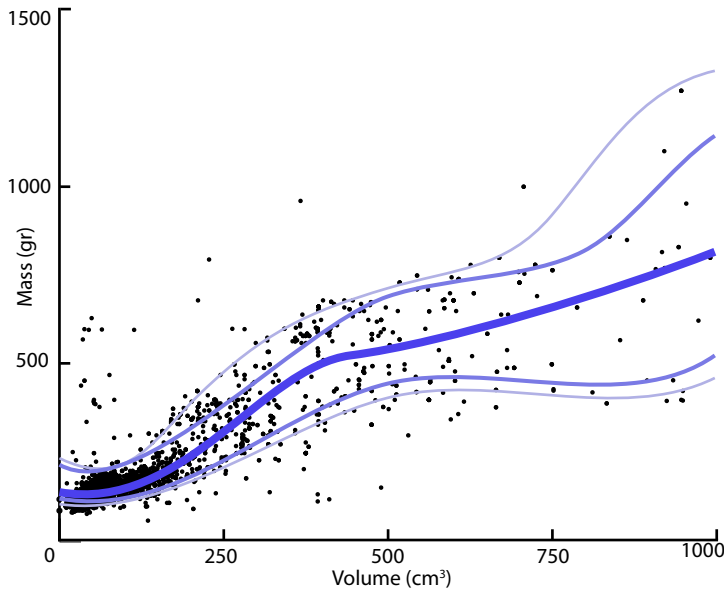


Figure 3: Expectations about the relationship between volume and mass within the domain of mobile devices.

The degree to which a particular device affects that tree structure can then be measured, indicating the amount by which it transformed the system’s knowledge of the domain. The most unexpected device according to this measure were the Bluebird Pidion BIP-2010, a ruggedised mobile phone which caused a redrawing of the physical dimensions based boundary between “tablet” and “phone” and caused a large number of devices to be recategorised as one or the other (although it must be noted that such labels are not known to the system). The second most unexpected device was the ZTE U9810, a 2013 high-end smartphone which put the technical specs of a tablet into a much smaller form factor, challenging the system’s previous categorisation of large devices as also being powerful. The third most unexpected device was the original Apple iPad, which combined high length and width with a low thickness, and had more in common internally with previous mobile phones than with previous tablet-like devices.

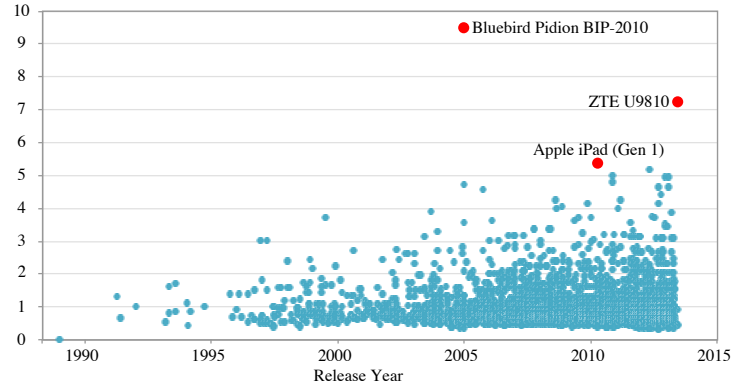


Figure 4: Incongruence of mobile devices with respect to their impact on learnt conceptual hierarchy. Three particularly unexpected devices are labelled.

This is a *reductionist, scope-complete, unconditional* expectation with congruence measured by *impact*. It is reductionist it does not predict all attributes of the device, only that there exists certain categories within the domain. It is scope-complete as it applies to all devices within the domain. It is unconditional as the prediction is not contingent on observing some attribute(s) of the device. The primary difference from the previous examples of expectation is the congruence measure, which measures not the accuracy of the prediction (which would be the classification error), but the degree to which the conceptual structure changes to accommodate the new observation.

Novelty, surprise, or transformativity?

Our categorisation framework demonstrates the complexity of the role of expectation in creativity evaluation, motivating the need for a deeper investigation. We argue that expectation underlies novelty, surprise, and transformativity, but further work is needed before there is consensus on what kinds of expectation constitute each notion.

Macedo and Cardoso (2001) adopt the definition from Ortony and Partridge (1987) in which surprise is an emotion elicited by the failure of confident expectations, whether those expectations were explicitly computed beforehand or generated in response to an observation. By this construction all forms of expectation can cause surprise, meaning that surprise and novelty have considerable overlap. Wiggins (2006a)

goes further, saying that surprise is always a response to novelty, and thus need not be modelled separately to evaluate creativity. Schmidhuber (2010) takes the opposite approach, stating that all novelty is grounded in unexpectedness, and that creativity can be evaluated by the union of usefulness and improvement in predictability (which would, under our framework, be a kind of *impact-based congruence*). Wiggins (2006b) would consider Schmidhuber’s “improvement in predictability” to be a kind of transformation as it is a measure of the degree of change in the creative system’s rules about the domain. Maher and Fisher (2012) state that the dividing line between novelty and surprise is temporality – surprise involves expectations about what will be observed next, while novelty involves expectations about what will be observed at all. Grace et al. (2014a) expand that notion of surprise to include any conditional expectation, regardless of temporality.

We do not offer a conclusive definition of what constitutes novelty, what constitutes surprise, and what constitutes transformativity, only that each can be thought of as expectation-based. It may well be that – even should we all come to a consensus set of definitions – the three categories are not at all exclusive. We offer some observations on the properties of each as described by our framework:

- Surprise captures some kinds of creativity-relevant expectation that extant models of novelty do not, namely those concerned with trends in the domain and relationships between attributes of artefacts.
- Models of surprise should be defined more specifically than “violation of expectations” if the intent is to avoid overlap with measures of novelty, as novelty can also be expressed as a violation of expectations.
- The unexpectedness of an observation and the degree of change in the system’s knowledge as a response to that observation can be measured for any unexpected event, making (P-)transformativity a continuous measure. Models of transformative creativity should specify the kind and degree of change that are necessary to constitute creativity.

Conclusion

We have sought to build theoretical bridges between the notions of novelty, surprise and transformation, reconceptualising all three as forms of expectation. This approach is designed to offer a new perspective on debates about the roles of those disparate notions in evaluating creativity. We have developed a framework for characterising expectations that apply to the evaluation of creativity, and demonstrated that each of novelty evaluation, surprise evaluation, and transformational creativity can be conceived in terms of this framework. Given the wide variety of kinds of expectation that should be considered creativity-relevant we argue that originality alone is not a sufficient accompaniment to value to constitute creativity. This insufficiency is a critical consideration for computational models that can recognise creativity. The expectation-centric approach provides a framing device for future investigations of creativity evaluation. Expectation both serves as a common language by which those seeking to computationally model creativity can compare their disparate work, and provides an avenue by which human judgements of creativity might be understood.

References

Abra, J. 1988. *Assaulting Parnassus: Theoretical views of creativity*. University Press of America Lanham, MD.

- Baldi, P., and Itti, L. 2010. Of bits and vows: a bayesian theory of surprise with applications to attention. *Neural Networks* 23(5):649–666.
- Berlyne, D. E. 1966. Curiosity and exploration. *Science* 153(3731):25–33.
- Berlyne, D. E. 1970. Novelty, complexity, and hedonic value. *Perception & Psychophysics* 8(5):279–286.
- Bishop, C. M. 1994. Novelty detection and neural network validation. In *Vision, Image and Signal Processing, IEE Proceedings-*, volume 141, 217–222. IET.
- Boden, M. A. 2003. *The creative mind: Myths and mechanisms*. Routledge.
- Csikszentmihalyi, M. 1988. *Society, culture, and person: A systems view of creativity*. Cambridge University Press.
- Dudek, S. Z. 1993. The morality of 20th-century transgressive art. *Creativity Research Journal* 6(1-2):145–152.
- Feyerabend, P. K. 1963. How to be a good empiricist: a plea for tolerance in matters epistemological. In *Philosophy of science: The Delaware seminar*, volume 2, 3–39. New York, Interscience Press.
- Fisher, D. H. 1987. Knowledge acquisition via incremental conceptual clustering. *Machine learning* 2(2):139–172.
- Florida, R. L. 2012. *The Rise of the Creative Class: Revisited*. Basic books.
- Grace, K.; Maher, M.; Fisher, D.; and Brady, K. 2014a. A data-intensive approach to predicting creative designs based on novelty, value and surprise. *International Journal of Design Creativity and Innovation* (to appear).
- Grace, K.; Maher, M.; Fisher, D.; and Brady, K. 2014b. Modeling expectation for evaluating surprise in design creativity. In *Proceedings of Design Computing and Cognition*, volume (to appear).
- Horvitz, E. J.; Apacible, J.; Sarin, R.; and Liao, L. 2012. Prediction, expectation, and surprise: Methods, designs, and study of a deployed traffic forecasting service. *arXiv preprint arXiv:1207.1352*.
- Itti, L., and Baldi, P. 2005. A principled approach to detecting surprising events in video. In *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, volume 1, 631–637. IEEE.
- Jennings, K. E. 2010. Developing creativity: Artificial barriers in artificial intelligence. *Minds and Machines* 20(4):489–501.
- Macedo, L., and Cardoso, A. 2001. Modeling forms of surprise in an artificial agent. *Structure* 1(C2):C3.
- Maher, M. L., and Fisher, D. H. 2012. Using ai to evaluate creative designs. In *2nd International conference on design creativity (ICDC)*, 17–19.
- 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*, 147.
- 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*, 22–28. Desire Network.
- Newell, A.; Shaw, J.; and Simon, H. A. 1959. *The processes of creative thinking*. Rand Corporation.

- O'Quin, K., and Besemer, S. P. 1989. The development, reliability, and validity of the revised creative product semantic scale. *Creativity Research Journal* 2(4):267–278.
- 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.
- Pearce, M. T.; Ruiz, M. H.; Kapasi, S.; Wiggins, G. A.; and Bhattacharya, J. 2010. Unsupervised statistical learning underpins computational, behavioural, and neural manifestations of musical expectation. *NeuroImage* 50(1):302–313.
- Peters, M. 1998. Towards artificial forms of intelligence, creativity, and surprise. In *Proceedings of the twentieth annual conference of the cognitive science society*, 836–841. Citeseer.
- Runco, M. A. 2010. *Creativity: Theories and themes: Research, development, and practice*. Access Online via Elsevier.
- Saunders, R., and Gero, J. S. 2001a. The digital clockwork muse: A computational model of aesthetic evolution. In *Proceedings of the AISB*, volume 1, 12–21.
- Saunders, R., and Gero, J. S. 2001b. Artificial creativity: A synthetic approach to the study of creative behaviour. *Computational and Cognitive Models of Creative Design V*, *Key Centre of Design Computing and Cognition, University of Sydney, Sydney* 113–139.
- Saunders, R. 2012. Towards autonomous creative systems: A computational approach. *Cognitive Computation* 4(3):216–225.
- Schmidhuber, J. 2010. Formal theory of creativity, fun, and intrinsic motivation (1990–2010). *Autonomous Mental Development, IEEE Transactions on* 2(3):230–247.
- Schumpeter, J. 1942. Creative destruction. *Capitalism, socialism and democracy*.
- Silberschatz, A., and Tuzhilin, A. 1995. On subjective measures of interestingness in knowledge discovery. In *KDD*, volume 95, 275–281.
- Sosa, R., and Gero, J. S. 2005. A computational study of creativity in design: the role of society. *AI EDAM* 19(4):229–244.
- Strzalecki, A. 2000. Creativity in design: General model and its verification. *Technological Forecasting and Social Change* 64(2):241–260.
- Wiggins, G. A. 2006a. A preliminary framework for description, analysis and comparison of creative systems. *Knowledge-Based Systems* 19(7):449–458.
- Wiggins, G. A. 2006b. Searching for computational creativity. *New Generation Computing* 24(3):209–222.