What is deemed computationally creative?

S. Agrawal
A. Sankaran
A. Laha
S. A. Chemmengath
D. Shrivastava
K. Sankaranarayanan

In this new era of computational creativity, where artificial intelligence (AI) systems are attempting to achieve human-level creativity, a set of golden questions needs to be answered, such as "What kind of systems are creative systems?" and "When does a system qualify as truly creative?" The existing generation of AI-driven cognitive systems is based on the goal of achieving human-level intelligence, not human-level creativity. Creativity is considered subjective with respect to both the application domains and the perceiving end-user. In this paper, we postulate the dimensions and factors that distinguish creativity and intelligence, such as novelty, value, surprise, influence, coherence, correctness, and comprehensibility. We group the application domains into time-dependent and time-independent ones and define a framework to describe these dimensions in each application. In addition to defining the factors that determine creativity, we also suggest ideas on how to evaluate these factors. We strongly believe that the proposed framework would act as the basis for building and evaluating creative systems and also provide us with the ultimate goal for achieving human-level creativity.

Introduction

In the past few decades, psychologists have taken additional interest in studying the different aspects of human creativity [1, 2]. The quintessential argument that has been lingering through the literature centers around two questions: *What tasks are considered creative?* and *How to measure the extent of creativity in these tasks?* The most generic definition includes a wide range of tasks such as composing music, designing fashion, creating works of arts and painting, story writing, movie making, teaching students, playing competitive and noncompetitive sports, and of course, writing research papers. Extending on this definition, there has been a lot of studies and arguments on the relationship between creativity and intelligence [3–5]. There are the following five possible relationships between human creativity and intelligence, according to Sternberg's framework [2]:

1) creativity as a subset of intelligence;
2) intelligence as a subset of creativity;
3) creativity and intelligence as overlapping sets;
4) creativity and intelligence as coincident sets;
5) creativity and intelligence as disjoint sets. Plucker and Esping [6] provide examples for all five cases, establishing a strong relationship between creativity and intelligence.

Computationally, the primary aim of artificial intelligence (AI) is to understand and reproduce the capabilities of human intelligence in a machine. In the last two decades, machine learning has pushed the boundaries of AI, automating several tasks that were erstwhile considered challenging [7]. These include tasks such as virtual personal assistants, prediction and recommendation systems, financial services, healthcare systems, and even automated code generation systems. These intelligent systems aid human intelligence in day-to-day life activities and in many critical life-saving activities. On a comparative scale, there is a lack of creative systems that could amplify humans in their creative tasks. Garry Kasparov [8], explaining his 1997 game series against IBM's supercomputer Deep Blue in his book *Deep Thinking: Where Machine Intelligence Ends and Human Creativity Begins,* states that creativity could be the highest form of intelligence. Thus, the grand challenge for machine learning would be to develop creativity-augmenting models and systems, addressing the challenges in computational

Digital Object Identifier: 10.1147/JRD.2019.2900640

creativity. The primary challenges involved in building creative systems (akin to intelligent systems) are summarized as follows.

1) *Subjectivity:* Defining creativity is considered highly subjective. For instance, the same work of art might appeal for a selected audience while not to others due to different social and cultural influences of the population [9]. Thus, a creative system should consider the target audience's background. However, subjectivity adds beauty to any creative work as the individualization of the creator is expressed and appreciated. This individualization also needs to be considered while assessing creativity.
2) *Domain dependency:* The explanation and expectations of creativity change across domains and tasks. The creativity of a fashion designer, a chef, and a musician will be diversely different. Thus, a creative system should consider the expectation of the different tasks.
3) *Evaluation:* Currently, comparison of different creative systems is more qualitative than quantitative. This is due to lack of metrics, measures, and evaluation methods of creative systems. A creative system should be able to measure its own creativity and improve upon it.

In the literature, researchers and psychologists have discussed few techniques to address these challenges. Measurement of creativity is split into two parts [10]: first, p-creativity, which measures the creative cognition of an individual; and second, h-creativity, which measures the global creative cognition.

There exist creative systems in the literature that are custom developed for a specific task or domain. Recently, IBM's Watson Beat [11] could automatically compose music for video footage and augment a human's creative capabilities. Google's Project Magenta [12] aims at generating music and art (images) using machine learning aiding human creative experts. Sony's Flow Machines [13] generated a pop song called "Daddy's Car," impersonating the style of The Beatles. However, the way creativity is perceived and evaluated across all these systems varies.

To mitigate the challenges discussed above, in this paper, we propose seven necessary and sufficient conditions (dimensions) for a system to be creative, which include novelty, value, surprise, influence, coherence, correctness, and comprehensibility. The different dimensions explained in this paper are driven by and based on various psychoanalysis results showcased in the literature. The sets of all possible creative tasks (or domains) are grouped into two categories: 1) time-dependent and 2) time-independent. Some examples of time-dependent tasks are story writing [14], music composition, and movie creation [15]. Sculpting or painting, food recipe creation, fashion design [16], and manufacturing designing process are categorized as time-independent tasks. The conditions for creativity and the definitions for these conditions can vary across different tasks. This paper aims to provide a comprehensive framework, laying out the essential criteria for creativity in varied domains or tasks. This framework is then used to evaluate quantitatively the creativity of any system in those domains.

The remainder of this paper is organized as follows: In the first section, we detail the seven dimensions of creativity. In the second section, we explain the definition of these creativity measure dimensions for six different tasks/domains: three time-dependent and three time-independent. In the fourth section, we provide a computational method of mathematically measuring these dimensional values. In the fifth and sixth sections, we discuss the related work in the literature and conclude this research work, respectively. In this paper, the word *artifact* is used to refer to any item that might be generated by an algorithm. The item itself could belong to any domain. In addition, we use the terms *tasks* and *domains* interchangeably.

Conditions for creativity

Creativity is defined as the act of doing something new to form something valuable [17]. Sternberg and Lubart [18] define creativity as follows:

A product is creative when it is (a) novel and (b) appropriate. A novel product is original not predictable. The bigger the concept, and the more the product stimulates further work and ideas, the more the product is creative.

Creativity is often measured in terms of novelty and value. While these two are the necessity criteria for defining creativity, it is an open research to study the different sufficient conditions to measure the creativity of a system.

The basic idea is to look at the works that are usually called creative by the experts in those fields and distill the factors and conditions that they subconsciously impose on these artifacts to label them creative. Another factor that is taken into consideration is whether these conditions could be quantitatively measured. In the following subsections, we discuss the following seven conditions for creativity: novelty, value, surprise, influence, coherence, correctness, and comprehensibility.

Novelty

The degree of novelty of an artifact is defined as the extent of dissimilarity of it with the existing corpus of artifacts in that particular domain and scope. This measure of dissimilarity is very subjective and could vary from one domain to the next. For example, coming up with a fusion recipe for *Honey Paneer Chicken Butter Masala Roulade* might be considered a novel idea in the food domain; on the

other hand, a fusion of jazz and rock might not be novel in the music domain. It is to be remembered that novelty is dependent on the context and the target audience for which it is measured. For example, a person who has not read the *Harry Potter* books might consider the movies to be a novel idea, whereas a person who has read the books would not.

Value

Value is the perception of the utility or performance or attractiveness of a particular artifact as seen by the target audience in a given scope [19, 20]. Utility, performance, and attractiveness could take different meanings across different domains. For example, if a computer produces a design of laptop with its screen placed backward, this would be considered novel but most certainly will not be called valuable, as this would not be useful for the end-user in any possible scope. As another example, if a computer produces music with a series of high-frequency notes, it would turn out to be discordant and could be called unattractive.

Surprise

Surprise can be given as the deviation of a new artifact from the existing expectations from an artifact with a given scope and history [21, 22]. As an example, this could be seen as the presence of abnormal and new combinations of ingredients in a particular recipe. This would deviate from the normal and present notions of combinations of ingredients, yet, might turn out to be both negative and positive. Therefore, the presence of surprise would not give us any indication of it being positive or negative.

Influence

Influence largely depends on the analysis of the future effects of a particular artifact on all the artifacts produced in the domain. For example, Shakespeare's works are said to be influential because his style of writing has managed to spawn an entire genre of writing. Hence, evaluating influence [20, 23] would entail looking at similar artifacts from the domain before and after the generation of the given artifact. In our opinion, while influence is not a primary contributor toward measuring the creativity of an entity, it would surely increase the creativity score.

Coherence

Coherence would best be explored in the time domain, as in the continuation of a particular thought or event over multiple parts of artifact without any contradictions. It could also be viewed as the satisfaction of any new events by the target audience. For example, a story in which a person is currently in London and is in Tokyo in the next scene without any explanation would be considered incoherent.

Table 1 List of different time-dependent and time-independent creative tasks are characterized in this research.

Time Dependent	Time Independent
Writing Stories	Painting or Artwork
Composing Music	Food Recipes
Making Movies	Industrial Design

Correctness

Correctness is the semantic and logical appropriateness of a particular artifact. This could be understood when you would find it acceptable to portray Harry Potter with a wand, but at the same time, to portray James Bond with a wand would not be logically appropriate. Correctness is best evaluated by experts of that domain, as some facts might make sense to creative experts only. For example, some of the deeper physics-based rationale behind events portrayed in *Interstellar* movie might make sense to physicists but not to everyday moviegoers. Correctness is usually considered to evaluate the artifacts in time-independent factors.

Comprehensibility

Comprehensibility is highly scope-dependent and refers to the ability of a particular audience to understand a particular artifact. For example, an expert in finance would find a movie about the 2008 financial crash completely understandable, while another person with no knowledge of the domain might find the movie too difficult to understand.

It is to be noted that all the seven dimensions independently contribute to the creativity of a system; however, individually they may not have a significant impact in every domain. The basic principle behind the framework is to suggest that all the seven dimensions, in a cumulative fashion, contribute toward the evaluation of a creative system.

Creativity across multiple domains

We consider the existing domain of applications and categorize into two divisions: 1) time-dependent domains and 2) time-independent domains. We broadly segregate the domains as time-dependent and time-independent. The domains are summarized in **Table 1**. Time-dependent mostly refers to domains or tasks that would require a level of temporal analysis while analyzing their data. Time-independent refers to those domains wherein the artifacts are generated independent of the other artifacts. Here,

coherence is not considered as a parameter, as it has a temporal aspect.

We now discuss the various conditions on each of the creativity domains that would be necessary for an artifact to be called a part of that domain, the definitions of which would have to be appropriately extended to accommodate for future developments in the fields.

Writing stories

A story is defined as a sequence of sentences used to depict a particular event, more often in a detailed fashion [14]. We attempt to obtain a particular definition of each of the conditions while running through them with an appropriate example of the *Harry Potter* books, which are certainly considered creative.

We can consider a story to be novel if it is different from all other existing stories with respect to its plot/storyboard, character maps, space, and theme. Each of these measures can be calculated using various techniques [24]. For example, while studying the impact of character maps in the *Harry Potter* book series, the story revolves around a central character, Harry Potter [25]. Fay [26] studied that in *Lord of the Rings*, which is considered closest to *Harry Potter* in terms of overall plot, and does not have one single central character, except Bilbo Baggins. This could be further quantitatively estimated using learnable character models, which map characters across a range of traits such as vengeance, ambition, and compassion, using which similarity between characters is computed. The *Harry Potter* book also introduces a parallel universe and describes every detail of its elements, from the Gringotts, the magical bank, to Hogwarts, the magical school, to Azkaban, the prison. Thus, ensuring a unique setting for the story yields novelty in the storyboard.

Value is defined as the attractiveness created by this story in its particular genre. Different stories of a similar genre and the ratings of the stories could be obtained from book-curating websites such as Goodreads, Amazon, and Common Sense Media. Specifically, we avoid using the factor of revenue generated to measure value, as there have been a large number of stories and books that have been received critically when they were launched but have later gone on to become huge successes. In addition, a more influential book would be considered more valuable as it would lead to the creation of more value in the future. For example, the *Lord of the Rings* series received a lot of criticism when it was originally released, but later sold more than 200 million copies [27]. Therefore, it is very challenging to predict the true influence of a particular book by looking at its instantaneous sales. The weighted average of the corresponding ratings weighed with the similarity scores gives us the possible rating of the current story and correspondingly its perceived value. For example, *Harry Potter* is rated 4.8/5.0 by Amazon, 4.4/5.0 by Goodreads, and 5.0/5.0 by Common Sense Media. Considering that this is an old story and not a new one, a simple average of these values gives us a value of 4.73, which suggests the value of this story.

A story is said to be surprising if it deviates from what is expected of it at any point. The story could deviate along any of the elements proposed earlier. A book that is more surprising might also be more novel, as the expectation of events in some way are correlated to the past knowledge of books. In *Harry Potter*, there are several instances of surprise in the plot, such as when it is revealed in the first book itself that Harry, the protagonist, had a personal rivalry with Voldemort, the central villain of the wizardry world, having only met him as a child, and this goes on to become the premise for all the books that follow. These also make the story novel, as the readers would not have seen such a twist in the past. The screenplay of the story is responsible for conveying the surprise element.

The influence of a given story is measured by looking at the average similarity of this story with all the stories that come after it, scaled by its similarity to the books that came before it. Here, we would also want to consider the number of books in that particular genre before and after that given story. This would give us a sense of the change in popularity of a particular genre with time with respect to the story, given that no other major book was released in the same time frame. This would ensure that this story created a major influence in this genre and encouraged more stories to be written. The novelty of the book also increases its influence, as that aspect of the story helps it differentiate itself from the others behind it and act as a starting thread for the stories that follow. It would also include the effects of the popularity and value of the story. For example, *Harry Potter* not only inspired a lot of fan fiction, it also made (Young Adult + Magic) the genre popular apart from being made into a very famous movie series [28–30].

Coherence refers to the concept within a story that seems to move in the same direction or is at least well connected. There might be a tradeoff between coherence and surprise, wherein some jumps in the story might make it more interesting while reducing the coherence. For example, when Harry has visions from Voldemort's mind, it makes the story interesting as it leaves the reader puzzled about what the visions meant. If there are too many such events, it would make it difficult to follow the flow of thought of the story [31].

Correctness refers to the logical correctness of the story, wherein any new event must not contradict any of the rules set by the events occurring before it. For example, the characters acquiring powers to fly without their brooms would contradict the fact that wizards cannot fly and would thus be incorrect, unless the story justifies this directly or indirectly. The fundamental aim with this dimension is to ensure that the correctness of the artifact is not lost in the process of increasing its creativity.

Comprehensibility can be defined as the level of the language of a story keeping in mind its scope. This can be measured by creating and comparing the vocabulary of the existing stories with that of our story and finding the similarity between each of them. Different vocabularies can also be constructed based on preclassified stories according to their age group, demography, and technical prowess. The given story's vocabulary should be closest to the other vocabularies obtained from stories in the same genre/scope. For example, a toddler would certainly not be able to read the *Harry Potter* books as they do not have as evolved a system of language and would instead prefer a children's book or a picture book. On the other hand, a satire on the Indian political scenario may probably be targeted at an Indian older than 18 years of age, having a decent amount of general knowledge.

On the other hand, if we consider a book like a technical manual, it would not be considered creative, as it fails across most parameters. This type of book is not novel because it simply describes the instructions to use a machine and would probably be similar for similar machines. The manual does have utility but no attractiveness and thus minimal value. There is no element of surprise or any influence on future work as the task of manual writing is fairly mechanical and would not change much. The manual does have coherence, correctness, and comprehensibility, as it is meant to be understood by a person using the machine and also for them to be able to use it by following the instructions. Thus, the manual would not have novelty, surprise, and influence and minimal value and thus cannot be called creative. This also indicates that coherence, comprehensibility, and correctness are the minimum criteria that need to be satisfied by any writing, creative, or otherwise. This suggests that a subset of dimensions is not enough to compute creativity in a satisfactory manner. The relation between different dimensions to evaluate creativity in writing stories is shown in **Figure 1**.

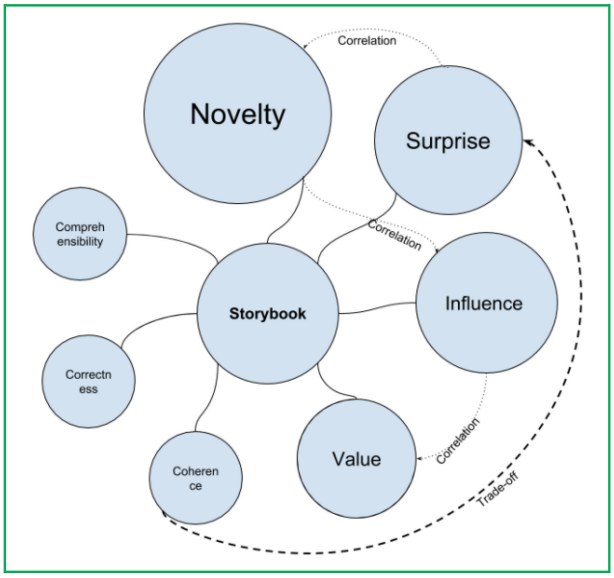

Figure 1

Relationship and the importance of the different dimensions in evaluating the creativity of stories and books.

Composing music

Music in its most elementary form refers to a sequence of sounds in harmony. This does not make any presumptions about the periodicity or the likability of the sounds. To consider an example, over the years Bob Marley has been recognized as one of the best reggae artists across the globe, and his songs can be considered creative, indisputably.

Novelty of a song is measured along the various elements of music, namely pitch, timbre, loudness, and other such factors, which are computed for small subsegments of the song of a few microseconds in length. These are compared using various techniques developed for comparison of music for the purpose of intellectual property rights. The similarity scores thus obtained can be used to decide on the novelty of the given song [32–34]. For example, the song "Buffalo Soldier" by Bob Marley and The Wailers is considered similar to "The Tra-La-La Song" by the Banana Splits, as the choruses of both songs are similar [35].

The value of a song is represented by its attractiveness and can be seen as the cumulative sum of many different factors as already done by organizations like Billboard for the calculation of top songs using factors. Additional stats that indicate the total airplay of a particular song, like the sales, online streaming stats, and radio presence, are also considered important for value. These factors combined provide for a quantitative measure for the value of a song. All of Bob Marley's songs, for example, were very famous in all areas including but not limited to radio airtime and are even today some of the most famous tracks for a reggae radio station.

Surprise in a song is very closely related to novelty and would be measured by the similarity of a track to other tracks generated by itself. The system should thus be able to generate vastly different tracks for the same set of inputs. For example, it would be surprising if the system generates a jazz mix of a song when asked for a song by Bob Marley.

A song's influence can be measured in many ways including but not limited to the similarity of a song to the songs released after it and also the change in the popularity of the people associated with the song. The album *The Best of The Wailers* is considered an influential album because it completely modified the reggae beats by moving to a slower beat.

Coherence and correctness as factors in music is a topic of debate among experts, as some believe that music cannot

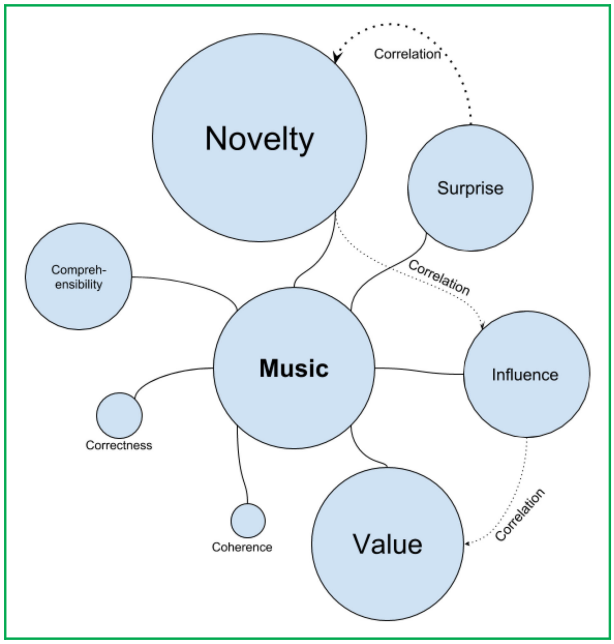

Figure 2

Relationship and the importance of the different dimensions in evaluating the creativity of composing music.

have any rules, whereas others believe that music plays by a set of underlying rules, thus giving an indication about its correctness. For example, the Indian Carnatic music has a very strict and vast grammar for music generation, while genres such as blues or a more recent neofusion have a less restrictive grammar.

The comprehensibility of a certain song can be given by its likability within the people of that scope. For example, rock music might not be appreciated by a group of people at a classical concert, but might still be appreciated at a hip-hop concert with which it shares some level of similarity.

If we consider noise like the cacophony of traffic, we certainly would not call it music. This fits well with our theory as it is novel as the combination of the various factors is certainly unique. The sound would have no value or surprise, as it is discordant and is most certainly not comprehensible by anyone as people usually try to avoid it. Thus, the sound fails to match any criteria apart from novelty and hence cannot be called novel.

The relationship between the different dimensions in computing the creativity of composed music is shown in **Figure 2**.

Movies

Movies are defined, in the most simplistic form, as a type of visual communication that uses moving pictures and sound to tell stories or inform (help people to learn). We consider the basic example of the movie *Fight Club* (1999), which is currently ranked 14th in the list of Top Movies in the Internet Movie Database (IMDb) [36]. This domain draws its influence from stories but is differentiated from them by the manner in which they portray each of the scenes in the story, a major example of this being the remakes of various movies that have been released in the recent past.

Novelty in movies is defined as how different a particular movie is from the movies preceding it in terms of its three most basic components: the storyline, visualizations of that storyline, and the sound that puts all of these together. Each of these components is then individually split into multiple factors, similar to how stories are evaluated as discussed earlier; for example, the visualizations can be evaluated along the technical components of *mise en scene*, while sound is analyzed by looking at how it ties the story and the visuals together [37]. Essentially, we would look at the novelty of the story at the most elementary level, which is followed by looking at how different the visualizations are given the underlying storyline. This is followed by analyzing how the sound plays a role in bringing them together—that is, is it in the traditional format wherein the dialogues start and finish at the boundaries of the corresponding scene, or is it done in a different manner? These factors become evident when we look at the example of the movie *Fight Club*, wherein the story is in itself novel due to the basic factor in which the main character is seen as two different people throughout the movie and the viewers are not given a hint about the fact that they represent the same person until the very end. The jumpy and flaky timeline (screenplay) in the story also lends a lot of novelty to it, as there are very few stories that follow such a pattern but are still coherent. The visualizations are also unique in the sense that they manage to bring out the contrast in the two personalities' lives very appropriately by showing the normal person's surroundings as drab and dull while that of Tyler, or the violent one, with very high contrast and lots of colors. This ensures that the difference of personalities becomes more evident to the viewer. Lastly, the sound is also considered very important, wherein the fight scenes were made even more realistic and believable using the novel sound effects for the punches instead of the standard Foley sounds [38].

The value of the movie can be looked at by three most popular techniques, such as critics' reviews, movie earnings, and user reviews. A combination of these ratings ensures that none of the creative movies is underrepresented due to the inherent biases in any of these factors. On the other hand, some of these techniques, like movie earnings, might be heavily dependent on the cast and crew of the movie and would have to be appropriately normalized for those factors. Rotten Tomatoes [39] reviews act as an aggregation of movie critic reviews and thus their proxy, while IMDb ratings normalized appropriately using a technique like Wilson's score to account for the number of

reviewers are a representative of the average user ratings. We can see the value in our *Fight Club* example, as it is 14th on IMDb's top movies list and also has a 78% Rotten Tomatoes rating. On the other hand, the movie has barely broken even over the years, as it had a budget of more than $63 million USD and has lifetime earnings of roughly $100 million USD. Hence, this factor alone would be a poor indicator of the film's popularity [40].

The surprise in a movie would be very similar to that in a story, wherein a movie would be deemed surprising if at any point, it deviates from the notions of expectations. The difference here is that the notion of expectation would be in the sense of the story, the visuals, and the sound. A classic example of surprise is horror movies, where neither of these factors individually surprises the user, whereas it is the combination of these factors that manages to surprise the viewer. In our example of the movie *Fight Club*, we find multiple moments of surprise whenever the scene jumps the timeline, as the moment in time we would land in is always unexpected. In addition, the final realization of the dissociated personalities is very surprising, as it shatters the previous assumptions and notions about the movie in the mind of the viewer.

The influence [20] for movies would be defined by looking at both the similarity of movies before and after this particular movie, coupled with the increase in the number of the movies of this particular genre after the given movie. In addition, in the context of documentaries, this factor would be most important and could be considered as the level of awareness raised by a particular documentary. This factor becomes especially important, as many movies might be creative, but most of them are unable to change the way movies are made or accepted, in general. For example, movies with more realistic representations of the world having both the positive and negative aspects of life showcased with equal importance were made famous by David Fincher, the director of *Fight Club*, not only using *Fight Club*, but also with the movie *Se7en*. This has led to a wider acceptance of such movies with true representations.

Coherence in movies would refer to the general flow of concept in a movie, both visually and semantically in the story. Here, coherence can have a significant tradeoff with surprise and novelty, wherein the correlation between multiple scenes might be implicit and not explicit. Considering the example of *Fight Club*, this tradeoff becomes very evident as each time the movie jumps in the timeline, the viewer is expected to understand the context of the new scene with no explicit context being provided. This makes the movie difficult to understand for some people, but also makes the movie enjoyable and creative for other people.

Correctness can be defined in the context of movies, wherein there are no changes or no preset rules are violated that are not explained in the movie either explicitly or implicitly. We make this definition broader to account for the possibility of movies in genres such as horror, wherein a person might randomly appear and disappear in a fraction of a second. However, here this would be correct, as there is a basic underlying assumption of ghosts and their properties. In addition, in *Fight Club*, jumping between timelines would not make sense until and unless the basic assumption of different time periods was properly established.

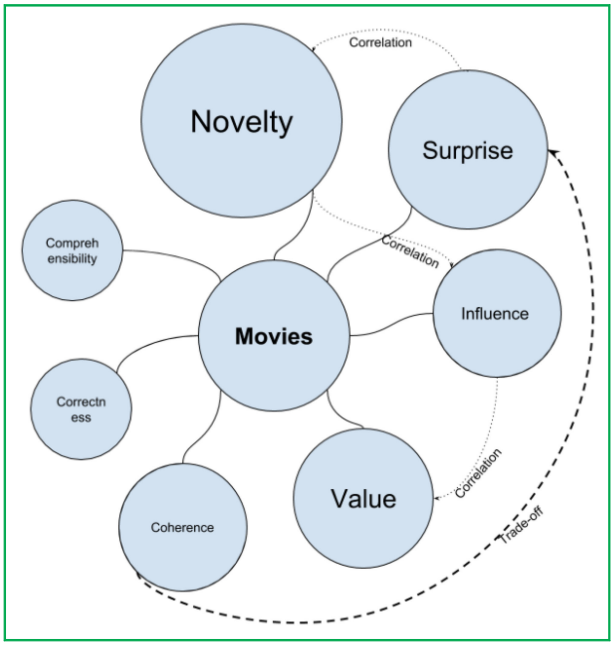

Figure 3

Relationship and the importance of the different dimensions in evaluating the creativity of movie making.

Comprehensibility refers to the ability of the group of people to properly understand any given movie. This pertains mostly to incoherent movies and technical documentaries, as a movie like *Fight Club* might be incoherent for some people, but is comprehensible for others.

Let us consider the example of the dash-cam footage of a police officer on an uneventful day played back to you at a completely normal speed. This would surely be comprehensible, correct, and coherent. However, it would not be novel, as it is similar to the thousands of hours of footage from police dash-cams in the past. In addition, it would not have value because no one would pay to see it and no user would rate it highly. Though it could have some value in crime rate detection, which is not a value for the majority. Lastly, there would be no surprise or influence, as it would be very monotonous. Hence, using our framework, we can safely call this footage noncreative.

The correlation among different dimensions in evaluating creativity in movies is shown in **Figure 3**.

Painting and artwork

A painting is the application of a paint-like substance on a surface. In this domain, we use Pablo Picasso as an example to evaluate creativity along the various dimensions [23, 41].

Novelty in a painting can be evaluated by considering a range of factors such as color, texture, edges, light, and localized segments of the image, which can be compared with other paintings to help determine its similarity. Some high-level features that could be used are the brush strokes and the subject matter, which are usually unique to an artist's painting style. Picasso's style was novel, as he invented cubism at the confluence of Greek, Iberian, and African art [42]. He also dabbled in a form of Surrealism, making his paintings vastly different from the prevalent style of painting.

The value of a painting can be quantitatively defined by the amount it could sell for and could be estimated by roughly estimating the total area of a particular painting and then multiplying it by a factor that would be dependent on factors such as the fame of the artist, age of the painting, and its uniqueness [43]. Picasso's paintings have repeatedly surpassed auction records, with his painting *Les femmes d'Alger (Version 'O')* selling for $179 million at an auction in 2015, and many of Picasso's other paintings are on the list of the most expensive paintings of all time [44].

Surprise in artwork can be described as the unexpectedness of subject or emotion in it. This is closely related to novelty, as any such surprise would contribute a new element to it. For example, the painting *The Young Ladies of Avignon* by Picasso was considered surprising when it was first released, as it ushered in the era of cubism and is considered as a proto-cubist painting [45].

Artistic styles are very heavily dependent on the type of art the painter is exposed to growing up. This can also be influenced by the prevalent styles of painting and the creation of a new style. Influence can again be measured by looking at the ratio of the similarity of an object with those that come after it to those that came before. For example, Picasso grew up in France and studied Symbolism, which helped him develop imagery drawn from mythology and dreams. The cubist movement was also drawn from tribal and archaic influences that he learned from Cézanne and Rousseau. There was a wave of cubism after Picasso became famous and was thus very influential [46].

Correctness is irrelevant for art, as any form of art might be appreciated by a certain segment of people in the appropriate scope. For example, abstract art is heavily critiqued in some circles and appreciated in others and cannot be called relevant.

Comprehensibility would be a scope-dependent factor considering that some types of paintings are appreciated by certain types of people, while others are appreciated by a completely different sector. For example, some people might appreciate abstract art, while others might not. This does not indicate whether it is creative.

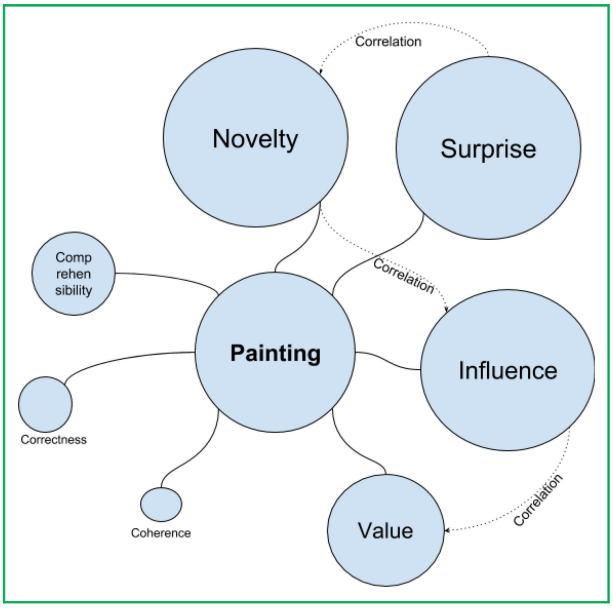

Figure 4

Relationship and the importance of the different dimensions in evaluating the creativity of painting and artwork.

The relationship between the different dimensions in the domain of artwork and painting to evaluate creativity is shown in **Figure 4**.

Food

Food in its crudest definition is defined as any substance consumed to provide nutritional support for an organism. However, from a creative perspective, we can define food as any combination of one or more ingredients that when combined using some sequence of actions represents a final product [47]. The only constraint is the fact that the final product needs to be edible to satisfy the nutrition precondition of food. We can use the example of *Ample Hills Creamery*, a cookbook by Jackie Cuscuna and Brian Smith, which combines different combinations of ingredients to create varieties of ice cream.

Novelty in the context of food can be evaluated by looking at the different ingredients in the dish, as well as their usual cooccurrence along with the techniques applied on each of them [48]. Hence, ingredients and techniques could be used as pairwise features for this purpose. This would help us define the posterior probability of any given artifact. For example, the Ample Hills Creamery website contains a flavor called "The Munchies," which is a pretzel-infused ice cream with clusters of Ritz crackers, potato chips, and pretzels. These items would not be novel as such, but the technique applied over them to convert them into an

ice cream would inherently make them novel. Hence, just the use of ingredients as a factor is not considered sufficiently novel, but the final food product is.

Value of a particular dish can be seen as the perception of taste of any particular dish [47]. Thus, the tastier dish would be considered more valuable or attractive compared with another dish. This could be approximated using the sales of a particular item in a restaurant. It has also been found that taste is almost a linear combination of the individual items in it. Taste fails to take into account the technique or preparation process applied on the item. Thus, this could be replaced with a weighted sum of the ingredients and their corresponding techniques. In the previous example of "The Munchies" ice cream, the artifact would be considered valuable along both the filters listed here. The item as such is a popular selling item on the Ample Hills Creamery website, and the item-technique combination would also lead to a high score as such.

Surprise in a dish is primarily dictated by the unexpectedness of any particular combination of texture, appearance, and flavor in that class of items. For example, in the case of "The Munchies," the presence of salty and crunchy flavors would be considered as surprising by the consumer. This would drive the value as well as the novelty of any particular item, as our minds react favorably to new experiences in most cases.

The influence of any dish or culinary effort can be estimated by looking at the prevalence of any particular ingredient, technique combination before and after the introduction of that particular dish. For example, the introduction of fried chicken in the modern-day format by Kentucky Fried Chicken has ensured that it has become a popular fast-food item around the world.

Coherence can be weakly defined in the context of food when we consider the act of eating to be a temporal experience that starts with the appearance, both visual and smell based, followed by texture we experience on consuming it, and finally followed by the flavor of that dish. Any incoherence here can have a significant impact on the overall surprise of the dish. For example, the Durian fruit in Malaysia has a very horrid smell but is very tasty at the same time.

Incorrectness in the context of food can be defined as the forceful presence of some techniques with certain ingredients, i.e., the inability of each ingredient to be prepared using any given technique. However, absence of such techniques would be considered correct. This is dictated by the inherent properties of that particular ingredient. For example, egg can be prepared using a set combination of techniques, and thus a recipe asking for egg to be, say, aerosolized would be incorrect.

Comprehensibility in the context of food would be irrelevant, except for the boundary scenario of a person with allergies, which would limit the ability of that person to eat some particular classes of food.

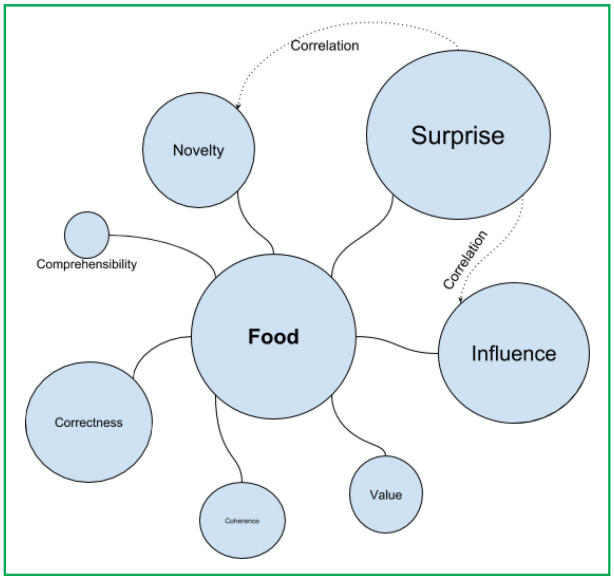

Figure 5

Relationship and the importance of the different dimensions in evaluating the creativity of recipe creation in food.

Overall, taste is very subjective, and hence, scope would again play a very important role. This becomes important, especially when some people are open to trying new things while others are not. This would greatly skew the results of any survey on humans.

Let us take the example of a simple and standard (most frequently occurring) recipe of mashed potatoes. It is edible and is hence correct and would also be comprehensible to most people. However, the recipe is a very common thing and is hence not novel. In addition, the taste of the final item and hence the value if prepared perfectly would be good but not high as can be estimated by looking at the individual ingredients as we mentioned earlier. Finally, the surprise and influence would be minimal, as the experience would be uneventful. Hence, we can conclude that the recipe is not creative.

The correlation among the different dimensions for evaluating creativity in making food recipes is shown in **Figure 5**.

Industrial design

Industrial design is defined as the art or process of designing manufactured items. This includes the design of daily-use items such as phones and watches, as well as the design of industrial machines and goods such as scalpels, hammers, and machinery. Let us take the example of the Apple iMac G3, with its cherry drop design, which was considered very innovative and unique when it was initially launched.

Novel designs are characterized by analyzing the difference between designs of various items in the domain

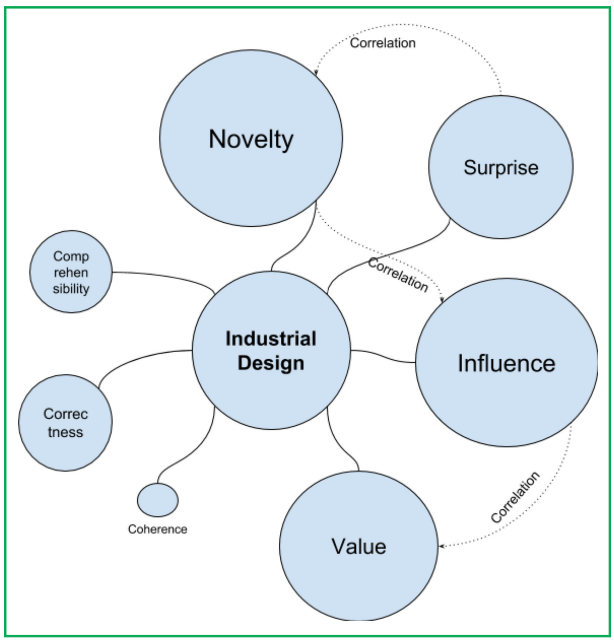

Figure 6

Relationship and the importance of the different dimensions in evaluating the creativity of industrial design.

of that particular item and finding the similarity between their designs. Here, we want to analyze the design in three-dimensional (3-D) space as any of the dimensions could introduce a uniqueness in our design. Considering the iMac G3, its translucent and cherry drop design was considered very unique, as most of the personal computers of that era were drab and built primarily for business use.

The value of a design can be thought of as the extra percentage value that we can attribute to a particular product just because of its design, i.e., how much would a customer pay for this product, or how many more customers would buy this product compared with another product with similar specifications. This was very evident in the case of the iMac, as more than 500,000 units were sold in the first year, increasing its market share in the PC market by 2% [49].

Surprise can be a very important element of design, as it lends a level of playfulness to it, increasing the product recall and recognition [50, 51]. We could define surprise in this domain as an emotional response to incongruent, unfamiliar, unexpected, or sudden stimuli. Here, the stimuli are usually visual or tactile as they are the primary ways in which you interact with an item and the values are measured by looking at the historical data and looking for similar combinations of elements in a similar or same class of products. This instantly lends novelty to the design, as the surprise element would have to be novel for it to elicit that particular response [52]. For example, the bold colors and translucent body of the iMac were surprising for most users, as they were used to seeing drab black and white boxes.

Influence would again be a very strong parameter for the design by looking at how heavily a particular design cue is used in the products that follow a particular design. Here, we would again see a clear example in the iMac G3, as it led to an increase in the use of translucent bodies not only within the PC industry but also in other industries. It became sufficiently famous to force Apple to not use that particular design anymore due to its commonplace nature.

Coherence would be irrelevant in this domain, as it is a completely time-independent domain.

Correctness of a particular design can be defined by whether it can be manufactured, however complex the process may be. The boundaries of this domain have been heavily pushed recently, considering the advent of 3-D printing technologies, which allow designers to build much more different and novel products than they previously did.

The comprehensibility of a design could be defined by the value attributed by people of a particular scope.

The relation among the different dimensions used in evaluating the creativity in design process is shown in **Figure 6**.

Computational measurements

In this section, we provide a broad approach to quantitatively measure the different dimensions. These measures provide a computational way to implement these dimensions and evaluate a creative system.

1) *Novelty:* Novelty of an artifact is measured as the divergence of the artifact against all the artifacts in that domain. There has to be novelty in the presence of surprise; however, the other way around is not necessary.
2) *Value:* Measurement of value directly captures the influence of the artifact in that domain. This is usually domain dependent and is a direct utility value in that domain for that artifact.
3) *Surprise:* In the temporal domain, surprise is the unexpectedness of a certain artifact point compared with the points preceding it:

$$\text{Surprise} = 1 - P(\text{Point}_t|\text{Point}_0{}^{(t-1)}).$$

In the time-independent domain, surprise is the uniqueness of every atomic element and their combination in the artifact from every other atomic element and combinations in the domain. The definition of atomic element of an artifact is domain dependent.

1) *Influence:* Influence is measured as the novelty of the artifacts that are generated after the given artifact is generated.

$$\text{Influence} = \frac{\sum_{New} Similarity * Novelty\ of\ artifacts}{\sum_{All\ time} Similarity * Novelty\ of\ artifacts}.$$

2) *Coherence:* Coherence of the artifact can be computed as the divergence from predefined rules of the domain.

The total computational score for creativity can be defined as

$$\text{Creativity score} = (\text{Novelty} + \text{Value} + \text{Surprise} + \text{Influence} + \text{Coherence}) * \text{Correctness} * \text{Comprehensibility}$$

where correctness and comprehensibility are necessary binary values.

Conclusion

In this paper, we have formulated a framework for evaluating whether a system could be categorized as computationally creative. This framework grounds as a fundamental work for building different creative systems. We proposed a set of seven dimensions that account for the necessary and sufficient conditions for evaluating creativity, such as novelty, value, surprise, influence, coherence, correctness, and comprehensibility. We provided an initial generic definition of these seven dimensions and justify their need in evaluating creativity. We further studied the adaptation of these dimensions in three time-dependent domains, such as story writing, music composition, and movie creation, and in three time-independent domains, such as painting, cooking, and industrial designing. We provided an abstract computational form of measuring these dimensions in different domains and provide an eventual creativity score for a system. We believe this research is the essential missing link in the domain of computational creativity and could foster active further research in building creative systems.

References

1. A. Brandt and D. Eagleman, *The Runaway Species: How Human Creativity Remakes the World*. Edinburgh, U.K.: Canongate, 2017.
2. R. J. Sternberg ed., *Handbook of Creativity*. Cambridge, U.K.: Cambridge Univ. Press, 1999.
3. M. Karwowski, J. Dul, J. Gralewski, et al., "Is creativity without intelligence possible? A necessary condition analysis," *Intelligence*, vol. 57, pp. 105–117, 2016.
4. E. Jauk, M. Benedek, B. Dunst, et al., "The relationship between intelligence and creativity: New support for the threshold hypothesis by means of empirical breakpoint detection," *Intelligence*, vol. 41, no. 4, pp. 212–221, 2013.
5. J. A. Plucker and J. S. Renzulli, "Psychometric approaches to the study of human creativity," in *Handbook of Creativity*. New York, NY, USA: Cambridge Univ. Press, 1999, pp. 35–61.
6. J. A. Plucker and A. Esping, "Intelligence and creativity: A complex but important relationship," *Asia Pac. Educ. Rev.*, vol. 16, no. 2, pp. 153–159, 2015.
7. M. I. Jordan and T. M. Mitchell, "Machine learning: Trends, perspectives, and prospects," *Science*, vol. 349, no. 6245, pp. 255–260, 2015.
8. G. Kasparov, *Deep Thinking: Where Machine Intelligence Ends and Human Creativity Begins*. New York, NY, USA: Public Affairs, 2017.
9. J. Domingues, *Social Creativity, Collective Subjectivity and Contemporary Modernity*. New York, NY, USA: Springer, 2000.
10. M. A. Boden, *The Creative Mind: Myths and Mechanisms*. New York, NY, USA: BasicBooks, 1991.
11. IBM Corporation, IBM Watson Beat, 2017. [Online]. Available: https://www.ibm.com/case-studies/ibm-watson-beat
12. Google, Inc., Magenta. [Online]. Available: https://ai.google/research/teams/brain/magenta/
13. Sony Computer Science Laboratories, Inc., Flow Machines. [Online]. Available: http://www.flow-machines.com/
14. P. Jain, P. Agrawal, A. Mishra, et al., "Story generation from sequence of independent short descriptions," in *Proc. SIGKDD Workshop on Mac. Learn. Creativity*," 2017.
15. N. Panwar, S. Khare, N. Gantayat, et al., "Movie amalgamation using neural imitation; Visualizing the movie while reading a book," in *Proc. SIGKDD Workshop Mach. Learn. Creativity*," 2017, pp. 1–8.
16. V. Munigala, S. Tamilselvam, and A. Sankaran, "Let me convince you to buy my product. . .: A case study of an automated persuasive system in fashion," in *Proc. SIGKDD Workshop Mach. Learn. Creativity*, 2017, pp. 1–4.
17. R. W. Weisberg, *Creativity: Beyond the Myth of Genius*. San Francisco, CA, USA: Freeman, 1993.
18. R. J. Sternberg and T. I. Lubart, *Defying the Crowd: Cultivating Creativity in a Culture of Conformity*. New York, NY, USA: Free Press, 1995.
19. R. W. Weisberg, "On the usefulness of "value" in the definition of creativity," *Creativity Res. J.*, vol. 27, no. 2, pp. 111–124, 2015.
20. D. Shrivastava, S. Chemmengath, A. Laha, et al., "A machine learning approach for evaluating creative artifacts," in *Proc. SIGKDD Workshop Mach. Learn. Creativity*, 2017, pp. 1–9.
21. M. L. Maher, "Evaluating creativity in humans, computers, and collectively intelligent systems," in *Proc. 1st DESIRE Netw. Conf. Creativity Innov. Des.*, Lancaster, U.K., 2010, pp. 22–28.
22. C. França, L. Goes, A. Amorim, et al., "Regent-dependent creativity: A domain independent metric for the assessment of creative artifacts," in *Proc. 7th Int. Conf. Comput. Creativity*, 2016, pp. 68–75.
23. A. Elgammal and B. Saleh, "Quantifying creativity in art networks," in *Proc. Sixth Int. Conf. Comput. Creativity*, Jun. 2015, pp. 39–46.
24. P. Karampiperis, A. Koukourikos, and E. Koliopoulou, "Towards machines for measuring creativity: The use of computational tools in storytelling activities," in *Proc. IEEE 14th Int. Conf. Adv. Learn. Technologies*, Jul. 2014, pp. 508–512.
25. M. Blum, "Great geek debates: The Lord of the Rings vs. Harry Potter," *Wired*, Sep. 4, 2009. [Online]. Available: https://www.wired.com/2009/09/great-geek-debates-the-lord-of-the-rings-vs-harry-potter/
26. M. P. Fay, "Driving story generation with learnable character models," Ph.D. dissertation, Comput. Sci. Artif. Intell. Lab., Massachusetts Inst. Technol., Cambridge, MA, USA, 2014.
27. J. della Quercia, "6 great novels that were hated in their time," Jul. 2010. [Online]. Available: http://www.cracked.com/article_18645_6-great-novels-that-were-hated-in-their-time.html
28. R. S. De Vera, "How Harry Potter changed the world," Inquirer.net, Jul. 16, 2011. [Online]. Available: http://lifestyle.inquirer.net/5583/how-harry-potter-changed-the-world/
29. "The cultural impact of Harry Potter," Nov. 2014. [Online]. Available: http://www.ericrettberg.com/datacultj1/longer-essays/the-cultural-impact-of-harry-potter/
30. H. Kozlowska, "Can 'Harry Potter' change the world?," *New York Times*, Sep. 17, 2014. [Online]. Available: https://op-talk.blogs.nytimes.com/2014/09/17/can-harry-potter-change-the-world/
31. "Coherence," [Online]. Available: https://literaryterms.net/coherence/
32. M. Slaney, K. Weinberger, and W. White, "Learning a metric for music similarity," in *Proc. Int. Symp. Music Inf. Retrieval*, 2008, pp. 1–6.
33. M. W. Park and E. C. Lee, "Similarity measurement method between two songs by using the conditional Euclidean distance," *WSEAS Trans. Inf. Sci. Appl.*, vol. 10, no. 12, pp. 381–388, 2013.

34. J. J. Aucouturier and F. Pachet, "Finding songs that sound the same," in *Proc. IEEE Benelux Workshop Model Process. Coding Audio*, Nov. 2002, pp. 1–8.
35. O. Seguna, "Buffalo Soldier (Bob Marley & The Wailers) sounds like The Tra-La-La Song (Banana Splits)," Samethattune.com, Jul. 21, 2014. [Online]. Available: http://www.samethattune.com/buffalo-soldier-bob-marley-wailers-sounds-tra-la-la-song-banana-splits
36. IMDb, *Fight Club*, Oct. 15, 1999. [Online]. Available: https://www.imdb.com/title/tt0137523/
37. "Components of film for study and analysis." 2019. [Online]. Available: https://www.westga.edu/~bjames/ComponentsofFilm.htm
38. "Fight Club: The beauty of sound design," vimeo.com, Mar. 12, 2017. [Online]. Available: https://vimeo.com/208025022
39. Rotten Tomatoes.com. [Online]. Available: https://www.rottentomatoes.com/
40. Box Office Mojo, "Fight Club." [Online]. Available: http://www.boxofficemojo.com/movies/?id = fightclub.htm
41. B. Saleh, K. Abe, R. S. Arora, et al., "Toward automated discovery of artistic influence," *Multimedia Tools Appl.*, vol. 75, no. 7, pp. 3565–3591, 2016.
42. "Pablo Picasso," 2000. [Online]. Available: http://www.mcs.csueastbay.edu/~malek/Artfolder/Pablo.html
43. "A simple formula for pricing artwork," Artists Network. 2009. [Online]. Available: https://www.artistsnetwork.com/artist-life/a-simple-formula-for-pricing-artwork/
44. H. Timmons, "This $179 million Picasso is now the most expensive painting ever sold at auction," qz.com, May 11, 2015. [Online]. Available: https://qz.com/402938/this-179-million-picasso-is-now-the-most-expensive-painting-ever-sold-at-auction/
45. "30 most expensive paintings of all time-Inspiring showcase," 2018. [Online]. Available: http://webneel.com/webneel/blog/30-most-expensive-paintings-all-time-inspiring-showcase
46. "Who were Pablo Picasso's influences?" 2015. [Online]. Available: https://www.reference.com/art-literature/were-pablo-picasso-s-influences-907ff3e1db82c8e3
47. F. Pinel and L. R. Varshney, "Computational creativity for culinary recipes," in *Proc. Extended Abstr. Human Factors Comput. Syst.*, Apr. 2014, pp. 439–442.
48. N. Shao, P. Murali, and A. Sheopuri, "New developments in culinary computational creativity," in *Proc. Int. Conf. Comput. Creativity*, 2014, pp. 324–327.
49. J. Swartz, "Apple Profit Up – iMac Sales Cited," SFGate, Jan. 14, 1999. [Online]. Available: https://www.sfgate.com/business/article/Apple-Profit-Up-iMac-Sales-Cited-2952619.php
50. G. D. Ludden, B. M. Kudrowitz, H. N. Schifferstein, et al., "Surprise and humor in product design," *Int. J. Humor Res.*, vol. 25, pp. 215–374, 2012.
51. E. R. R. Ramírez, "Industrial design strategies for eliciting surprise," *Des. Stud.*, vol. 35, no. 3, pp. 273–297, 2014.
52. G. D. Ludden, H. N. Schifferstein, and P. Hekkert, "Surprise as a design strategy," *Des. Issues*, vol. 24, no. 2, pp. 28–38, 2008.

Received February 6, 2018; accepted for publication February 7, 2019

Shavak Agrawal *Birla Institute of Technology and Science, Pilani 333031, India (shavak1997@gmail.com).* Mr. Agrawal received a dual graduate degree in computer science from Birla Institute of Technology and Science, Pilani, India. He is currently working with Microsoft, Redmond, WA, USA, as a Data Scientist with the Bing team. Prior to joining Microsoft, he was an intern with Flipkart and IBM Research with their Machine Learning teams. His research interests include machine learning, natural language processing, and artificial intelligence.

Anush Sankaran *IBM Research, Bangalore 560045, India (anussank@in.ibm.com).* Dr. Sankaran received a B.Tech. degree (Gold Medal) in computer science from the Coimbatore Institute of Technology, Coimbatore, India, in 2010, and a Ph.D. degree in biometrics and machine learning from the IIIT Delhi, New Delhi, India. He is a Researcher with IBM Research AI, Bangalore, India. He is currently primarily leading the technical efforts in two projects—IBM Deep Learning IDE and Machine Learning for Creativity. His research interests include deep learning, image processing, human cognition, and their applications. He was the recipient of the TCS Ph.D. Research Fellowship from 2010 to 2015. He has authored more than 20 peer-reviewed conferences and journals and is also the recipient of the Best Poster Awards in the IEEE BTAS 2013 and the IEEE IJCB 2014.

Anirban Laha *IBM Research, Bangalore 560045, India (anirlaha@in.ibm.com).* Mr. Laha received an M.S. degree in computer science and automation from Indian Institute of Science, Bangalore, India. He is currently working with the AI Tech Group, IBM Research AI, Bangalore. His research interests include machine learning/deep learning in natural language processing. He has been working with the natural language generation project, IBM, since the last two years and has published papers on abstractive summarization both from unstructured and structured data in top conferences like NeurIPS, ACL, and NAACL. At IBM, he has also worked on argumentation mining (IBM Project Debater), which received news coverage worldwide recently because of a live machine-versus-human debate. He was also briefly associated with machine learning for creativity project at IBM (SIGKDD workshop), during which he has worked on story generation. Before joining IBM, he had spent some time as an Applied Scientist in Microsoft and SDE at Amazon.com. He had given talks on NLG, particularly NLG from structured data in multiple venues. The full list of his publications and talks are available in his website.

Saneem Ahmed Chemmengath *IBM Research, Bangalore 560045, India (saneem.cg@in.ibm.com).* Mr. Chemmengath received an M.Sc. degree in feature selection algorithms (by research) from IISc, Bangalore, India. He is a Research Software Engineer with IBM Research, Bangalore. In IBM, he worked on various projects including recommender systems, image detection and classification, data interpolation, time series data analysis, etc. He has been working in machine learning and related areas for the past nine years and has authored or coauthored papers in top machine learning conferences such as NeurIPS, UAI, and WACV. His research interests include core machine learning, statistics, and deep learning for computer vision.

Disha Shrivastava *IBM Research, Bangalore 560045, India (dishu.905@gmail.com).* Ms. Shrivastava received a B.E. degree in electronics and communication engineering from Birla Institute of Technology, Mesra, India, in 2012, and a master's degree in computer technology from Indian Institute of Technology, Delhi, India, in 2016. She has been working as a Research Engineer with the Reasoning group in the AI Science division of IBM Research, Bangalore, India, since August 2016. She is currently working on complex question answering systems which involve reasoning. Her research interests include deep learning, natural language processing, computer vision, and large-scale distributed machine learning.

Karthik Sankaranarayanan *IBM Research, Bangalore 560045, India (kartsank@in.ibm.com).* Dr. Sankaranarayanan received a Ph.D. degree in computer science from The Ohio State University, Columbus, OH, USA, where he was a U.S. Department of Energy Graduate Fellow between 2009 and 2011. During his Ph.D. studies, he worked with the intersection of machine learning and computer vision for real-time object localization and tracking for surveillance applications. He is a Research Staff Member with the Cognitive Solutions Department, IBM Research, Bangalore, India. His research interests include statistical machine learning applied to natural language processing and computer vision. His current work is focused around question-answering from domain-specific knowledge graphs, natural language generation, and computational creativity working primarily with IBM's Watson Group. Previously, he worked with IBM's Global Business Services unit for building ticket analytics applications to improve agent productivity using novel ensemble text clustering approaches as part of the Catapult project. For this work, he was the recipient of an IBM Corporate Award in 2013 and an Outstanding Technical Accomplishment Award in 2012.