A big data approach to computational creativity: The curious case of Chef Watson

L. R. Varshney
F. Pinel
K. R. Varshney
D. Bhattacharjya
A. Schörgendorfer
Y.-M. Chee

Computational creativity is an emerging branch of artificial intelligence that places computers in the center of the creative process. Broadly, creativity involves a generative step to produce many ideas and a selective step to determine the ones that are the best. Many previous attempts at computational creativity, however, have not been able to achieve a valid selective step. This paper shows how bringing data sources from the creative domain and from hedonic psychophysics together with machine learning and data analytics techniques can overcome this shortcoming to yield a system that can produce novel and high-quality creative artifacts.
To demonstrate our data-driven approach, we developed and deployed a computational creativity system for culinary recipes and menus, Chef Watson, which can operate either autonomously or semiautonomously with human interaction. We present the basic system architecture, data engineering, and algorithms that are involved. Experimental results demonstrate the system passes the test for creativity based on the consensual assessment technique, producing a novel and flavorful recipe. Large-scale deployments are also discussed.

1 Introduction

Creativity is defined as the generation of a product or service that is judged to be novel and also to be appropriate, useful, or valuable by a knowledgeable social group [1], and is often said to be the pinnacle of intelligence [2]. Creativity is the basis for "disruptive innovation and continuous re-invention" [3]. Due to greater competitiveness in global markets in all industries, there is a need to make product/service development cycles more efficient. Given the limited availability of human creativity resources, it is important to develop intelligent systems and technologies for greater creativity, either operating autonomously or in collaboration with people. Furthermore, such technologies may provide insight into human creativity itself.

Computational creativity is an emerging branch of artificial intelligence that places computers in the center of the creative process [2, 4–8], concerned with machine systems that produce novel and high-quality artifacts for people. In this paper, we focus primarily on culinary recipes, which include both the set and quantities of ingredients to be used as well as the methods and procedures of preparation. We also discuss menus, which can be viewed as sequences of culinary recipes. Cutting-edge chefs must have an impeccable culinary technique, but become renowned for their creative recipe designs [9]; hence, it is clear that culinary is an appropriate creative domain of study.

The broad technological contributions of this paper, such as system architecture and data-driven analytic methods, are grounded in a computational creativity system, called Chef Watson, we built and have deployed at scale to create culinary dishes that have been served to thousands of people, with largely positive assessments. Using the consensual assessment technique (CAT) [10–12] (the standard evaluation method in creativity research), we show that artifacts produced by the system are rated as very creative by domain experts, more so than similar human-created artifacts.

The general system structure for our computational creativity system was first presented in [13] and related

Digital Object Identifier: 10.1147/JRD.2019.2893905

ideas on the mixed initiative approach for semiautonomous operation were first presented in [14], but this paper represents an integration of and significant expansion on the conference papers. Formal definitions of creativity, the description of how our system architecture is inspired by stages in human creativity, descriptions of natural language processing and the integration of data from chemistry and psychophysics, and menu design have not been published previously. Moreover, experimental system deployment results, both to domain experts and in large-scale practice, have not previously been presented. Concurrent and subsequent to our initial work, there have been further culinary creativity systems either restricted to specific dishes, such as soups and drinks, [15, 16] or based very closely on our approach [17].

We focus on a specific domain because at least with human creativity, there is substantial evidence that this cognitive ability has some domain-specificity [18], in the sense that a good engineer may not be a good poet. Notwithstanding, psychometric testing has indicated correlations of ability that allow domains to be grouped into categories: expressive creativity (visual arts, writing, humor); performance creativity (dance, drama, music); and scientific creativity (invention, science, culinary), with architecture not related to any of these [19]. Most previous computational creativity systems have focused on either expressive or performance creativity [6, 20, 21], whereas we consider a form of scientific creativity (but see [22, 23]).

We believe that scientific creativity has a major advantage for machines as compared to other categories of creativity: By applying data-driven learning techniques to the vast amount of technical information available and understandable by a computer, evaluating the quality of the products generated in the creative process becomes more achievable.

Indeed, many previous attempts at computational creativity did not achieve a valid selective step [1], as we do here; rather, such systems only generated artifacts and required a human to select the best ones. A central contribution of this paper is in showing how bringing together data sources from the creative domain and from hedonic psychophysics can overcome this shortcoming. It is worth noting that selection based on supervised learning algorithms trained on complete artifacts is not appropriate for creativity, as it is for search, since the goal is to create novel artifacts each time rather than to find existing ones. Basic models of human perception, applied at the constituent part level, and techniques for building up predictions of quality and novelty for whole artifacts is critical to our approach. Predicting the perceptual properties of completely new whole artifacts from data on parts and combining rules is also a novel contribution to computational creativity.

We operate with the so-called four Vs of big data: variety, veracity, volume, and velocity. For culinary creativity, we draw on a variety of datasets: large repositories of existing recipes as inspiration, chemoinformatics data to understand food at the molecular level, and hedonic flavor psychophysics data to predict which compounds, ingredients, and dishes people will like and dislike. Since these datasets arise from noisy sensors such as gas chromatography and from noisy data preprocessing steps such as natural language processing, algorithms must be robust to issues of veracity, cf. [24]. These datasets are used to develop generative algorithms that intelligently produce thousands or millions of new ideas from the recipe design space, which, for particular dishes and regional cuisine influences, has a size in excess of 10^{24} just for listing ingredients. The large volume of intermediate ideas generated by the system must then be evaluated to select the best ones. Evaluative metrics are based on principled models of human perception structured according to ideas from neurogastronomy and derived from recipe, chemical, and psychophysical data. The information-theoretic functional *Bayesian surprise* [25–28] is used to measure attraction of human attention and novelty. Since the system is meant to support real-time interaction with human creators and make the product design cycle faster, the system must compute with the velocity required for human–computer interaction.

Again note that the computer-generated culinary recipe design problem is not just one of locating existing recipes and recommending them [29], but of creating new ones. It is different from web search and product recommendation, and is truly a part of an emerging computing paradigm distinct from fields such as information retrieval and statistical learning.

Our computational creativity system operates in stages that are modeled after stages in human creativity [1, 30]: find the problem, acquire relevant knowledge, gather related information, incubate, generate ideas, combine ideas, select best ideas, and externalize ideas. The staged approach not only leads to modular system design, but also improves computer–human interaction when operating semiautonomously. Developing algorithms and systems is essential, but interaction with and presentation of results to users in ways that allow them to trust insights are also important [31]. In a semiautonomous mode, our system takes a mixed-initiative approach: The human and computer have a creation conversation in which each contributes ideas [32], rather than the computer acting as a nanny, coach, or pen pal for the human creator [33]. This human-inspired usage flow (and modular system architecture that supports it) is a central contribution of this paper.

Building on individual recipe design, complete menus can also be created using ideas from topic modeling. By using the principle of variety across dishes in a menu, measured using a stochastic distance function, input parameters for dish design may be generated and selected.

The remainder of the paper discusses details of datasets, data engineering, system architecture, data analytics algorithms, and results indicating that the system is indeed creative. Although we use culinary recipes as the example domain herein, the basic concepts are generally applicable to data-driven approaches to computational creativity in any domain.

2 Background on creativity

This section reviews past work in creativity research, which puts forth definitions of creativity, methods to evaluate creativity, and psychological models of how human creativity is thought to proceed. It also discusses models for predicting creativity assessments.

2.1 Defining creativity

Deductive and inductive reasoning are easily assessed since there is often ground truth, but not so with creativity. Creativity involves reasoning about things that have never previously been imagined.

One definitional approach is to list several properties of a creative output, such as being novel, being useful, rejecting previously held ideas, and providing clarity [34–36]. Viewing creativity as a relationship between the creator/creation and an observer [20], if a human evaluator deems something creative, one can say it is creative [37]; in this definition, creativity is only meaningful in the presence of an audience perceiving the creation. To formalize in a way that can be operationalized, we use a definition of creativity from human creativity research.

Definition 1 ([1]). *Creativity is the generation of a product that is judged to be novel and also to be appropriate, useful, or valuable by a suitably knowledgeable social group.*

This definition describes two dimensions of creativity: novelty and quality. It also specifies that creativity is fundamentally socially constructed; a computational creativity system has little meaning in a closed universe devoid of people.

An alternate definition for computational creativity would be by analogy to the Turing test—a system is creative if it produces artifacts indistinguishable from those produced by humans or having as much aesthetic value as those produced by humans [21]. We do not use this definition.

2.2 Assessing creativity

The most common way to assess creativity of an artifact under Definition 1 is the CAT [10–12], where the creativity of an artifact is rated by two or more experts in the field. The measured creativity is the average rating of the judges. Although it may seem this methodology is too subjective, many studies have demonstrated that ratings of experts are generally highly correlated, yielding good inter-rater reliability [38–41]. In contrast, novice ratings are not highly correlated, so novices should not be used for the CAT [42]; for example, [41] found expert ratings had correlation coefficient 0.93, whereas novice ratings only had correlation coefficient 0.53. Although CAT is an expensive and time-consuming creativity test to administer—requiring finding and engaging a panel of experts—it is the standard, most widely used evaluation method in creativity research [1].

Another operational test for creativity based on Definition 1 is to measure the impact of a created artifact in large-scale deployment. For example, the number of citations for scientific papers is often used to measure their creativity [1, 43].

2.3 Models for predicting human creativity assessment

Our view is that a computational creativity machine without the capability to evaluate its potential outputs is incomplete because generation and assessment must coexist for proper functioning. In the same way information cannot be encoded without a model of the receiver decoding that information [44], artifacts cannot be created without a model of human evaluators.

The lack of evaluative and selective ability has been a primary criticism of many previous computational creativity systems. Consider the computational creativity system for visual art AARON. AARON generates 150 pieces a night, but Cohen decides which 5 to print by viewing them all: "AARON doesn't choose its own criteria for what counts as a good painting. . . To be considered truly creative, the program would have to develop its own selection criteria; Cohen was skeptical that this could ever happen" [1, p. 146]. A computational creativity system for mathematical proofs AM suffers in the same way: "The first and biggest problem is that AM generates a huge number of ideas, and most of them are boring or worthless; Lenat has to sort through all of the new ideas and select the ones that are good" [1, p. 147].

We will fix the operational view of Definition 1 [35] and develop a data-driven evaluative/selective component as part of a computational creativity system, a key contribution of this paper. Note that such a component is not the final arbiter of creativity as that is a human determination but is useful nonetheless.

2.4 Stages of human creativity

We use stages of the human creative process to guide our computational creativity system design. This will lead to a

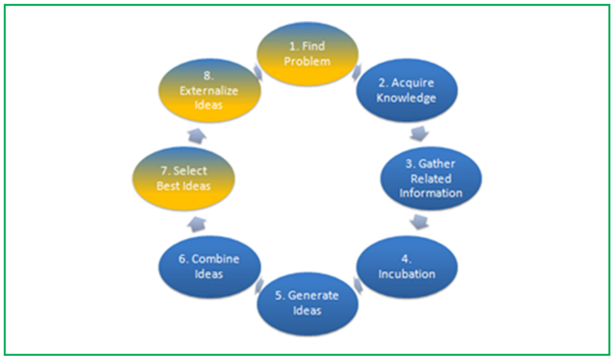


Figure 1

Stages in the human cognitive process of creativity [1], which are useful to understand for system and human–computer interaction design. In computational creativity, blue/gold stages may have significant human–computer interaction, whereas blue stages may involve autonomous computer operation.

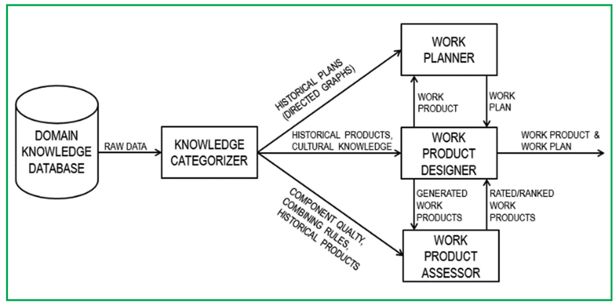


Figure 2

Block diagram of the computational creativity system that produces a work product and a work plan. The three main data-driven analytic components are on the right side. Arrows indicate the main information flows among components.

modular system architecture. Note that stages are not always followed sequentially by human creators; there can be backtracking and jumping around.

When the system operates in a semiautonomous mode, the computer acts as a colleague or partner to the human, and so following the natural human process improves computer–human interaction. Indeed there is an emerging consensus that even in purely human contexts, interacting groups are more creative than individuals, hence the value of computer–human interaction.

We review stages of creativity delineated by Sawyer [1], given in **Figure 1** (which also depicts stages where human interaction is most useful).

1) *Find the problem*: For ill-defined problems such as creating new products, the first step is to actually identify and formulate the problem using *divergent thinking*. Exceptional creativity is more likely when people work in areas where problems are not specified *a priori*.
2) *Acquire knowledge*: The second stage is to learn everything there is to know about the problem, especially in terms of past creative artifacts. Without knowing what has already been done, there is no inspiration set or way to judge novelty. Since it is impossible to be creative without first internalizing the creative domain, data intake is necessary for creativity.
3) *Gather related information*: Besides learning about past examples of creative artifacts within the domain, it is important to absorb information from a wide variety of other sources, so as to link new information with existing problems.
4) *Incubation*: In human creativity, it is important to give the mind time to process gathered information, and to let the subconscious search for new and appropriate combinations.
5) *Generate ideas*: After incubation, the mind is ready to generate ideas. The generation of ideas is often considered the key step in creativity and is rather different from other forms of reasoning such as induction or deduction.
6) *Combine ideas*: There is often value in cross fertilization of ideas across problems and domains. Approaches to combining concepts across domains include attribute inheritance, property mapping, and concept specialization.
7) *Select best ideas*: After a new idea or insight emerges, the creator must determine whether it really is good. This stage is sometimes referred to as *convergent thinking*. In two-stage models, convergent thinking follows divergent thinking. The evaluation stage is fully conscious, drawing on much domain knowledge to assess novelty and quality.
8) *Externalize the idea*: Successful creation requires not only ideas but also execution of those ideas by identifying necessary resources to make them successful, forming plans for implementing the ideas, etc. This final stage is mostly conscious and directed.

This staged view of creativity forms the starting point for system architecture design.

3 New creativity system architecture

In this section, we propose a new system architecture for computational creativity that includes a data absorption and organization component, as well as a data-driven assessment component that models human perception.

A block diagram is given in **Figure 2**, with three main computational components: a work planner, a work product designer, and a work product assessor, which interact to output a work product and work plan. These components are fed by a domain knowledge database and knowledge categorizer. It is important to note that in our system, the work planner and the work product assessor do not directly interact, but only do so through the work product designer.

The domain knowledge database represents information collected on the creative field of interest, including information on styles, tastes, constituents, combinations, evolution, regionality, culture, and methods of preparation. It also includes a repository of existing artifacts that have been deemed creative by human audiences. This knowledge is resolved and organized by the knowledge categorizer. It is the source of data that the designer, planner, and assessor components draw from. Information from related but distinct fields to the creative domain is also kept in the database. As we will see, significant data engineering and natural language processing is required for creating and using this knowledge database.

The designer generates new ideas for artifacts. The assessor evaluates those potential design ideas for creativity, and the planner determines the methods by which the ideas could be externalized. All three components take input from the categorized database: the designer to draw inspiration for new ideas, the planner to learn from extant methods of preparation, and the assessor to evaluate a design idea against the repository of existing artifacts as well as against properties of constituents and combinations for creativity.

The designer is the lead component of the system. Although it is possible to use human-like generative processes, a generation or design procedure wholly different from the human approach is valuable precisely because it creates things different from what a human would. It may have different kinds of "illusions" or "blindspots" than a human and, thus, would supplement and support human creativity. These differences enlarge the hypothesis space and allow the machine to break new creative ground. There are several possible algorithmic approaches to generation.

The assessor component models human perception, taste, and culture using data-driven models. It examines creative ideas produced by the designer along two main dimensions: novelty and quality. These metrics are defined on the basis of datasets within the creative domain, information related to the domain, and experimental data from hedonic psychophysics. Note that computational creativity is fundamentally not a supervised learning problem: One must decompose artifacts into parts and have assessment methods for the parts and for the recombination rules to predict quality of completely new artifacts. There will not be training data available for novel complete artifacts.

Novelty can be assessed via information-theoretic or other similar quantifications of innovation within the context of all other existing artifacts in the domain of interest. The novelty dimension is less specific to the particular creative domain of interest, whereas the quality dimension is intimately tied to it. We provide details for a specific domain, the flavor of food, in the following sections.

The final component, the work planner, determines steps needed to take the concept to externalization. The work plan provides constraints on what designs are possible and can be optimized for efficient production, e.g., using techniques from planning and operations research. Generating the plan is itself a creative act and may be judged as such if an audience observes production. However, artifacts can be deemed creative even if the work plan used to produce the artifact is not observed.

3.1 Computational creativity in the culinary domain

In the food domain, a dish is the basic unit of creation. A recipe is a work plan for how to cook a dish, but it is also a description of the work product, as it describes the ingredients to be used, their quantities, and their transformations and combinations. A menu is a set of dishes that together constitute a meal, a kind of narrative sequence [45] to be created.

The overall culinary recipe design problem has many facets. Through the lens of Figure 2, the first is to design and construct a suitable *domain knowledge database*. This requires a data model enabling the system to reason about food and support algorithms for design, assessment, and planning. In particular, it should be a repository of food ingredients and existing recipes, but also include knowledge about culinary styles and techniques, regional and seasonal cuisines, flavor compounds and their combinations, etc. We propose and discuss a data model for food in Section 4.1.

A related aspect is ingesting and processing raw data to populate the domain knowledge database structured according to the data model with the *knowledge categorizer*. Sources include cookbooks and other repositories of recipes, culinary guides that explicate the culture of food, repositories of culinary techniques, and chemical databases of food ingredient constituents.

Given a designed and populated domain knowledge database, a next step is developing a way for the *work product designer* to generate recipe ideas. Since cuisine naturally has evolutionary properties [46], genetic algorithms are one approach to recipe generation [47]. Our approaches include using stochastic sampling [48] and associative generation [49].

Besides random recombination of recipes, there are some prominent culinary design principles that can be utilized. For example, two principles focused on the chemosenses are the flavor pairing hypothesis [50] and olfactory pleasantness maximization [51]. Additional principles

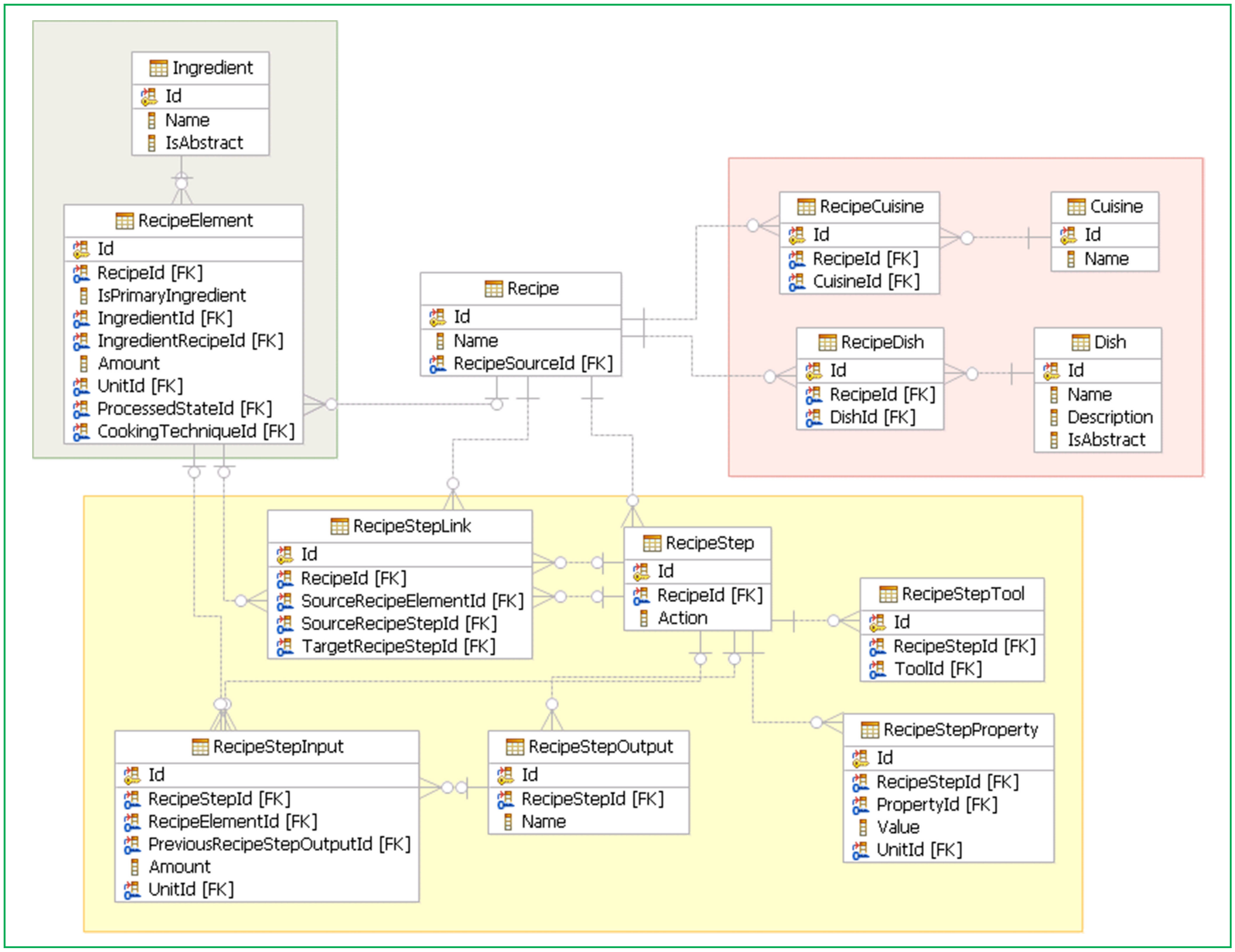


Figure 3

Knowledge representation schema for culinary recipes. The ingredient component is expanded upon in Figure 4.

center around similarity of ingredients in properties, such as geographic origin and seasonal origin. Chefs may also want to maintain balance (in terms of tastes, temperatures, or textures), or on the contrary accentuate a given characteristic, e.g., create the beefiest burger or the crunchiest cookie.

Finally, a recipe is not only a work product but also a rudimentary work plan. Therefore, in the culinary domain, a plan to produce the dish (or menu) is a must. The *work planner*, utilizing a machine system's strengths, can optimize and parallelize this plan by formulating an operations research problem [52].

As detailed in Section 6, the temporal ordering of how the various components of the system are used roughly follows the stages of human creativity (see Figure 1) to facilitate interaction with people. In operation, the three analytics components—work product designer, work product assessor, and work planner—are used cyclically in that order by people to converge on a final design.

4 Data engineering

4.1 Artifact data model

Here, we propose a data model to capture the salient pieces of domain knowledge in supporting all of the components of machine-generated creative recipe design.

As in Section 3.1, the basic unit of cuisine is the dish, which is represented as a recipe. We propose a representational model for culinary computational creativity that also has a recipe as the basic unit, yielding a schema for cuisine shown in **Figures 3** and **4**. Within this representation, we first capture the basic factors of the recipe, including the ingredients and their quantities, the tools required, and the sequence of cooking steps with

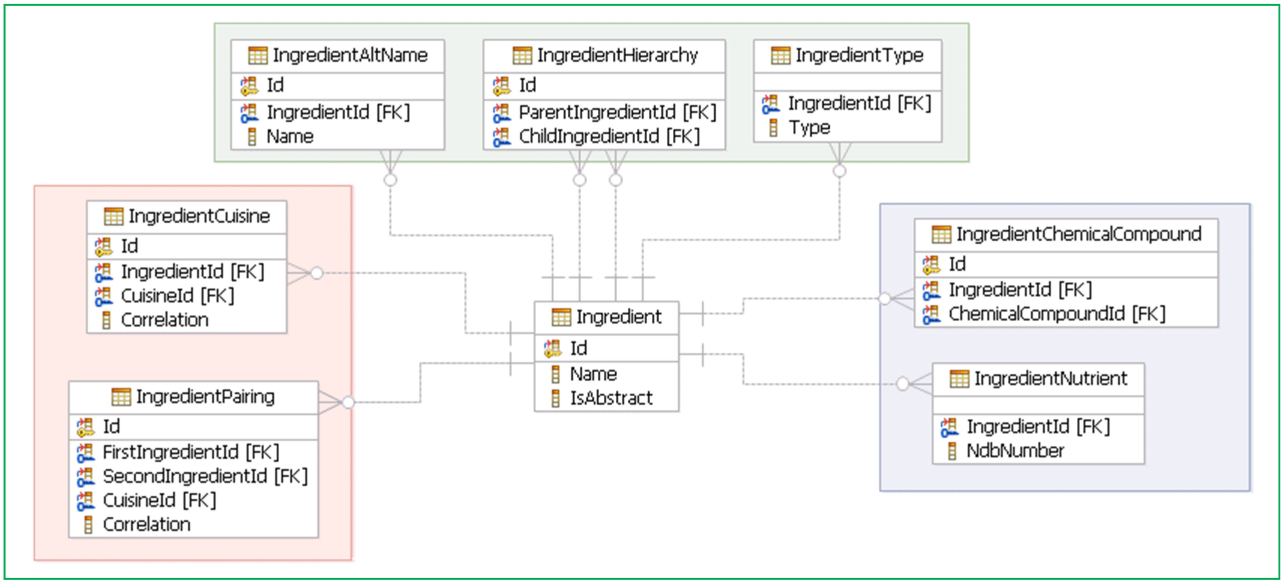


Figure 4

Knowledge representation schema for culinary ingredients that is a part of the overall schema for recipes given in Figure 3.

input, output, tool, and duration specified. These basic factors are enough to be able to produce the artifact, i.e., the dish.

A creative culinary system's knowledge representation, however, needs much more than simply a recounting of the ingredient list and cooking steps. We must include knowledge about cultural context, human ratings, chemical analysis of ingredients and processes, etc., to be able to characterize and emulate flavor perception. For example, we include the name of the dish because it relates to the influence of cortical language circuits on flavor perception. We include the regional cuisine to which the dish belongs because regionality is a prominent design principle in cooking. Similarly, we include the chemical flavor compound constituents of ingredients to enable modeling of flavor perception.

Idea generation can only use attributes in the data model and nothing more. It truly is the case that the way the world is internally represented determines what can be created. Creation, in our view, is the process of decomposing artifacts into their constituents as depicted in the data model, and then recomposing and reconstituting new artifact ideas. Without a simplified universe containing blindspots, the deployment of reasoning resources becomes untenable. In the culinary case, we certainly have blindspots in our proposed schema. For example, we do not include a data element about sound even though it is a contributor to human flavor perception, e.g., sizzling sounds often enhance flavor perception. Since sound is not in the schema, it is also outside the universe of reasoning for the system. Importantly, it is purposefully not in the schema because capturing every component of flavor would be unmanageable.

4.2 Natural language processing

Recipes originally written in a human-readable form must be parsed to extract key knowledge for the data model [53]. Beyond just ingredient lists as in other work [29, 50], our system needs to understand the inputs, outputs, tools, times, and techniques of the recipe steps.

This is performed using natural language processing; our approach for processing ingredient amounts, units of measure, names, and processed states is rule-based. We enumerated possible constructions of statements such as "1 tablespoon cinnamon" as amount is 1, unit of measure is tablespoon, and ingredient name is cinnamon; or "1/2 cup (120 g) chopped walnuts" as amount is 1/2, unit of measure is cup, ignored statement is "(120 g)," processed state is chopped, and ingredient name is walnut. This was used to construct tens of rules for parsing, which completely covered approximately 95% of ingredients in recipe datasets produced through peer production (as in the 25,000 recipes available on Wikia); the remainder were processed, in part, manually.

Our approach for processing the recipe instructions is based on statistical parsing with domain-specific tokens, including ingredient names, tools, and techniques. Crowdsourcing has been used to develop an initial labeled corpus that can be bootstrapped for improved statistical parsing. A key element of the algorithm is part-of-speech tagging using an off-the-shelf application of the Stanford parser.[1] As compared to training on general corpora (e.g., *Wall Street Journal*), naive statistical parsers trained on both general and domain-specific corpora can have accuracy that improves from 65% to 85%, in terms of getting the task, tool, ingredient, and tip correct from a recipe instruction sentence.

Peer-produced recipes come in various styles and are not as structured as recipes in published cookbooks, presenting extra challenges. Some notable attributes include personal commentary, multilingual text, missing information, abstracted description, and implied temporal information, which are elaborated as follows:

1) *Personal commentary*—Since cooking is often personal, people share their emotional attachment to recipes (e.g., child, travel memories) and indicate special occasions to prepare the dish (e.g., romantic dinner, Christmas treat, family dish).
2) *Multilingual text*—Although nominally an English language resource, a substantial number of Wikia recipes are written in more than one language, often using languages such as Spanish, Italian, or French to emphasize a particular ingredient or step.
3) *Missing information*—Text often has missing ingredients, measurements, or steps: Oft forgotten ingredients are sugar, salt, and pepper.
4) *Abstracted description*—A listed ingredient may be described by a more general term in the instructions, e.g., "wash the vegetables."
5) *Implied temporal information*—Step duration is more often stated in terms of stopping conditions, e.g., "cook until tender," in Wikia than in other sources.

We use domain-specific tokens to filter personal commentary, and dictionaries deal with multilingual text. Using elements of commonsense reasoning [54], missing information is inferred (e.g., salt can be added back to ingredient list when parsing the instructions) and abstracted descriptions are resolved to more specific descriptions using an ingredient ontology. Computational creativity algorithms must be robust to these properties of the inspiration set.

4.3 Related information

Besides ingesting repositories of extant recipes, it is also important to gather related information. One source is Wikipedia, as a description of regional cuisines. Again, natural language processing is needed to convert text into insight, e.g., which ingredients are typical or canonical for a given region. There are hundreds of regional cuisines to be understood. As an example, using the domain-specific token dictionary described above, we can extract regional ingredient lists from Wikipedia pages on given regional cuisines.

Another source of data, especially important for computational creativity, is hedonic psychophysics data [55] linked to chemical informatics data. This provides characterizations of which flavor compounds are present in which ingredients, and how much people like those flavors according to human psychophysics experiments. Each ingredient may contain hundreds of flavor compounds in varying concentrations, as determined in the Volatile Compounds in Food 14.1 database (VCF) and in Fenaroli's Handbook of Flavor Ingredients as processed and released in [50].

Since experimental psychophysics data [55] may be sparse with respect to the thousands of flavor compounds present in foods, data is also needed to predict the hedonic percepts of unmeasured compounds. This requires further physicochemical data on the various compounds; there can be hundreds or thousands of physicochemical descriptors such as the number of atoms or the molecular complexity.[2]

Much of the data from these sources is already well-structured in databases. When mapping certain named entities (such as names of ingredients and names of flavor compounds) across databases, however, we resorted to a manual approach when exact automatic matching failed.

5 Data-driven assessment

We now turn to data-driven assessment of novelty and flavor, which draw from human flavor perception science and operate within the universe set forth by the data model and related data. We begin with a computational proposal for novelty, which can also be applied more generally to other creative endeavors. We then develop a computational quantification of pleasantness for food. A creative recipe should have large values for novelty and pleasantness quantifications.

5.1 Novelty

An artifact that is novel is surprising and changes the observer's world view. Novelty can be quantified by considering a prior probability distribution of existing artifacts and the change in that probability distribution after the new artifact is observed, i.e., the posterior probability distribution. At the level of observable representation of artifacts, the difference between these probability distributions describes exactly how much the observer's world view has changed. In recent work, such a quantitation has been given the name *Bayesian surprise* and has been shown empirically to capture human notions of novelty and

[1] http://nlp.stanford.edu/software/lex-parser.shtml

[2] See, e.g., https://pubchem.ncbi.nlm.nih.gov/.

saliency across different modalities and levels of abstraction [26–28].

Surprise and novelty depend heavily on the observer's existing world view, and thus the same artifact may be novel to one observer and not novel to another observer. That is why Bayesian surprise is measured as a change in the observer's specific prior probability distribution of known artifacts.

Bayesian surprise is defined as follows. Let $\mathcal{M}$ be the set of artifacts known to the observer, with each artifact in this repository being $M \in \mathcal{M}$, which should be thought of as a chance variable. Furthermore, a new artifact that is observed is denoted as A, which is also thought of as a chance variable. The probability of an existing artifact is denoted as $p(M)$, the conditional probability of the new artifact given the existing artifacts is $p(A|M)$, and via Bayes' theorem the conditional probability of the existing artifacts given the new artifact is $p(M|A)$. The Bayesian surprise is defined as the following Kullback–Leibler divergence (relative entropy) between the two probability distributions:

$$\begin{aligned}\text{Bayesian surprise} &= D(p(M|A)||p(M)) \\ &= \int_{\mathcal{M}} p(M|A) \log \frac{p(M|A)}{p(M)} dM.\end{aligned} \tag{1}$$

Thinking of an artifact as an unordered tuple of N ingredients $A = \{I_1, \ldots, I_N\}$, combinatorial expressions for probability distributions are found. This is done by looking at occurrence probabilities of single ingredients, pairs of ingredients, triples of ingredients, etc. Although there are sophisticated techniques for estimating information-theoretic functionals from data [56], we find plug-in estimators to often be sufficient. So to estimate $D(p(M|A)||p(M))$, we use the empirical distribution of the existing artifacts $\mathcal{M}$ as $p(M)$, and for $p(M|A)$, we use the empirical distribution of $\mathcal{M} \cup A$. The *unseen elements* problem in statistical estimation [57] is critical in creativity since the goal is to create completely novel artifacts; to handle it, we introduce a small probability mass for unseen objects.

5.2 Flavor pleasantness

Human flavor perception is very complicated, involving a variety of external sensory stimuli and internal states [58]. Not only does it involve the five classical senses, but also sensing through the gut, and the emotional, memory-related, motivational, and linguistic aspects of food. First, there are the basic tastes: sweet, sour, salty, bitter, and umami. The smell (both orthonasal and retronasal olfaction) of foods is the key contributor to flavor perception, which is in turn a property of the chemical compounds in the ingredients [59]. There are typically tens to hundreds of different flavor compounds per food ingredient [50] and olfactory perception is integrative rather than analytic, yielding unified percepts [60].

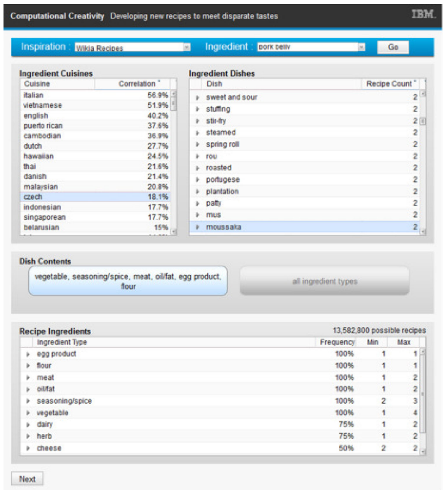
Other contributors to flavor perception are the temperature, texture, astringency, and creaminess of the food; the color and shape of food; and the sound that the food makes. The digestive system detects autonomic and metabolic properties of food. Moreover, there are emotion, motivation, and craving circuits in the brain that influence flavor perception, which are in turn related to language, feeding, conscious flavor perception, and memory circuits. Furthermore, stimuli beyond the food itself, such as ambience of the room, influence flavor.

The complication in flavor perception is due to the interconnection and interplay between a multitude of neural systems, many of them not memoryless. Recreating such a flavor perception system in a computer is an ambitious goal, but any progress is an advance toward a viable computational creativity system for food. Also, note that simply describing the factors and pathways of flavor perception fails to consider the settings of those factors that make food flavorful.

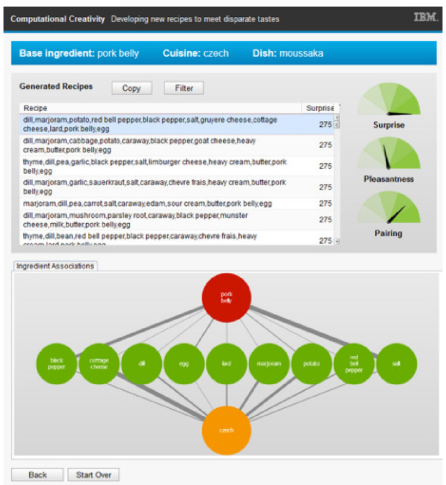
To evaluate the flavor pleasantness of generated recipes, we focus on the fact that constituent flavor compounds sensed by the olfactory system are the key to flavor perception. Thus, a tractable step toward a data-driven model for flavor pleasantness is a model for odor pleasantness.

Recent work has shown that there is a low-dimensional, almost scalar, hedonic quantity that describes the pleasantness of odors to humans, regardless of culture or other subjectivity [51, 55]. Moreover, this pleasantness is statistically associated with the physicochemical properties of compounds [61]. Hence, we develop regression models to predict human-rated odor pleasantness of chemical compounds using their properties such as topological polar surface area, heavy atom count, complexity, rotatable bond count, and hydrogen bond acceptor count. Starting with tens of physicochemical features for 70 observations in a pleasantness-labeled training dataset [51], multiple linear regression with model selection based on smallest prediction error in either tenfold or leave-one-out cross validation yielded the small set of features used in the final regression model. The regression model achieved $R^2 = 0.374$.

There is evidence that pleasantness is an approximately linear property of compounds [62]. If two compounds are mixed together and smelled, the hypothesis is that the odor pleasantness of the mixture is approximately a linear combination of the pleasantness values of the individual compounds. With such linearity, one can predict the pleasantness of food ingredients that contain several flavor compounds and of dishes that in turn contain several ingredients. Some of the individual ingredients that our



(a)



(b)

(c)

Figure 5

(a) Interface for problem finding, showing common and uncommon choices. (b) Interface for selection, showing design reasoning and ratings along dimensions of novelty and quality. (c) Bottom of interface for externalization, showing recipe steps and their partial ordering. Green boxes are recipe ingredients, and orange boxes represent actions performed. For the example quiche recipe depicted: Vegetables are cut and fried together, wet ingredients are mixed, pie crust dough is rolled, etc. Steps can be performed by multiple cooks in parallel, until all elements are put together, baked, and cooled.

model predicts to be most pleasant include black tea, bantu beer, beer, strawberry, white wine, and cooked apple.

The chemical properties of flavor compounds are well-catalogued, and there is a growing body of literature cataloguing the flavor compound constituents of food ingredients [50]. In fact, flavor compound data can also be used to compute the level of flavor pairing among the ingredients; this measure has been indicated to give a regional cuisine-specific sense of overall human flavor judgement [50].

5.3 Recipe assessment

Thus, if the recipe assessor is given a proposed idea by the recipe designer in a computational creativity system, it can calculate its novelty using Bayesian surprise and calculate its flavorfulness using an olfactory pleasantness regression model applied to its constituent ingredients and flavor compounds in those ingredients. Ideas can also be assessed according to flavor pairing [50]. Such scoring represents a data-driven approach to assessing artifacts that have been newly created and have never existed before.

Users can either be shown the several performance metrics in a disaggregated form, e.g., via gauges in **Figure 5(b)**, or some aggregation function can be developed to provide a single score. One approach that users found intuitive was to have a slider that allowed the input of weights for a linear combination.

6 Computer–human interaction for semiautonomy

Although the computational creativity system defined thus far can operate autonomously, it can have a greater impact as part of an integrated collaborative workflow with human creators. We implement an interactive interface, taking a mixed-initiative approach to human–computer interaction via turns between human and computer [32].

The first step in creativity is problem finding. Mediated by a novel interactive interface design, this may be accomplished jointly by the human and the machine, by picking a key ingredient, one or more regional cuisines to influence flavor, and a dish type such as soup or quiche. Machine learning is used to suggest ingredient types, though this can be modified by the human. The problem-finding input screen, **Figure 5(a)**, sets parameters for the generative algorithm to create thousands or millions of ingredient list ideas.

The penultimate stage in creativity is selecting the best idea(s). The computer predicts which generated recipes will be the most surprising to a human observer, will be perceived as the most flavorful, and will have the best pairings of ingredients (see [50]). These metrics are used to rank the generated ideas, and then a human makes the final selection [see the selection screen in Figure 5(b)]. In our experience, humans select one of the top ten ideas, rather than looking through hundreds or thousands of possibilities. Hence, selection is truly a collaboration between human and machine.

Visualizations at the bottom of the screen help the human understand the reasoning used by the computer in generating and ranking ideas, so as to provide confidence. This includes visualizing the design process, as well as the metrics of pleasantness, pairing, and surprise.

The final stage of creativity is externalizing. In recipe creation, this involves coming up with not just the list of ingredients (the focus of idea generation and selection), but also proportions and recipe steps. Professional chefs often operate without computer support for externalizing, but amateurs appreciate guidance since it too requires significant creativity and much domain knowledge. The final screen shows proportions and steps in the form of a directed acyclic graph; see **Figure 5(c)**. Possible actions are abstracted to improve reasoning. Details of an algorithm for creating proportions based on distributional matching of ingredient types and nutrients, as well as an algorithm for creating recipe steps based on subgraph remixing, are detailed elsewhere [48]. The directed acyclic graph may also be converted into a natural language text.

7 Menus of recipes

So far, we have discussed creating a single recipe at a time, and in Section 6, problem-finding was cast as human–machine interaction for picking a key ingredient, regional cuisines, and dish types. When creating a sequence of dishes, such parameters should be linked across dishes. Here, we introduce the notion of dish *variety* based on topic modeling. Topic models are used to identify underlying latent topics in a set of documents; we apply them to a repository of recipes.

7.1 Topic modeling

Definition 1 requires creative artifacts to be novel and of high quality. For menus, the novelty and quality of the set is partially determined by its constituent dish recipes, but variety is a property of multiple artifacts: It is an emergent property for collections and is not definable for individual artifacts.

Topic models are machine learning algorithms that discover the main underlying themes that pervade a large collection of documents through a generative model assuming documents are probabilistic mixtures of a set of underlying latent variables, i.e., "topics," and the "words" that compose a document are probabilistically generated from these topics [63]. Here, we treat recipes as documents, and apply the latent Dirichlet allocation (LDA) method of topic modeling [64] to the Wikia corpus of recipes. While applying this method, we assume that a recipe is adequately summarized as a set of ingredients, but see Section 4.1.

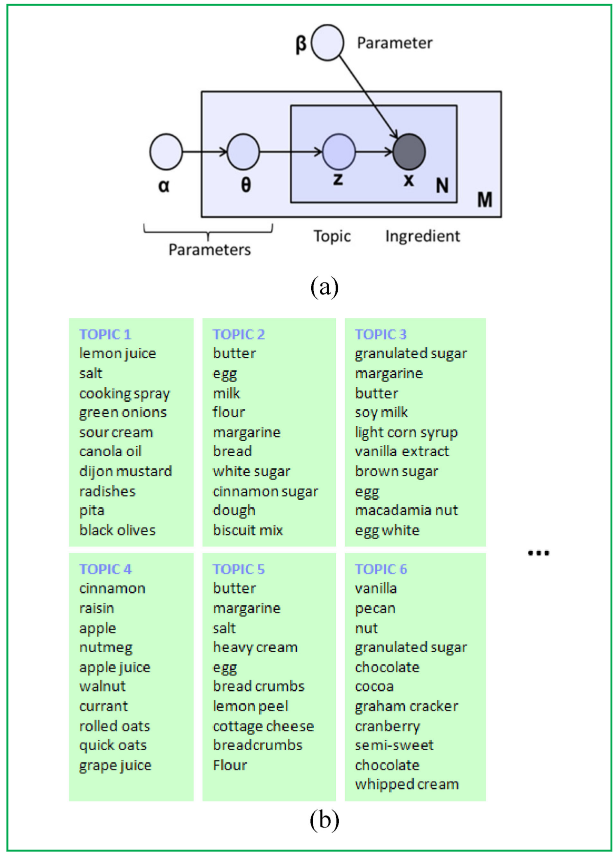


Figure 6

(a) Plate diagram for the LDA topic modeling method. (b) Example topics when the LDA method is applied to Wikia recipes.

Topic modeling has previously been applied to recipes for other purposes [65].

Figure 6(a) presents a plate model depiction of the underlying generative LDA model as applied to a corpus of M recipes. The model assumes that ingredients in recipes in the corpus are generated by first selecting a topic z based on the topic distribution θ, which is a function of hyperparameter α. Then, an ingredient x is generated based on the ingredient distribution over topics, denoted by hyperparameter β. The outer plate in the figure represents recipes, whereas the inner plate represents the repeated generation of recipe topics and ingredients within a recipe. Note that a topic distribution and N ingredients are generated for every recipe. The model hyperparameters α and β are estimated using data.

Learned recipe topics can be interpreted as some ingredient combinations that either go well together (such as sauces) or that can be substituted (such as a set of fruits). To generate a new recipe, one selects an ingredient by first choosing an underlying recipe topic, and then drawing the ingredient from the recipe topic-specific distribution.

Figure 6(b) shows a list of a selected few topics that were generated using the LDA method applied to the Wikia corpus, along with some of the most likely ingredients in these topics.

7.2 Assessing variety

Here, we propose a topic-based approach for assessing variety in menus. Consider a menu with K recipes, where the $k\mathrm{th}$ recipe is $A_k = \{I_{k_1}, \ldots, I_{k_N}\}$ and I_{k_n} is the $n\mathrm{th}$ ingredient in the $k\mathrm{th}$ recipe. For notational convenience, assume all recipes in the menu have the same number of ingredients N, but the method applies to the general case where recipes have varying numbers of ingredients. Suppose that a topic model with L underlying recipe topics has been used to model the generative process by which the corpus of recipes was created. Let T denote the random variable for the marginal distribution of the recipe topics from which an ingredient is selected for a recipe, and let t_ℓ be the $\ell\mathrm{th}$ topic.

Note that a topic model is a Bayesian model that considers the relationship between the parameters, the recipe topics, and the ingredients that are chosen in recipes. Therefore, we can use the Bayes' rule and perform inference to compute the probability that a particular recipe topic was selected in picking a particular ingredient. Let $P(T|I_{k_n})$ denote the probability distribution over recipe topics for the $n\mathrm{th}$ ingredient in the $k\mathrm{th}$ recipe. This probability measures how an ingredient is associated with the underlying themes in the recipe database.

To measure the variety in menus, we compare how various recipe topics are spanned by the constituent recipes through the notion of a topic spanning metric for a recipe, which measures how that particular recipe is associated with the various recipe topics. This *topic spanning metric* for a recipe can be any function of topic probabilities for its constituent ingredients:

$$s_k = s(A_k) = f\big[P\big(T|I_{k_1}\big), \ldots, P\big(T|I_{k_N}\big)\big]. \tag{2}$$

An example is when a topic is said to be covered by a recipe when at least one ingredient in that recipe was selected from that topic. Let s_{k_ℓ} be the probability that the $\ell\mathrm{th}$ topic was used by at least 1 ingredient in the $k\mathrm{th}$recipe, in which case

$$s_{k_\ell} = 1 - \prod_{n=1}^{N} [1 - P(T = \ell | I_{k_n})]. \tag{3}$$

The following vector is a potential topic spanning metric; it measures the degree to which every recipe topic is associated with the kth recipe: $s_k = \{s_{k_\ell}\}_{\ell=1}^{L}$. We can now score menu variety based on the distance between spanning metrics for recipes in the menu. Note that all recipes have vectors of the same dimension, which is the number of recipe topics

$$\mathrm{Variety} = D[s_1, \ldots, s_K] \tag{4}$$

where $D[\cdot]$ is a distance metric such as Euclidean distance. The variety score computed in this fashion assesses how recipes in the menu differ from each other in terms of the fundamental underlying themes from which they were generated. Thus, the topic modeling approach allows for the effective use of data to identify and model variety in menus and leads to pleasing sequences of individual recipes.

8 Experimental validation

The computational creativity system described herein has been used to create hundreds of novel and flavorful recipes that have been served to thousands of people. Recipes range from Indian Turmeric Paella, Baltic Apple Pie, and Ecuadoran Strawberry Dessert to Creole Shrimp Dumpling, Swiss-Thai Asparagus Quiche, and Turkish Bruschetta. Recipes for some of these and many other dishes created by the system are available elsewhere [66]. As an example, a Caymanian Plantain Dessert recipe generated and selected by the system[3] is given in Appendix A, together with a photograph.

8.1 Consensual assessment technique

Although recipes created by the computational creativity system have been tested by professional chefs at the Institute of Culinary Education (ICE), we performed the CAT using professional chefs that were not involved in the project development. Each of three chefs was asked to evaluate the Caymanian Plantain Dessert (see Appendix A) generated by the system, along with two other similar recipes. One of these recipes, Plantain Tart (see Appendix B), is from a peer-produced repository of recipes that served as an inspiration set for the computational creativity system, whereas the other recipe, Dulce de Platanos (see Appendix C), is from a professional online repository of recipes.

For each of these recipes, panelists were asked to compare the recipe to all recipes they had seen as professional chefs, providing ratings on 1-to-5 hedonic scales for the following dimensions: creativity, novel combinations of ingredients, flavor, and well-paired ingredients. The results are plotted in **Figure 7(a)**. Notice that the Caymanian Plantain Dessert scored a 4:33, which is near the top of the hedonic scale and much higher than the two human-created desserts. Results indicate that the Caymanian Plantain Dessert is creative and therefore demonstrate the system can achieve creativity. Even on the contributing subdimensions of novelty, flavor, and pairing, the computationally created recipe scores better than or equal to the human-created recipes.

To give some more insight into the process by which the Caymanian Plantain Dessert recipe was created, **Figure 7(b)** shows a rank-frequency plot for the data-driven psychophysical pleasantness scores for 10,000 Caymanian

[3] The ingredient list was created fully autonomously, and the recipe steps and ingredient proportions were determined through the mixed-initiative approach by an author of this paper.

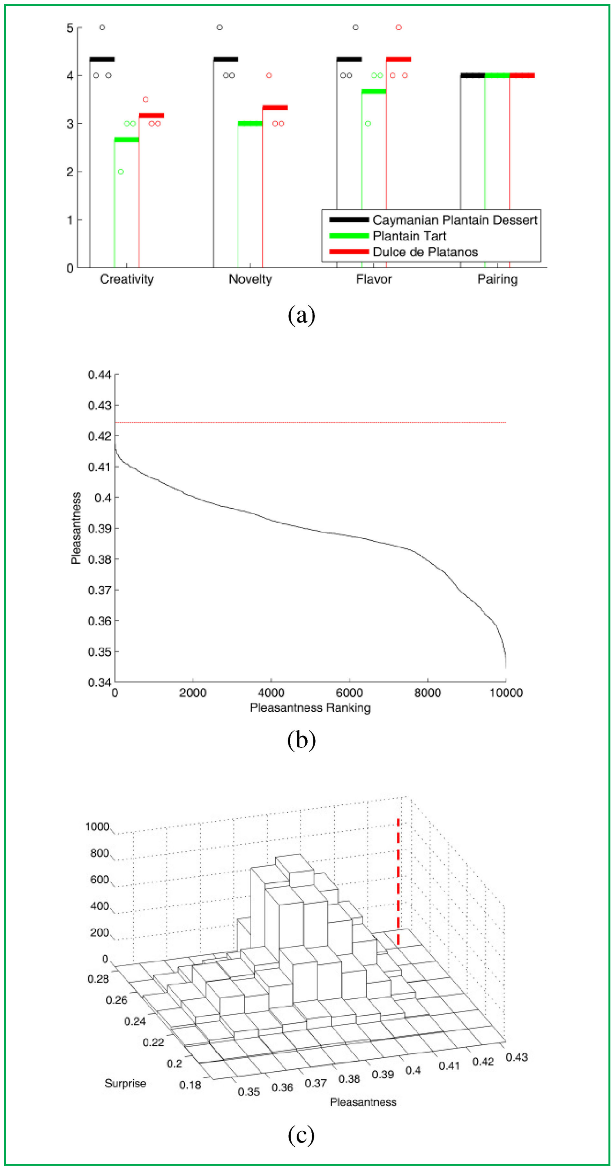


Figure 7

(a) Assessment of creativity using the CAT, which uses the average scores of the panelists. The first group of bars is for creativity for the three recipes assessed, and the remaining groups of bars are for novel combination of ingredients, flavor, and well-paired ingredients. Circles indicate the individual ratings by the panelists. (b) Rank-frequency plot of pleasantness scores for 10,000 Caymanian Plantain Dessert recipes generated with a creativity system (solid line) and pleasantness score for the selected and tested recipe (dashed line). (c) Joint histogram of surprise and pleasantness for 10,000 Caymanian Plantain Dessert recipes generated with a creativity system. Values for the selected and tested recipe are indicated with dashed line.

Plantain Dessert recipes from the system's culture-based generative algorithm (solid line). Due to the stochastic sampling structure of the algorithm, these 10,000 recipes are representative of any larger set that would be generated. Figure 5(b) also shows the pleasantness value for the recipe selected for maximal pleasantness under a minimal surprise constraint (dashed line). **Figure 7(c)** shows a joint histogram of surprise (in units of wows arising from base-2 logarithms) and pleasantness for these 10,000 Caymanian Plantain Dessert recipes; the dashed line indicates the values for the final recipe selected. As both plots demonstrate, the generative algorithm produces many possibilities, but the selective algorithm draws out the best one for human selection and verification through flavor testing.

Testing by experts validates that the computational creativity system is indeed creative under Definition 1 via the CAT.[4]

8.2 Impact test

To further corroborate the results of the CAT, we also tested for creativity via impact in large-scale deployment. It is an alternative operational test based on Definition 1.

At the 2014 South by Southwest festival, IBM and ICE introduced a food truck powered by the system; see **Figure 8(a)**. Every day, the truck let the audience vote for a type of dish on Twitter, created a never-before-seen recipe for the winning dish using the system, and served the results to visitors the next day. On average, 500 portions were given each day with overwhelmingly positive feedback. Moreover, professional chefs at various hotels, restaurants, food companies, and culinary schools have indicated that the system helps them explore new vistas in food, and have expressed a desire to have access to such a tool in their daily jobs. This result also indicates that the system was creative, providing validation for the data-driven approach to computational creativity.

Additionally, at the food truck, novice users were allowed to design recipes themselves and then provide a binary rating of whether the resulting recipe looked to be "yummy" or "yucky" just by reading the recipe. Of 236 creation instances, 77.1% were rated "yummy." Delving into the pleasantness (from hedonic psychophysics) and the chemical pairing (from the flavor pairing hypothesis) scores for yummy and yucky recipes, however, we see novices may be unable to make reasonable judgments of perceived flavor just from the text, when recipe ideas were far from their regular experience. **Figures 8(b)** and **(c)** show the empirical cumulative distribution functions of pleasantness and chemical pairing scores for recipes that were rated "yummy" and "yucky." As determined by the two-sample Kolmogorov–Smirnov goodness-of-fit hypothesis test, for each pleasantness and pairing, we do not reject the null hypothesis at the 5% significance level that the two distributions are the same. This result points to the fact that

[4] Cf. [67] for an independent CAT for the same recipe that also led to a determination of creativity.

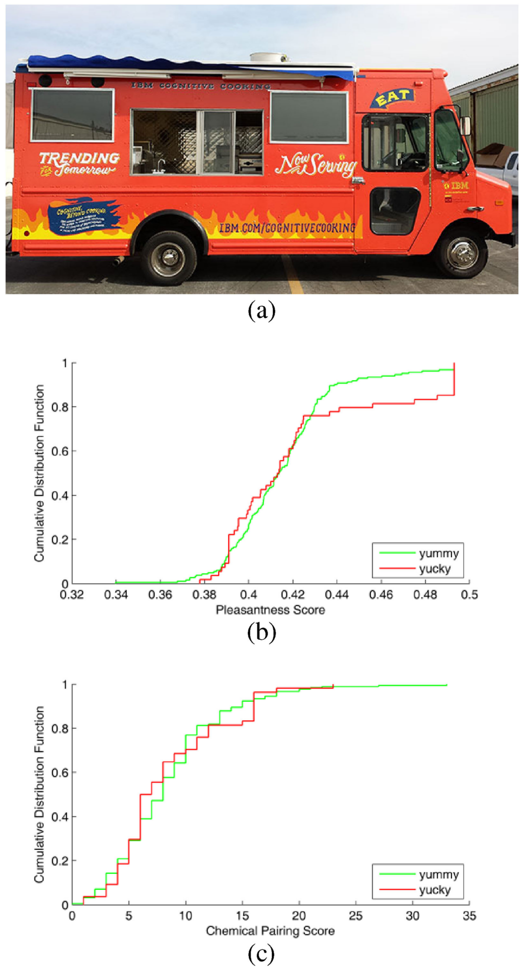


Figure 8

(a) Food truck that implements the computational creativity system, which was used to serve over 2,000 dishes created and to perform the impact test. (b) Empirical cumulative distribution functions of pleasantness scores for recipes rated "yummy" and "yucky" by novices. (c) Empirical cumulative distribution functions of chemical pairing scores for recipes rated "yummy" and "yucky" by novices.

the computational creativity system may have more ability in predicting flavor perception from written data than novice humans.

9 Summary and outlook

In this paper, we have described a computational creativity system that can automatically or semiautomatically design and discover culinary recipes that are novel and flavorful. This is done through artificial intelligence algorithms based on Bayesian probability and regression analysis, as well as disparate data sources from culinary traditions, chemoinformatics, and hedonic psychophysics. We proposed a structure for a computational creativity system that contains three main components: a designer, an assessor, and a planner, all fed by a domain knowledge database. Furthermore, we discussed the role of the domain knowledge database in structuring and setting the bounds for creativity. Our experimental results demonstrate the system and validate the efficacy of the system design: The system can indeed produce artifacts judged as creative.

Creativity is easy neither for people nor for machines, but the challenges are different. Without taking advantage of modularity, people often have trouble being creative and innovative because they are overwhelmed by the combinatorial complexity of large design spaces [68]. Since people end up thinking modularly, progression of creative thought is often evolutionary [69]. A computational creativity system can test quadrillions of ideas at once without needing to invoke modularity and may thus offer solutions that completely redefine an art. Such creations may offer advantages by being completely "outside the box" through large jumps in thought rather than gradual evolutionary changes.

Although we chose a particular creative application domain—culinary recipe design—as an example, the system architectures, approaches, and insights garnered in facing the challenges should be applicable across creative domains.

Appendix A

Caymanian Plantain Dessert Recipe

Caramelized bananas
21 g butter
28 g molasses
1 tsp (5 g) pure vanilla extract
about 0.3 g nutmeg
170 g peeled very ripe bananas, medium dice
85 g milk

- Heat butter and molasses in a saucepan over medium heat.
- Add vanilla extract and nutmeg, then the bananas, and cook for 2 minutes, stirring regularly with a spatula.
- Add the milk, stir, and bring to a simmer. Remove from heat. Adjust nutmeg as needed: You should be able to taste just a hint of it.
- Pass the mixture through a sieve. Process half of the bananas with the liquid in a blender until smooth.

Transfer to a container, mix in rest of the banana chunks, and let cool for 30 minutes.

- Pour into verrines and refrigerate for at least 30 minutes. Once cold, mixture should not be liquid anymore.

Coconut and lime pastry cream
3 egg yolks
45 g light brown sugar
14 g flour
170 g milk
17 g lime juice
28 g coconut flakes
3 g butter, diced

- In bowl, mix egg yolks and half of the sugar with whisk for 1–2 minutes, then mix in sifted flour.
- In a small saucepan over high heat, place milk, lime juice, coconut flakes, and the rest of the sugar, and bring to a simmer. Remove from heat and let steep 5 minutes.
- Process milk mixture in a blender, and pass through a conical sieve, pressing with a ladle to get all liquid out of coconut residue. Return liquid to saucepan, and bring back to a simmer.
- Slowly pour milk over egg yolk mixture to temper it, whisking constantly. Return to the saucepan, and bubble gently for 2 minutes, still whisking constantly. Transfer to a container, mix in butter, and let cool for 15 minutes.
- Pour into verrines and refrigerate for at least 30 minutes. Once cold, mixture should not be liquid anymore.

Papaya and orange salad
113 g orange juice
20 g butter, diced
about 0.1 g cayenne pepper
128 g papaya, small dice (1/4")

- In a saucepan over high heat, reduce the orange juice to $\frac{1}{4}$.
- Whisk in the butter and cayenne pepper. Adjust the pepper as needed: You should be able to taste just a hint of it.
- Toss the papaya, and remove from heat. Transfer to container, and let cool for 15 minutes.
- Pour into verrines and refrigerate for at least 30 minutes.

Plantain chips
Corn oil, for deep-frying
1 plantain, very cold
salt

- In a deep-fryer, heat the oil in to 375 F / 190 C.
- Peel the plantain, then cut it in half. Using a mandoline, slice each half very thinly: the slices should be just thick enough so they do not break. Cut each slice in half lengthwise (into two long strips).
- Proceeding in small batches, deep-fry the plantain strips until golden brown, then drain on paper towels and let cool.
- Deep-fry the strips a second time for about 10 seconds to make them crispy. Drain on a paper towel, and season with salt.
- Place two chips on each verrine just before serving.

Appendix B

Plantain Tart Recipe

Available at http://recipes.wikia.com/wiki/Plantain_Tart
2 cups all-purpose flour
1 teaspoon salt
1/4 cup cold butter, cut into 1/2 inch pieces
3 tablespoons shortening, chilled and diced
1 egg, beaten
1 tablespoon ice-cold water
3 very ripe (black) plantains
1/4 cup white sugar
1 teaspoon vanilla extract
1 teaspoon grated nutmeg
2 drops red food coloring (optional)
1 egg white, beaten
white sugar for decoration

- Prepare the pastry by combining the flour and salt in a bowl. Rub in the butter and shortening until incorporated and the mixture takes on a sandy appearance. Combine the egg and water, and stir into the flour mixture until a dough forms, then knead for a few turns to bring the dough together. Wrap well, and chill for 3 hours in the refrigerator.
- While dough is chilling, peel plantains and cut into thirds. Place into a small saucepan with a little water. Bring to a simmer and steam until tender, 5–10 minutes depending on how ripe your plantains are. Once soft, pour out the water, and mash plantains with sugar, vanilla, nutmeg, and red food coloring. Set aside to cool.
- Preheat oven to 350 °F (175 °C).
- Roll dough out on a lightly floured surface to 1/4 inch thick. Cut into circles using a 4- or 5-inch round cookie cutter. Spoon a little of the plantain filling into the center of each circle, then fold in half, to form a half-moon shape. Place the tarts on a baking sheet, brush with beaten egg white, and sprinkle with sugar.
- Bake in preheated oven for 20–25 minutes until golden brown. Allow tarts to cool to room temperature before serving.

Appendix C

Dulce de Platanos Recipe

Available at http://www.epicurious.com/recipes/food/views/Dulce-de-Platanos-15745

2 very ripe (brown to black) plantains
1/2 stick (1/4 cup) unsalted butter
2 tablespoons dark rum
1/2 cup well-shaken canned unsweetened coconut milk
1 cup sugar
1/4 cup heavy cream
Accompaniment: vanilla ice cream

- Cut ends from plantains and peel fruit. Diagonally cut plantains into 1/2-inch-thick slices. In a 12-inch nonstick skillet, heat butter over moderate heat until foam subsides and cook plantains until golden, about 3 minutes on each side. With a slotted spatula, transfer plantains to a plate, reserving butter in skillet.
- In a very small saucepan, heat rum and coconut milk until warm. Add sugar to reserved butter and cook over moderate heat, stirring, until caramelized, about 5 minutes. Remove skillet from heat and carefully whisk in warm coconut milk mixture (mixture will vigorously steam and caramel will harden). Cook mixture over low heat, whisking, until caramel is dissolved. Add plantains and cook, without stirring, until heated through and tender, about 5 minutes. In very small saucepan, heat cream until warm and pour over plantains. Gently shake skillet to incorporate cream into sauce.
- Cool plantains slightly and serve over ice cream.

Acknowledgment

We thank L. Tounsi, K. Ratakonda, J. Wang, E.-E. Jan, and T. Heath from IBM, as well as J. Briscione and M. Laiskonis from ICE. We also thank the panel of experts in the Consensual Assessment Technique.

References

1. R. K. Sawyer, *Explaining Creativity: The Science of Human Innovation*. Oxford, U.K.: Oxford Univ. Press, 2012.
2. M. A. Boden, *The Creative Mind: Myths and Mechanisms*, 2nd ed. London, U.K.: Routledge, 2004.
3. *Capitalizing on Complexity: Insights From the Global Chief Executive Officer Study*, IBM, New York, NY, USA, May 2010.
4. B. G. Buchanan, "Creativity at the metalevel," *AI Mag.*, vol. 22, no. 3, pp. 13–28, Fall 2001.
5. S. Colton, R. López de Mántaras, and O. Stock, "Computational creativity: Coming of age," *AI Mag.*, vol. 30, no. 3, pp. 11–14, Fall 2009.
6. A. Cardoso, T. Veale, and G. A. Wiggins, "Converging on the divergent: The history (and future) of the international joint workshops in computational creativity," *AI Mag.*, vol. 30, no. 3, pp. 15–22, Fall 2009.
7. S. Colton and G. A. Wiggins, "Computational creativity: The final frontier?," in *Proc. 20th Eur. Conf. Artif. Intell.*, 2012, pp. 21–26.
8. M. A. Boden, "Foreword," in *Computational Creativity Research: Towards Creative Machines*, T. R. Besold, M. Schorlemmer, and A. Smaill, Eds. New York, NY, USA: Springer, 2015, pp. v–xiii.
9. V. Leschziner, *At the Chef's Table: Culinary Creativity in Elite Restaurants*. Stanford, CA, USA: Stanford Univ. Press, 2015.
10. M. Csikszentmihalyi, "Artistic problems and their solutions: An exploration of creativity in the arts," Ph.D. dissertation, Committee on Human Dev. Univ. Chicago, Chicago, IL, USA, Jun. 1965.
11. J. W. Getzels and M. Csikszentmihalyi, *The Creative Vision: Longitudinal Study of Problem Finding in Art*. New York, NY, USA: Wiley, 1976.
12. T. M. Amabile, "Social psychology of creativity: A consensual assessment technique," *J. Pers. Soc. Psychol.*, vol. 43, no. 5, pp. 997–1013, Nov. 1982.
13. L. R. Varshney, F. Pinel, K. R. Varshney, et al., "Cognition as a part of computational creativity," in *Proc. 12th IEEE Int. Conf. Cogn. Inform. Cogn. Comput.*, Jul. 2013, pp. 36–43.
14. F. Pinel and L. R. Varshney, "Computational creativity for culinary recipes," in *Proc. SIGCHI Conf. Human Factors Comput. Syst.*, Apr. 2014, pp. 439–442.
15. R. G. Mo rris, S. H. Burton, P. Bodily, et al., "Soup over beans of pure joy: Culinary ruminations of an artificial chef," in *Proc. Int. Conf. Comput. Creativity*, May 2012, pp. 119–125.
16. J. Pagnutti and J. Whitehead, "Generative mixology: An engine for creating cocktails," in *Proc. Int. Conf. Comput. Creativity*, Jun. 2015, pp. 212–219.
17. Á. Amorim, C. França, L. F. W. Góes, et al., "Creative flavor pairing: Using RDC metric to generate and assess ingredients combinations," in *Proc. Int. Conf. Comput. Creativity*, Jun. 2017.
18. J. C. Kaufman and J. Baer, *Creativity Across Domains: Faces of the Muse*. London, U.K.: Psychology Press, 2004.
19. S. H. Carson, J. B. Peterson, and D. M. Higgins, "Reliability, validity, and factor structure of the creative achievement questionnaire," *Creativity Res. J.*, vol. 17, no. 1, pp. 37–50, 2005.
20. G. A. Wiggins, "Searching for computational creativity," *New Gener. Comput.*, vol. 24, no. 3, pp. 209–222, Sep. 2006.
21. M. A. Boden, "The Turing test and artistic creativity," *Kybernetes*, vol. 39, no. 3, pp. 409–413, 2010.
22. M. Schmidt and H. Lipson, "Distilling free-form natural laws from experimental data," *Science*, vol. 324, no. 5923, pp. 81–85, Apr. 2009.
23. R. D. King, J. Rowland, S. G. Oliver, et al., "The automation of science," *Science*, vol. 324, no. 5923, pp. 85–89, Apr. 2009.
24. K. R. Varshney, L. R. Varshney, J. Wang, et al., "Flavor pairing in Medieval European cuisine: A study in cooking with dirty data," in *Proc. Int. Joint Conf. Artif. Intell. Workshop Cooking Comput.*, Aug. 2013, pp. 3–12.
25. L. Itti and P. Baldi, "Bayesian surprise attracts human attention," in *Proc. Adv. Neural Inf. Process. Syst. 18*, 2006, pp. 547–554.
26. L. Itti and P. Baldi, "Bayesian surprise attracts human attention," *Vis. Res.*, vol. 49, no. 10, pp. 1295–1306, Jun. 2009.
27. P. Baldi and L. Itti, "Of bits and wows: A Bayesian theory of surprise with applications to attention," *Neural Netw.*, vol. 23, no. 5, pp. 649–666, Jun. 2010.
28. L. R. Varshney, "To surprise and inform," in *Proc. IEEE Int. Symp. Inf. Theory*, Jul. 2013, pp. 3145–3149.
29. C.-Y. Teng, Y.-R. Lin, and L. A. Adamic, "Recipe recommendation using ingredient networks," in *Proc. 3rd Annu. ACM Web Sci. Conf.*, Jun. 2012, pp. 298–307.
30. J.-S. Horng and M.-L. Hu, "The mystery in the kitchen: Culinary creativity," *Creativity Res. J.*, vol. 20, no. 2, pp. 221–230, May 2008.
31. C. Fry, "Closing the gap between analytics and action," *INFORMS Analytics Mag.*, vol. 4, no. 6, pp. 4–5, Nov./Dec. 2011.
32. G. Smith, J. Whitehead, and M. Mateas, "Computers as design collaborators: Interacting with mixed-initiative tools," in *Proc. ACM Creativity Cognition Worshop Semi-Autom. Creativity*, Nov. 2011.
33. T. Lubart, "How can computers be partners in the creative process," *Int. J. Human-Comput. Studies*, vol. 63, no. 4–5, pp. 365–369, Oct. 2005.

34. A. Newell, J. C. Shaw, and H. A. Simon, "The process of creative thinking," in *Contemporary Approaches to Creative Thinking*, H. E. Gruber, G. Terrell, and M. Wertheimer, Eds. New York, NY, USA: Atherton, 1963, pp. 63–119.
35. A. K. Jordanous, "Evaluating computational creativity: A standardised procedure for evaluating creative systems and its application," Ph.D. dissertation, Dept. Informatics, Univ. Sussex, Brighton, U.K., Dec. 2012.
36. D. Bhattacharjya, "Preference models for creative artifacts and systems," in *Proc. Int. Conf. Comput. Creativity*, Jun. 2016, pp. 52–59.
37. A. Pease and S. Colton, "On impact and evaluation in computational creativity: A discussion of the Turing test and an alternative proposal," in *Proc. Symp. Comput. Philosophy*, Apr. 2011, pp. 15–22.
38. T. M. Amabile, "Children's artistic creativity: Detrimental effects of competition in a field setting," *Pers. Soc. Psychol. Bull.*, vol. 8, no. 3, pp. 573–578, Sep. 1982.
39. J. Baer, *Creativity and Divergent Thinking: A Task-Specific Approach*. Hillsdale, NJ, USA: Lawrence Erlbaum, 1993.
40. J. C. Kaufman, J. Lee, J. Baer, and S. Lee, "Captions, consistency, creativity, and the consensual assessment technique: New evidence of reliability," *Thinking Skills Creativity*, vol. 2, no. 2, pp. 96–106, Nov. 2007.
41. J. C. Kaufman, J. Baer, and J. C. Cole, "Expertise, domains, and the consensual assessment technique," *J. Creative Behav.*, vol. 43, no. 4, pp. 223–233, Dec. 2009.
42. J. C. Kaufman, J. Baer, J. C. Cole, et al., "A comparison of expert and nonexpert raters using the consensual assessment technique," *Creativity Res. J.*, vol. 20, no. 2, pp. 171–178, 2008.
43. B. Uzzi, S. Mukherjee, M. Stringer, et al., "Atypical combinations and scientific impact," *Science*, vol. 342, no. 6157, pp. 468–472, Oct. 2013.
44. C. E. Shannon, "The redundancy of English," in *Proc. Trans. 7th Conf. Cybern.*, Mar. 1950, pp. 123–158.
45. M. O. Riedl and R. M. Young, "Narrative planning: Balancing plot and character," *J. Artif. Intell. Res.*, vol. 39, pp. 217–268, 2010.
46. O. Kinouchi, R. W. Diez-Garcia, A. J. Holanda, et al., "The non-equilibrium nature of culinary evolution," *New J. Phys.*, vol. 10, 073020, 2008.
47. K. Veeramachaneni, E. Vladislavleva, and U.-M. O'Reilly, "Knowledge mining sensory evaluation data: Genetic programming, statistical techniques, and swarm optimization," *Genet. Program. Evolvable Mach.*, vol. 13, no. 1, pp. 103–133, Mar. 2012.
48. F. Pinel, L. R. Varshney, and D. Bhattacharjya, "A culinary computational creativity system," in *Computational Creativity Research: Towards Creative Machines*, T. R. Besold, M. Schorlemmer, and A. Smaill, Eds. New York, NY, USA: Springer, 2015, pp. 327–346.
49. L. R. Varshney, J. Wang, and K. R. Varshney, "Associative algorithms for computational creativity," *J. Creative Behav.*, vol. 50, no. 3, pp. 211–223, Sep. 2016.
50. Y.-Y. Ahn, S. E. Ahnert, J. P. Bagrow, et al., "Flavor network and the principles of food pairing," *Sci. Rep.*, vol. 1, Dec. 2011, Art. no. 196.
51. R. Haddad, A. Medhanie, Y. Roth, et al., "Predicting odor pleasantness with an electronic nose," *PLoS Comput. Biol.*, vol. 6, no. 4, Apr. 2010, Art. no. e1000740.
52. W. Swart and L. Donno, "Simulation modeling improves operations, planning, and productivity of fast food restaurants," *Interfaces*, vol. 11, no. 6, pp. 35–47, Dec. 1981.
53. D. Tasse and N. A. Smith, "SOUR CREAM: Toward semantic processing of recipes," Carnegie Mellon Univ., Pittsburgh, PA, USA, Tech. Rep. CMU-LTI-08-005, May 2008.
54. E. T. Mueller, *Commonsense Reasoning*. San Francisco, CA, USA: Morgan Kaufmann, 2006.
55. R. M. Khan, C. H. Luk, A. Flinker, et al., "Predicting odor pleasantness from odorant structure: Pleasantness as a reflection of the physical world," *J. Neurosci.*, vol. 27, no. 37, pp. 10 015–10 023, Sep. 2007.
56. Q. Wang, S. R. Kulkarni, and S. Verdú, "Universal estimation of information measures for analog sources," *Found. Trends Commun. Inf. Theory*, vol. 5, no. 3, pp. 265–353, 2009.
57. A. Orlitsky, N. P. Santhanam, and J. Zhang, "Always Good Turing: Asymptotically optimal probability estimation," *Science*, vol. 302, no. 5644, pp. 427–431, Oct. 2003.
58. G. M. Shepherd, "Smell images and the flavour system in the human brain," *Nature*, vol. 444, no. 7117, pp. 316–321, Nov. 2006.
59. G. A. Burdock, *Fenaroli's Handbook of Flavor Ingredients*. Boca Raton, FL, USA: CRC Press, 2009.
60. K. Snitz, A. Yablonka, T. Weiss, et al., "Predicting odor perceptual similarity from odor structure," *PLoS Comput. Biol.*, vol. 9, no. 9, Sep. 2013, Art. no. e1003184.
61. A. Keller, R. C. Gerkin, Y. Guan, et al., "Predicting human olfactory perception from chemical features of odor molecules," *Science*, vol. 355, no. 6327, pp. 820–826, Feb. 2017.
62. H. Lapid, D. Harel, and N. Sobel, "Prediction models for the pleasantness of binary mixtures in olfaction," *Chem. Senses*, vol. 33, no. 7, pp. 599–609, Sep. 2008.
63. D. M. Blei, "Probabilistic topic modeling," *Commun. ACM*, vol. 55, no. 4, pp. 77–84, Apr. 2012.
64. D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent Dirichlet allocation," *J. Mach. Learn. Res.*, vol. 3, pp. 993–1022, Mar. 2003.
65. V. Nedovic, "Learning recipe ingredient space using generative probabilistic models," in *Proc. Int. Joint Conf. Artif. Intell. Workshop Cooking Comput.*, Aug. 2013, pp. 13–18.
66. IBM and Institute of Culinary Education, *Cognitive Cooking with Chef Watson*. Naperville, IL, USA: Sourcebooks, 2015.
67. A. P. Davis, "Digital gastronomy: When an IBM algorithm cooks, things get complicated–and tasty," *Wired*, vol. 20, no. 10, Oct. 2013.
68. J. McNerney, J. D. Farmer, S. Redner, et al., "Role of design complexity in technology improvement," *Proc. Nat. Acad. Sci. USA*, vol. 108, no. 22, pp. 9008–9013, May 2011.
69. G. Basalla, *The Evolution of Technology*. New York, NY, USA: Cambridge Univ. Press, 1988.

Received December 11, 2017; accepted for publication December 21, 2018

Lav R. Varshney *University of Illinois at Urbana-Champaign, Urbana, IL 61801 USA (varshney@illinois.edu).* Dr. Varshney received a B.S. degree in electrical and computer engineering from Cornell University, Ithaca, NY, USA, in 2004, and S.M., E.E., and Ph.D. degrees in electrical engineering and computer science from the Massachusetts Institute of Technology, Cambridge, MA, USA, in 2006, 2008, and 2010, respectively. He is an Assistant Professor with the Department of Electrical and Computer Engineering, the Coordinated Science Laboratory, the Department of Computer Science (by courtesy), and the Neuroscience Program, University of Illinois at Urbana-Champaign, Urbana, IL, USA. His research focuses on information theory, neuroscience, and artificial intelligence. He appears on the List of Teachers Ranked as Excellent and has been named a Center for Advanced Study Fellow at the University of Illinois. He currently serves on the advisory board of the AI XPRIZE.

Florian Pinel *IBM Watson, Yorktown Heights, NY 10598 USA (pinel@us.ibm.com).* Mr. Pinel received a Culinary Arts diploma from the Institute of Culinary Education, New York, NY, USA, and a M.S. degree in computer science and engineering from Ecole Centrale de Paris, Paris, France. He is a Senior Technical Staff Member and Master Inventor with the Watson Group, IBM, Yorktown Heights, NY, USA. Before joining Watson, he was with the IBM T. J. Watson Research Center, focusing on business process management, IT services management, and software as a service. He is one of the inventors of *IBM Chef Watson*, an application that uses machine learning and natural language processing to demonstrate computational creativity and suggest original recipe ideas. He currently works on intelligent video analytics and machine learning model management. He coauthored two books, *Cognitive Computing with Chef Watson*, and *Towards Creative Machines*, and holds about 20 patents, mostly in the field of artificial intelligence and its applications. He is a member of the IBM Academy of Technology and of the IBM Industry Academy. He also writes a blog about Eastern European cuisine, *Food Perestroika* (www.foodperestroika.com).

Kush R. Varshney *IBM Research, Thomas J. Watson Research Center, Yorktown Heights, NY 10598 USA (krvarshn@us.ibm.com)*. Dr. Varshney was born in Syracuse, NY, USA, in 1982. He received a B.S. degree (magna cum laude) in electrical and computer engineering with honors from Cornell University, Ithaca, NY, USA, in 2004, and S.M. and Ph.D. degrees in electrical engineering and computer science from the Massachusetts Institute of Technology (MIT), Cambridge, MA, USA, in 2006 and 2010, respectively. While at MIT, he was a National Science Foundation Graduate Research Fellow. He is a Principal Research Staff Member and Manager with IBM Research AI, Thomas J. Watson Research Center, Yorktown Heights, NY, USA. He is the founding Co-Director of the IBM Science for Social Good initiative. He applies data science and predictive analytics to human capital management, healthcare, olfaction, computational creativity, public affairs, international development, and algorithmic fairness, which has led to recognitions such as the 2013 Gerstner Award for Client Excellence for contributions to the WellPoint team and the Extraordinary IBM Research Technical Accomplishment for contributions to workforce innovation and enterprise transformation. He conducts academic research on the theory and methods of statistical signal processing and machine learning. His work has been recognized through best paper awards at the Fusion 2009, SOLI 2013, KDD 2014, and SDM 2015 conferences. He is a Senior Member of the IEEE and a member of the Partnership on AI's Safety-Critical AI working group.

Debarun Bhattacharjya *IBM Research, Thomas J. Watson Research Center, Yorktown Heights, NY 10598 USA (debarunb@us.ibm.com)*. Dr. Bhattacharjya received a B. Tech. degree in industrial and production engineering from the Indian Institute of Technology (IIT), Hauz Khas, Delhi, India, and M.S and Ph.D. degrees in management science and engineering from Stanford University, Stanford, CA, USA. He is a Research Staff Member with IBM Research AI, Thomas J. Watson Research Center, Yorktown Heights, NY, USA. His primary research interests lie in probabilistic modeling and decision analysis, as applied to various problems in artificial intelligence, machine learning, and management science. His research interests include topics such as graphical models, event models, value of information, preference elicitation, and computational creativity. His applied work has spanned several domains, including energy, sales, business services, transportation, consumer products, and public policy. He has coauthored one book, *Value of Information in the Earth Sciences,* and more than 30 publications in highly refereed journals and conference proceedings.

Angela Schörgendorfer *IBM Research, Thomas J. Watson Research Center, Yorktown Heights, NY 10598 USA (angelasch@google.com)*. Dr. Schörgendorfer received an M.S. degree in computer science from the Vienna University of Technology, Vienna, Austria, and a Ph.D. degree in statistics from the University of Kentucky, Lexington, KY, USA. She is a Senior Data Scientist with Google Inc., Mountain View, CA, USA, leading a data science team working on Google Maps. Previously, she was a Researcher with the IBM T. J. Watson Research Center. Her research has focused on spatial and spatiotemporal statistical modeling, environmental research, and Bayesian nonparametric methodology.

Yi-Min Chee *IBM Research, Thomas J. Watson Research Center, Yorktown Heights, NY 10598 USA (ymchee@us.ibm.com)*. Mr. Chee received a bachelor's degree from the Massachusetts Institute of Technology, Cambridge, MA, USA, and a master's degree from Columbia University, New York, NY, USA, both in computer science. He is a Senior Technical Staff Member with Blockchain Solutions, IBM Industry Platforms Group, Yorktown Heights, NY, USA. Since joining IBM, he has worked in a variety of areas with the IBM T. J. Watson Research Center, ranging from incremental compilers and programming environments for C++, to software development tools, interfaces, and standards for pen-based devices, high-performance computing applications for game processors, design tools and delivery environments for software architects and consultants, and cognitive analytics for learning and computational creativity.