

# Thoughtful Surprise Generation as a Computational Creativity Challenge

Alexandra Coman

Erik T. Mueller

Margaret Mayer

Capital One

8020 Towers Crescent Dr.

Vienna, VA 22182 USA

alexandra.coman | erik.mueller | margaret.mayer@capitalone.com

## Abstract

Thoughtful acts are among the most valued products of human social behavior. They serve enculturation and the perpetuation of kindness, and often exemplify everyday creativity. We propose the thoughtful surprise generation problem as a computational creativity problem, and briefly outline a Turing Test–alternative challenge for evaluating AI agents’ ability to produce thoughtful acts.

## Introduction

We begin with a fictional but plausible story. Emma, a customer service assistant at a bank, receives a call from Claire, who has lost her credit card while traveling overseas. The conversation begins banally but leads to a childhood memory of Claire’s, which reminds Emma of a passage from a favorite book. She thinks Claire may appreciate the connection. She considers mentioning this to Claire, but then comes up with what she thinks is a better idea. After the banking problem is resolved and the conversation ends, Emma orders a copy of the book and sends it to Claire as a present, with an explanatory note. Maybe it will brighten Claire’s mood after having had to deal with the lost card issue. Maybe it will inspire her to do something kind for someone else. While it may seem unusual for such an interaction culminating in a thoughtful surprise to take place between a bank representative and a customer, consider the story (WTVR.com, 2016), reported by multiple media outlets, of a Capital One customer service representative who sent flowers and gifted travel miles to a customer who had revealed her difficult personal situation during a conversation about a banking issue. The surprise was extremely well received by the customer, who reported that it had “changed [her] life”. This is not an isolated occurrence; the bank empowers their customer service representatives to take actions like these when appropriate given the rapport established during the conversation.

The ability to come up with thoughtful acts for others (loved ones, acquaintances, customers, near strangers) leads to some of the most valued instances of human social behavior. It requires using knowledge of various types, reasoning, and emotional intelligence to identify situations in which such acts are opportune, adapt acts to the person(s) they are directed toward and to the situations that prompt them, and

behave cautiously so as to maintain unexpectedness. Usually, such acts are spontaneous, autotelic, and drawn from an unconstrained solution space. We believe that they are often described, informally, by receivers and observers, as “creative”.

So far, the generative subfield of computational creativity has dealt mostly with producing artistic artifacts and performances, e.g., narratives, music, visual art, poetry, choreography, and various aspects of computer games (Loughran and O’Neill 2017). While more mundane than art, thoughtful surprise generation is arguably more universally human, as it does not require exceptional skills or talent (although, if available, skills and talent can serve to enhance surprises, e.g., Emma might have written her note in verse). Thoughtful surprises are products of everyday creativity (O’Neill and Riedl 2011).

We propose thoughtful surprise generation as a computational creativity problem, and briefly outline a modular challenge for evaluating an AI agent’s ability to spontaneously generate thoughtful acts based on customer stories or dialogue. This challenge is to be included in a broader financial dialogue challenge for AI banking assistants, an alternative to the Turing Test (Turing 1950). Our focus is, hence, on the characteristics of the problem rather than on any particular solution, though we use hypothetical agents with various AI capabilities for exemplification throughout the paper.

In terms of practical relevance, virtual assistants with thoughtful surprise generation capabilities could create value for companies through richer customer interaction. More broadly, progress in this direction is also progress toward machine enculturation (Riedl 2016), as it requires AI to be informed by social norms and aligned with human goals.

We define a thoughtful surprise as an act that is (a) directed toward another person, (b) intended to have a positive impact on the person it is directed toward, and (c) intended to be unexpected by the person it is directed toward. In human interaction, such acts include: offering gifts, creating personalized mixtapes, and writing poetry inspired by the recipient. In addition, we require that the act be accompanied by framing (Charnley, Pease, and Colton 2012), both customer-directed (in the form of a note addressed to the customer) and process-related (revealing the system’s creative processes, thus demonstrating its intentionality). While

we use banking-related conversations in our examples (hence, the “customer” and “agent” terms we use to refer to the two conversation partners), the challenge is generalizable to any dialogue context.

In the following sections, we (1) briefly survey related work, (2) describe our proposed challenge problem, (3) propose several different challenge modules, (4) describe the types of required framing, (5) show how thoughtful surprise generation qualifies as a computational creativity problem, (6) describe a general process for surprise-preserving dialogue, (7) propose evaluation methods for the challenge, and (8) end with several open issues.

## Related Work

Various alternatives to the Turing Test have been proposed for evaluating abilities that can be characterized as types of creativity, such as the ability to generate stories (e.g., Riedl 2014). Jarrold and Yeh (2016) have outlined a social-emotional Turing Challenge; the evaluated AI agent must, among others, attempt to identify the feeling most likely to be experienced by a character in a short story presented to it. Such AI empathy is highly relevant to our challenge, as thoughtful acts should improve their recipient’s mood.

Pease and Colton (2011) argue against the appropriateness of the Turing Test for computational creativity tasks, stating, among others, that “there are huge philosophical problems with using a test based on imitation to evaluate competence in an area of thought which is based on originality”. We note that the originality requirements of thoughtful acts are more modest than those of artistic artifacts. Even at their most original, such acts must be socially understandable and palatable, so it could be argued that this problem is more amenable to Turing Test-like approaches than general artistic creativity. Still, our challenge, while not yet calculated to perfectly fit into any preexisting framework, shares with the FACE and IDEAS (Colton, Pease, and Charnley 2011), and SPECS (Jordanous 2012) models the approach of systematically describing task-relevant creativity aspects at which to target evaluation approaches, rather than requiring vague indistinguishability from humanly-generated artifacts.

Gil (2017) uses the term “thoughtful AI” in a broader sense than we do herein. Further related work will be mentioned where relevant throughout the paper.

## The Thoughtful Surprise Generation Problem

The main input of the proposed thoughtful surprise generation challenge is an input discourse which makes it possible to identify relevant information about the customer, including their preferences, biographical information, and current life circumstances. This information should allow the agent to (1) identify opportunities for thoughtful acts (e.g., finding out about a customer’s upcoming anniversary or about their favorite childhood candy that they have not been able to find in a while), and (2) generate suitable thoughtful acts. In certain variants of the problem, a solution space will also be part of the input. In addition, the input should also include

any necessary constraints (e.g., company guidelines restricting what a customer service representative may do in terms of thoughtful acts).

The output includes the thoughtful act itself (as a list of features or a natural language description, depending on the solution space) and framing.

## Problem Dimensions

**Input discourse.** We propose two types of input discourse: stories and dialogue. Stories are of a particular type: customer stories in the first person. In the case of dialogue, one of the participants is the customer.

**Solution Space.** In terms of solution space, a thoughtful surprise generation task can be constrained or unconstrained.

When the solution space is constrained, a thoughtful act needs to be selected from a provided solution space, small or large. For example, a customer service representative might be required to choose a gift for a customer from the available stock of an approved vendor. The types of acceptable acts are restricted in this case as well (e.g., to giving gifts). We propose two variants of the constrained solution space: the multiple-choice variant and the full-solution space variant. In the full-solution space variant, the agent may choose an act from the entire solution space. In the multiple-choice variant, the agent is required to select the most appropriate act out of several available options, preselected from the full solution space. This variant has implications for evaluation, e.g., one of the options might already have been identified as being “the best”, and “trap” options for thwarting known strategies for gaming the test may have been included (e.g., in Fig. 1 (1.2)(A), options (e), (g), and (j) are among the ones meant to trick unsophisticated, bag-of-words-type approaches to the challenge). The example selection in Fig. 1 (1.2)(A), (h), sets a rather high standard in that it requires complex application of cultural knowledge (as explained in Fig. 3), but it is included herein to exemplify the range of creativity that could potentially be demonstrated by contestant agents.

In the unconstrained variant, no solution space is provided as part of the input; the act may consist of any sequence of actions at all, just like human thoughtful acts do, e.g., writing and/or reciting a poem, or creating a mixtape of songs relevant to the recipient. The full spectrum of creativity is now at the agent’s disposal, should the agent be able to make use of it. In this case, the generated act will be a natural language construct describing a sequence of actions, which, semantically, is equivalent to a plan (Ghallab, Nau, and Traverso 2004). While contestants may adopt planning approaches for act generation, we will not provide planning-domain information (e.g., operators, with preconditions and postconditions), so any such information would have to be acquired by the solution designers and/or agents.

**Interactivity.** The two values of this dimension are: interactive and non-interactive.

In the non-interactive variant, the agent is presented with a static text, either a story or a dialogue snippet, and must produce a thoughtful act and framing based on the

story/dialogue. The agent is, therefore, not involved in the production of the input discourse.

In the interactive variant, the agent is actively engaged in dialogue with the customer, and can use this interaction to elicit additional information that can help it better adapt the thoughtful act to the customer (e.g., Fig. 1 (1.1)(b), assuming A is the actual agent that produces the acts). The agent thus has the ability to influence the input discourse. Different questions or remarks at any point in the conversation can lead to different dialogue paths. In the example in Fig. 1, the agent, on being told about the trip to France, might have asked (instead of “Any good food?”) “What was your favorite thing about France?” or “Does your sister live in France or did she just have her wedding there?”, possibly leading the conversation toward other, more or less specific and salient, customer information.

It should be noted that the interactive version of the challenge requires the evaluated system to be capable of conducting dialogue with a customer by processing and producing utterances in a goal-directed manner, i.e., it should be a dialogue system in its own right.

## Challenge Modules

We propose increasingly complex challenge modules, based on different combinations of values of two of the previously-introduced dimensions: input discourse and interactivity (Fig. 2). All three modules below can be administered with any of the three solution-space variants.

**Non-interactive, story-based.** The main input is a first-person customer story. The output is a thoughtful act accompanied by framing.

**Non-interactive, dialogue-based.** The input is a non-interactive dialogue in which one of the participants is the customer. The output is a thoughtful act accompanied by framing. The situation is similar to one in which a trainee demonstrates their ability to reason about a hypothetical scenario (e.g., “If I were this agent, I would do this”).

**Interactive, dialogue-based.** The agent is actively involved in the dialogue. In this case, we have two types of input/output: intermediary and final. On every step of the dialogue, the input is a customer utterance. As intermediary output, an utterance advancing the dialogue is presented to the customer, and the current intention regarding thoughtful acts is presented to an external-observer evaluator. The intention can indicate that the agent is (a) not currently planning thoughtful acts, or (b) in the process of generating thoughtful acts (candidate acts are also provided). Final output is provided after the conversation has ended, and consists of the finalized thoughtful act (if any) and framing.

## Framing Types

In the context of computational creativity theory, framing is defined by Colton, Pease, and Charnley (2011) as “a piece of natural language text that is comprehensible by people, which refers to [generative acts].”

(1.1) (a) I’d like to report that I lost my credit card. I’m sorry I didn’t do this sooner, but we were in France for my sister’s wedding, and I didn’t have my cellphone with me because I can’t use it overseas.

(b) 1//C: Hi! I’d like to report that I lost my credit card. I’m sorry I didn’t do this sooner, but we were in France for my sister’s wedding, and I didn’t have my cellphone with me because I can’t use it overseas.

A: [after eliciting C’s account information] France, huh? I’m jealous! Any good food?

2//C: Oh, the best cakes ever. And, um, this chicken, haha. With lots of vinegar. I think it’s the first dish with lots of vinegar in it that I’ve ever actually liked.

A: You usually dislike vinegar?

3//C: Hm. Maybe I like the smell more than the taste. It reminds me of Christmas ☺.

A: That’s unusual! Why?

4//C: Well ... my grandma used to douse all her jewelry in vinegar one week before Christmas, every year. Always one week before, I don’t know why. Her whole room would smell of it.

A: Wow, I think I’ll steal that jewelry-cleaning tip from your grandma ☺... that vinegar chicken you mentioned sounds good, too. You got the recipe ☺?

5//C: No, I’d never make it for myself. That’s no fun!

[The conversation continues, and the customer’s banking issue is resolved.]

(1.2)(A) **Search Space:** (a) French recipe cookbook, (b) France travel guide, (c) chicken recipe cookbook, (d) book of house-keeping tips, (e) bottle of jewelry-cleaning liquid, (f) bottle of vinegar, (g) copy of “A Christmas Carol” by Charles Dickens, (h) copy of “In Search of Lost Time: Vol. 1– Swann’s Way” by Marcel Proust, (i) strawberry cake, (j) cellphone, (k) bottle of perfume, (l) bouquet of flowers

**Thoughtful Act:** (h)

(B) **Search Space:** books on an e-commerce website, described by title and author name(s)

**Thoughtful Act:** (“In Search of Lost Time: Vol. 1 – Swann’s Way”, Marcel Proust)

(C) **Search space:** Unconstrained

**Thoughtful Act:** “I am going to send the customer a copy of “In Search of Lost Time: Vol. 1” by Marcel Proust as a gift.”

Figure 1. (1.1) The types of input discourse: (a) customer story and (b) dialogue (A - agent, C - customer). (1.2) The three types of solution spaces, with thoughtful act examples: (A) constrained, multiple-choice, (B) constrained, full solution space, (C) unconstrained.

Framing can include information about the creative process, among others. In our case, framing will be mostly external (Charnley, Pease, and Colton 2012), as thoughtful surprise generation is a particularly audience-centric creative act. All three areas of framing described by the authors (motivation, intention, and processes) are reflected in the framing we require. Their dually-creative approach to framing is

applicable here, but with a very significant requirement change: the framing must be factually correct.

The types of framing output that we require are: process-related (intermediary and final) and customer-directed (see Fig. 3 for examples).

**Process-related framing** is directed at evaluators acting as external observers of the creative process, and is inaccessible to the customer. Process-related framing must reflect the decision-making that occurs during surprise generation, including how the process was triggered. In the interactive version of the challenge, intermediary process-related framing can be provided during the interaction, thus illuminating the iterative refinement of surprises. This type of framing must be in natural language, but the language can be very simple. Other than that, we do not, at the moment, plan to impose any structural requirements onto process-related framing, as it will reflect the characteristics of the creative agent that generates it.

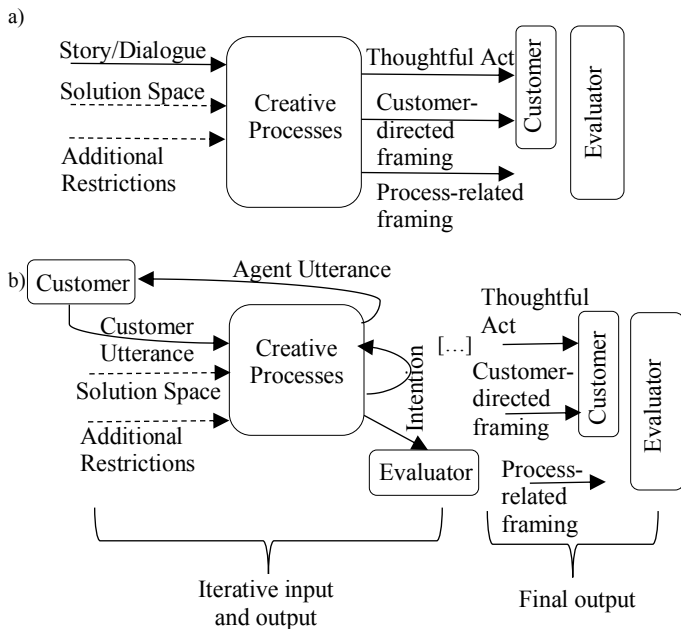


Figure 2. Input and output for (a) the non-interactive and (b) the interactive variants of the challenge. The dashed lines indicate optional input.

**Customer-directed framing** plays a role similar to that of notes accompanying gifts. For the purposes of the challenge, this type of framing must contain at least the following components: (1) acknowledgement of the conversation or story that triggered the thoughtful act generation, and (2) an explanation of the thoughtful act in relation to the content of the story/conversation. When based on static dialogue, customer-directed framing should be written from the perspective of the agent involved in the dialogue.

While process-related framing need merely be human-readable, customer-directed framing is held to the same standards as a gift note written by a human (e.g., it should flow well, be grammatically correct, and be sufficiently

informative). Certain pieces of information from process-related framing may be inappropriate for customer-directed framing. For example, it would probably be inappropriate for the note in Fig. 3 to contain the text: “I first thought about giving you a French recipe cookbook, but then I found out that you dislike cooking”.

**Internal representation (partial):**

Trigger: trip to France  
 Analogy (customer story, madeleine episode in “In Search of Lost Time”): smell of vinegar → taste of madeleine, memories of Christmases with grandma → memories of Aunt Léonie  
 Additional relevant feature: isSettingOf(“In Search of Lost Time”, France)

Additional salient customer information: dislikes (cooking) → rejected surprise “offer French cookbook”, likes(French\_food)

**Process-related framing (partial):** “I decided to initiate surprise generation when I heard about the customer’s trip to France. The surprise is relevant for this customer because [...]. I first thought of giving the customer a French recipe cookbook, but then I found out that she dislikes cooking. I think the customer’s mood will be improved by a gift that reminds her of France because, overall, she seemed to enjoy the trip.”

**Customer-directed framing:** “Dear Claire, [acknowledgement of the conversation containing the surprise trigger] I really enjoyed talking to you about your trip to France! [explanation of the surprise in relation to the conversation] Your story about how the smell of vinegar reminds you of Christmases with your grandma made me think of Proust’s story about how the taste of a madeleine dipped in tea brought back childhood memories of his aunt. I hope that you enjoy reading this book and that it reminds you of France 😊!”

Figure 3. Internal knowledge representation, process-related framing, and customer-directed framing examples for a hypothetical agent capable, among others, of simplified analogical mapping and commonsense reasoning, and in possession of cultural knowledge.

**Computational Creativity Criteria**

Creative processes are often described as producing artifacts that are novel, valuable, and unexpected (Boden 1990). Intentionality is an additional criterion in the literature (Ventura 2016). We now outline how these criteria are applicable to our challenge.

**Novelty.** The contestant agents are expected to be P-creative (Boden 1990), i.e., produce results that are novel to the agent producing them. The novelty required by the task is more obvious in the unconstrained-solution space variant, in which the agent is required to fully synthesize a thoughtful act from scratch. In the two constrained-solution space variants, in which surprises are selected, rather than fully synthesized, the novelty and, thus, creativity, lie in the connection between the input discourse and the surprise, as expressed in the framing.

**Value.** To be considered thoughtful, an act must at the very least: (1) be socio-emotionally positive, i.e., be likely, based on all available information, to have a positive effect on the

customer's mood (e.g., if the customer in Fig. 1 had indicated that she had not enjoyed the trip, a gift reminding her of it would have been inappropriate), and (2) be demonstrably rooted in the information provided by the customer, and appropriately justified. The act must demonstrate no misunderstanding or willful disregard of the provided information (e.g., in Fig. 1, the French recipe cookbook, although relevant to the conversation, would be a gift that shows disregard or ignorance of the customer's expressed preference not to cook).

**Unexpectedness.** The central role of unexpectedness in our challenge will be discussed in detail in the next subsection. Unexpectedness has been explored in computational creativity (e.g., Grace and Maher 2016; Yannakakis and Liapis 2016). Our challenge differs in that, in addition to generating surprising thoughtful acts, agents must maintain the very fact that a thoughtful act is being planned surprising. Agent utterances in the interactive version of the task need to be targeted both at increasing value (by acquiring relevant information for thoughtful act refinement) and at maintaining unexpectedness (by not revealing the thoughtful intentions). Also, there may be no obvious set of expectations against which to evaluate the unexpectedness of generated acts. Finally, like Pickering and Jordanous (2017) in storytelling, we are interested in surprising others, rather than in the creator's self-surprise.

**Intentionality** is defined by Ventura (2016) as “the fact of being deliberative or purposive; that is, the output of the system is the result of the system having a goal or objective—the system's product is correlated with its process.” Our agents are expected to demonstrate their intentionality through framing, particularly process-related framing (Fig. 3).

### The Thoughtful Surprise Generation Process

Without intending to constrain the ways in which an agent can approach the proposed tasks, we broadly envision a general thoughtful surprise generation process that might be conducted by agents engaged in the interactive version of the challenge. Of course, specific agents might approach parts of the process in other ways than exemplified herein, or their overall approach may be very different from what we anticipate. However, we believe that providing this general process can help guide the identification of capabilities needed by agents that might engage in such a challenge. We also do so in order to highlight the particular characteristics of this type of dialogue, which, among others, should be surprise-preserving.

For the exemplification purposes of this subsection, we assume a general conversation between an agent and a customer, not necessarily within the context of a competition. We do foresee contextual differences between “real-life” banking dialogue and competition situations: in the case of a regular banking-related conversation, any thoughtful act at all would likely be surprising; in a challenge context, the challenge would have to be framed in such a way as to

maintain unexpectedness, e.g. as a general banking dialogue challenge, with occasional thoughtful acts.

The process begins as a regular conversation regarding banking matters. At some point, a trigger identified in a customer utterance causes the agent to decide that a thoughtful act may be opportune, so the agent acquires a thoughtful intention. Either at the same time as acquiring the intention or later on during the dialogue, the agent comes up with one or more candidate thoughtful acts. On generating a candidate act, the agent may immediately be reasonably certain that it is appropriate for the customer (e.g., if the customer utterance is: “I love hazelnut chocolate!”, the agent might decide to order the mentioned treat for the customer). In this case, the agent does not elicit any additional information. However, the candidate act will be abandoned later on if the customer, on their own initiative, provides information that disproves the appropriateness of the act (e.g., C: “Unfortunately, I'm allergic to hazelnuts.”).

On the other hand, if the agent (a) is considerably unsure of the appropriateness of a candidate act (e.g., “Does the customer like chocolate?”), (b) needs more information to fully customize the thoughtful act (e.g., “I know the customer likes chocolate, but which kind?”), and/or (c) needs to choose between several different possible options (e.g., “I know the customer likes plain milk chocolate and hazelnut chocolate, but which does he like more?”), then surprise-preserving dialogue can be conducted, as shown below.

**Triggers.** Triggers are pieces of salient information from customer utterances that cause the initiation of a thoughtful surprise generation process. They are agent-specific, so information ignored by certain agents may be found salient by others. The trigger may immediately provide the agent with a more or less specific idea of what the act(s) might be, or it could simply signal an opportunity for a thoughtful act that the agent then needs to come up with. We do not currently plan to restrict what may constitute a trigger, but propose the following as possible trigger types (which can overlap):

- (1) highly emotionally-charged utterances, of positive or negative valence, identified as such because of (a) the use of emotion-related words, phrases, sentences, punctuation, capitalization, emoticons, etc. (e.g., “I'm having the WORST DAY EVER, you're my last hope ☹☹☹!!!”), or (b) narrative content with emotional implications (e.g., “My flight to France was canceled because of the weather, so I missed my sister's wedding.”)
- (2) utterances that express customer preferences (e.g., “I especially like reading very long books, the more volumes the better!”)
- (3) unexpected utterances (e.g., “the smell of vinegar reminds me of Christmas” as opposed to “the smell of cinnamon reminds me of Christmas”). The utterances may be unexpected in the context of the particular conversation or, more generally, in relation to the agent's entire world knowledge and/or conversational experience.

Alternatively, any of the types of utterances above can provide additional relevant information if they occur after the thoughtful surprise generation process has begun.

**Surprise-preserving dialogue.** While attempting to acquire additional information that can help it refine/select thoughtful acts and/or assess their suitability, the agent must also avoid revealing information that is likely to give away its thoughtful intentions and the specifics of the intended act(s). Therefore, unexpectedness plays two main parts in this process: (1) unexpectedness must be preserved by the information-eliciting utterances, and (2) unexpected utterances by the customer can act as triggers or other salient information for surprise generation. Conversely, the agent’s own information-eliciting utterances should not be unexpected, as this may raise suspicion. Assuming cognitive agents with the abilities to hold beliefs and to reason about the beliefs of others, the task is related to impression management targeted at “changing minds”, as explored by Bridewell and Bello (2014). However, our agents need not change any beliefs of customers. They merely need to avoid introducing two specific kinds of additional beliefs: (a) that the agent is planning a thoughtful act, and (b) what the planned thoughtful act is. With regard to reasoning about the shared context, shared mental models are also relevant (e.g., Magerko., Dohogne, and Fuller, 2011, whose work also exemplifies controlled communicative actions). In surprise-preserving dialogue, the relevant characteristics of agent utterances are: (1) informational content–eliciting potential, (2) surprise-preservation potential (related to Grice’s maxim of quantity, as it involves providing as much information as is needed, and no more (Grice 1967); specifically, if the agent violates Grice’s maxim of quantity, this might strike the customer as peculiar), and (3) context justifiability (i.e., is the utterance expected in the current dialogue context?), which contributes to surprise-preservation potential, and is related to Grice’s maxim of relation (Grice 1967).

The general process we envision includes the following types of steps (several of them exemplified in Fig. 4): (1) generating thoughtful intentions, (2) generating general information-eliciting utterances (necessary when the agent has a thoughtful intention but no partial thoughtful act candidates), (3) generating candidate acts, (4) identifying relevant missing act-related information, (5) generating act-specific information-eliciting utterances, (6) acquiring supporting evidence for a candidate act (such evidence includes customer positive preferences, or “likes”), (7) acquiring contrary evidence for a surprise (e.g., customer negative preferences, or “dislikes”, which may or may not be decisive in abandoning the surprise), (8) abandoning a candidate act, (9) refining a candidate act, (10) masking an utterance intention, so as to preserve the informational content–elicitation potential of the utterance while increasing its context justifiability, (11) abandoning a thoughtful intention, (12) reaching a commitment threshold (i.e., no further information elicitation needed; unless contrary information is provided by the customer, the agent will commit to this act after the conversation ends), and (13) committing to an act (which happens only after conversation has ended, as additional relevant information can come up at any time; reasoning could also occur after the end of the conversation).

[...]  
**[Thoughtful intention: yes; candidate act: FRP]**  
 4//C: Well ... my grandma used to douse all her jewelry in vinegar [...] Her whole room would smell of it.  
**[Generate ISoLT candidate act]**  
**[Generate relevant missing information for FRP: likes(C, cooking)?] [Possible information-eliciting utterance for FRP: “Do you like cooking?” [high ICEP, medium SPP]]**  
**[Mask utterance intention by linking it to conversation context]**  
 A: Wow, I think I’ll steal that jewelry-cleaning tip from your grandma ☺... that vinegar chicken you mentioned sounds good, too. You got the recipe ☺?  
 5//C: No, I’d never make it for myself. That’s no fun!  
**[Decisive contrary evidence acquired for FRP] [Abandon FRP] [Identify relevant missing information for ISoLT: (a) hasRead(C, ISoLT)?, (b) wouldLike(C, ISoLT)?, (b1) likesReading (C)? [...]]**  
**[Possible information-eliciting utterances for ISoLT: (a) “Have you read ISoLT?” [very high ICEP, very low SPP], (b) “Do you like reading?” [high ICEP, low CJ, hence low SPP]]**  
**[Mask utterance intention of b) by linking it to conversation context]**  
 A: Haha! So, I get the food was great. What else did you like about your trip? Any good vacation reading?  
 6//C: Nooo! That’s no fun for me either ☹.  
**[Decisive contrary evidence acquired for ISoLT]**  
**[Abandon ISoLT]**  
**[Abandon thoughtful intention]**

Figure 4. Surprise-preserving dialogue in which additional information is elicited for the refinement of two possible thoughtful acts. Potential utterances are evaluated in terms of informational content–eliciting potential (ICEP), surprise-preservation potential (SPP), and context justifiability (CJ). We do not exemplify how particular scores might be computed, as this should be agent-specific, but assume that SPP is valued higher than ICEP.

We assume that such agents would maintain levels of certainty about the suitability of various thoughtful acts, but choose not to represent these specifically in the examples, as their particularities will be agent-dependent.

We exemplify the surprise-preserving dialogue process in Fig. 4, which shows an extended version of the dialogue in Fig. 1. We assume that the agent has arrived at a point in the conversation where it is considering a French recipe book (FRP) as a possible gift, and will next encounter the customer utterance that makes it also consider “In Search of Lost Time” (ISoLT). The agent reasons that it needs more information to reach its certainty threshold for either one of the candidate acts. It conducts similar processes for the two acts, as shown in the figure. For brevity, we only discuss the process for ISoLT.

The agent reasons that three pieces of information could help it make its decision: (1) does the customer already own this book?, (2) has the customer already read this book?, and (3) more complexly, would this book constitute a good

preference match for the customer? For brevity, let us focus on (2) and (3). In order to resolve (2), the agent might ask “Have you read ISoLT?”. However, this question, while high in informational content–eliciting potential, would be minimally low in surprise-preservation potential. With respect to (3), there is no readily available question whose answer could resolve it. Instead, several questions for eliciting relevant information can be generated, e.g., “Do you like reading?” or, more specifically, “What kind of novels do you like?” or “Do you like early 20<sup>th</sup> century literature?” In our example, the agent settles on the more general question. However, asking this question in its raw form would be conversationally awkward at that stage in the dialogue, as it would have low context justifiability. Instead, the agent masks its intention by incorporating a related question more convincingly into the conversation (i.e., “Haha! So, I get the food was great. What else did you like about your trip? Any good vacation reading?”) The customer’s response causes the ISoLT candidate act to be dropped. Having dropped both candidate acts, the agent chooses to also drop its thoughtful intention. All this deliberation and decision-making should be described in process-related framing. Surprise-preserving dialogue is related to the strategic dialogue in games such as Werewolf, where agents attempt to acquire as much information as possible without revealing their own secrets (Pré-vot et al. 2015). Surprise-preserving dialogue can also be seen as recommendation dialogue with disguised intentions.

## Evaluation Methods

We now briefly describe evaluation methods to be integrated into modules of our challenge. The three types of output (thoughtful acts, customer-directed framing, and process-related framing) can be used in varied ways as part of the challenge evaluation. Herein, we exemplify a few possibilities. There is a major practical distinction between the non-interactive and interactive variants of the challenge, as the latter requires contestant agents to be full-fledged dialogue systems which also have surprise generation capabilities. Such a system would be assessed by humans playing two types of roles: (a) customers interacting directly with the dialogue system, and (b) observers of the conversational exchange and of the intermediary and final process-related framing. Steps would need to be taken to distinguish the evaluation of the quality of the thoughtful acts from that of the agent’s conversational capability.

For the non-interactive, unconstrained variant (Fig. 5), the human evaluators are first presented with the input discourse. After reading it completely, they are shown the thoughtful act. Then, they answer several survey questions based on the input dialogue and thoughtful act. They are then shown the customer-directed framing and (1) answer new questions, about the framing itself and about the connection between the framing and the act, and (2) re-answer the previous act-related questions. Finally, they are shown the process-related framing, and (1) answer questions about this additional framing, and (2) re-answer the initial act-related questions. A subset of the questions are re-asked because answers (e.g., regarding the clarity of the agent’s

reasons for choosing a surprise) may change after reading the framing. Process-related framing can help mitigate the placebo effect (Veale 2015) that can occur when the evaluator is exposed to the more emotionally-involving language of the input discourse and the note, potentially causing them to overestimate the intentionality reflected in an act. We provide several sample evaluation statements in Fig. 5. They assess value, unexpectedness, and the appearance of intentionality. We focus on questions that can be answered by non-expert human evaluators. Some aspects of creativity cannot be evaluated thus, e.g., whether the surprise is truly novel, given its generative process, or whether the agent’s generative process is accurately reflected in its process-related framing. The answers will be subjective (e.g., anticipated negative consequences are likely to be evaluator-specific), but this is the same sort of subjectivity with which thoughtful acts are received in inter-human relationships. In a more Turing Test–like variant of the evaluation, once sufficiently advanced AI agents have been developed, multiple agents, human and AI, can be exposed to the input discourse, and generate surprises and framing; then, human evaluators can attempt to distinguish between the surprise/framing pairs generated by humans and those generated by AI agents.

### a) Thoughtful act

The thoughtful act will have a positive effect on the customer’s mood. [V - Value]

It is clear why the agent chose the act. [I - Intentionality]

The choice of act is unexpected. [U - Unexpectedness]

The act is creative. [V, U, I]

The act is appropriate in the context of the conversation. [V]

One or more aspects of this act is/are inappropriate. [V]

The act demonstrates no misunderstandings of the information provided by the customer. [V]

The act is unlikely to be misunderstood by the customer. [V]

I can think of no unintended negative consequences of the act. [V]

[free-form] Here is a better act I thought of: [...] [V]

### b) Customer-directed framing

The note is well-written. [V]

The note meets the structural requirements. [V]

### c) Thoughtful act + customer-directed framing

The note is appropriate for the act. [V][I]

### d) Thoughtful act + process-related framing

The agent knew what it was doing when it came up with the act. [I]

Figure 5. Sample evaluation statements, tagged with the creativity aspects they are meant to evaluate. Answers are either on a five-level Likert scale or free-form.

## Conclusion and Further Research Directions

We have proposed thoughtful surprise generation as a computational creativity problem, and described initial steps toward modules of a financial dialogue challenge that evaluates AI agents’ abilities to generate thoughtful acts. We have

highlighted the surprise-preserving dialogue process that would need to be conducted by agents competing in a variant of the challenge.

Passing even simple versions of the proposed challenge requires complex AI capabilities. However, not even a hypothetical agent that can perform well in the most advanced version of the challenge is necessarily at the level of Emma from our introductory story. One notable reason is that Emma has autobiographical memories, preferences, and feelings, which, in combination with what she believes Claire's memories, preferences, and feelings to be, she uses to come up with the thoughtful act. Herein, we have made the simplifying assumption that thoughtful acts are always receiver-centric, i.e., they are based solely on what the agent infers the customer's preferences to be, not on any preferences or life history of the agent itself. In framing, the agent may talk about the customer's feelings, but not its own. However, in human social relationships (Schwartz 1967), gifts can be both giver-centric and receiver-centric. Even better, they can reflect commonalities of preference and life experience (e.g., "Here's a book by my favorite author, set in a country you enjoyed visiting!"). Developing such subjectivity-endowed agents for banking contexts raises not just practical issues but also ethical questions, which should be explored. An ethics-first approach to AI design should ensure that the customer is not (a) led to believe that they are talking to a human rather than an AI agent, and/or (b) intentionally deceived by the agent in any other way.

Unintended consequences and ambiguity of surprises have also been hinted at, but merit broader treatment. All of these point to future research directions.

### Acknowledgements

We thank the reviewers of the Ninth International Conference on Computational Creativity for their feedback and suggestions.

### References

- Boden, M.A. 1990. *The Creative Mind: Myths and Mechanisms*. Weidenfield and Nicholson, London.
- Bridewell, W., and Bello, P. 2014. Reasoning about Belief Revision to Change Minds: A Challenge for Cognitive Systems. *Advances in Cognitive Systems*, 3:107–122.
- Charnley, J.W., Pease, A., and Colton, S. 2012. On the Notion of Framing in Computational Creativity. In *Proc. of ICC 2012*, 77–81.
- Colton, S., Pease, A., and Charnley, J. 2011. Computational Creativity Theory: The FACE and IDEA Descriptive Models. In *Proc. of ICC 2011*, 90–95.
- Ghallab, M., Nau, D., and Traverso, P. 2004. *Automated Planning: Theory and Practice*. San Francisco: Elsevier.
- Gil, Y. 2017. Thoughtful Artificial Intelligence: Forging a New Partnership for Data Science and Scientific Discovery. *Data Science*, 1(1-2):119–129.
- Grace, K., and Maher, M.L. 2016. Surprise-Triggered Reformulation of Design Goals. In *Proc. of AAAI 2016*, 3726–3732.
- Grice, H. 1967. *Logic and Conversation*. Unpublished Manuscript. University of California, Berkeley.
- Jarrold, W., and Yeh, P.Z. 2016. The Social-Emotional Turing Challenge. *AI Magazine*, 37(1):31–38.
- Jordanous, A. 2012. *Evaluating Computational Creativity: A Standardised Procedure for Evaluating Creative Systems and its Application*. Ph.D. Dissertation, University of Sussex.
- Loughran, R., and O'Neill, M. 2017. Application Domains Considered in Computational Creativity. In *Proc. of ICC 2017*, 197–204.
- Magerko, B., Dohogne, P., Fuller, D. 2011. Shared Mental Models in Improvisational Digital Characters. In *Proc. of ICC 2011*, 33–35.
- O'Neill, B., and Riedl, M.O. 2011. Simulating the Everyday Creativity of Readers. In *Proc. of ICC 2011*, 153–158.
- Pease, A., and Colton, S. 2011. On Impact and Evaluation in Computational Creativity: A Discussion of the Turing Test and an Alternative Proposal. In *Proc. of AISB Symposium on AI and Philosophy*.
- Pickering, T., and Jordanous, A. 2017. Applying Narrative Theory to Aid Unexpectedness in a Story Generation System. In *Proc. of ICC 2017*, 213–220.
- Prévoit, L., Yao, Y., Gingold, A., Bel, B., and Chan, K.Y.J. 2015. Toward a Scary Comparative Corpus: The Werewolf Spoken Corpus. In *Proc. of SEMDIAL 2015*, 204–205.
- Riedl, M.O. 2014. The Lovelace 2.0 Test of Artificial Creativity and Intelligence. In *Proc. of "Beyond the Turing Test" AAAI Workshop*.
- Riedl, M.O. 2016. Computational Narrative Intelligence: A Human-Centered Goal for Artificial Intelligence. In *Proc. of CHI 2016 Workshop on Human-Centered Machine Learning*.
- Schwartz, B. 1967. The Social Psychology of the Gift. *The American Journal of Sociology* 73(1):1–11.
- Turing, A.M. 1950. Computing Machinery and Intelligence. *Mind*, 59:433–460.
- Veale, T. 2015. Game of Tropes: Exploring the Placebo Effect in Computational Creativity. In *Proc. of ICC 2015*, 78–85.
- Ventura, D. 2016. Mere Generation: Essential Barometer or Dated Concept? In *Proc. of ICC 2016*, 17–24.
- WTVR.com. 2016. <http://wtvr.com/2016/09/15/how-a-conversation-with-a-capital-one-rep-changed-a-womans-outlook-on-life/> [Retrieved 05/03/2018]
- Yannakakis, G.N., and Antonios Liapis, A. 2016. Searching for Surprise. In *Proc. of ICC 2016*, 25–32.