Personality and Emotion in Strong-Story Narrative Planning

Alireza Shirvani, Stephen G. Ware, *Fellow, IEEE*, and Lewis J. Baker

***Abstract*—Believable characters are core elements of a coherent story. Qualities that make story characters more believable include goals, beliefs, personality, and emotion. We propose computational models of emotion and personality by adapting the Ortony, Clore, and Collins (OCC) model of emotion and the Five-Factor personality model. Our models are formulated into multiagent strong-story narrative planning with the promise of being highly reusable and domain independent. We evaluate these models using multiple human subject studies. We show that our model's reasoning about character emotions matches the expectations of human readers, and using our emotion model, we can generate a larger set of stories than precedent narrative planners. We also demonstrate that human readers can perceive and recognize the personalities of story characters through their consistent behavior generated by our model. Our final experiment supports that human readers significantly find the behavior generated by our models of emotion and personality more believable than behavior that lacks either or both.**

***Index Terms*—Emotion, interactive narratives, narrative planning, personality, strong-story systems.**

I. Introduction

INTERACTIVE narratives allow users to be part of a fictional world and influence the storyline through their actions and interactions with nonplayer characters (NPCs) [1]. The applications of interactive narratives can be found in education [2], [3], [4], training [5], [6], [7], therapy [8], [9], and entertainment [10], [11], [12].

Whether a story is interactive or not, characters are its key component and crucial to its coherence. If we consider the user as one of the characters, character actions and interactions form a major portion of a story. Therefore, we focus mainly on character behavior with the goal of providing the user with an immersive and effective experience.

Narrative planning is one of the approaches to generating and adapting stories in interactive virtual environments. Narrative planners can explore story spaces too large for human authors to anticipate and can adapt the narrative to each individual user. To make their characters more believable, many narrative planners have considered the belief–desire–intention (BDI) model [13]. They focus on how character must appear to have goals and only act in pursuit of those goals [14], [15], [16], as well as having their individual (possibly wrong) beliefs about the state of the world [17], [18], [19], [20].

However, according to many models of believable characters [21], [22], [23], [24], [25], there are more qualities that make a character believable. In this article, we focus on two of the most notable qualities: personality and emotion. Characters should not only appear to think, but must also show emotions of their own [21], and have their individual personality contribute to the coherency, consistency, and predictability of their reactions and responses [26].

We propose a model of emotion, based on the Ortony, Clore, and Collins (OCC) theory of emotion [27], and personality, based on the Five-Factor model (FFM) [28], for strong-story state-space narrative planning. We previously *separately* introduced and evaluated these two models—our model of emotion [29] and personality [30]. In this article, we present both models together, and in addition to presenting the previous studies and their results, we also present new results from an unpublished study (Experiment 4) that combines both the models into a narrative planner. Combining our models of emotion and personality also allows us to update our personality features (see Section V-B) to leverage our definitions of emotion. We claim that our models enable strong-story narrative planners to generate more stories and more believable behavior, in terms of characters acting more consistently and human-like, than many of their precedent planners [14], [16], [19].

II. Related Work

Two main differences between our system and previous personality and emotion models can be summarized as follows. We focus on leveraging strong story in modeling personality and emotion, and we enable our system to both reactively and proactively reason about those qualities.

Manuscript received 26 September 2021; revised 8 June 2022; accepted 1 October 2022. Date of publication 6 December 2022; date of current version 15 December 2023. This work was supported by National Science Foundation under Grant IIS-1647427. *(Corresponding author: Alireza Shirvani.)*

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by Nonmedical Institutional Review Board (IRB), Office of research integrity, University of Kentucky Application No. 63947. The study meets federal criteria to qualify as an exempt study.

Alireza Shirvani was with the University of Kentucky, Lexington, KY 40506 USA. He is now with Tulane University, New Orleans, LA 70118 USA (e-mail: shirvani@tulane.edu).

Stephen G. Ware is with the University of Kentucky, Lexington, KY 40506 USA (e-mail: sgware@cs.uky.edu).

Lewis J. Baker is with Pymetrics, New York, NY 10001 USA (e-mail: lewisjbaker@gmail.com).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TG.2022.3227220.

Digital Object Identifier 10.1109/TG.2022.3227220

A. Strong Story Versus Strong Autonomy

An experience manager is an intelligent, omniscient, and disembodied agent that drives the narrative forward by intervening in the fictional world through coordination of NPCs and the environment [1]. Based on the NPCs' degree of independence from the experience manager, narrative generation systems fall on a spectrum from strong story to strong autonomy.

In strong-autonomy systems, NPCs decide about their actions independently and with little to no coordination, which results in producing an emergent narrative [31] and limits them to the material offered by a simulation. To find interesting stories, a *story sifter* sifts through raw simulated material to extract narrative artifacts with discernible story structure [32]. These characteristics make strong-autonomy systems suitable for applications that do not necessarily need to guide the user's experience toward a particular conclusion [1], such as simulation games and exploratory learning environments. An example of strong autonomy is *TALE-SPIN* [33], which generates stories where animals take actions to achieve their own goals regardless of author goals.

The Playground [34] is one of the extensions of the Oz project [22] that used a rather strong-autonomy approach to introduce a model of emotion and personality. Their methodology, however, is highly domain specific. Indeed, each of the decisions about how to incorporate personality must be made separately for each character [34]. Similar to emotion and adaptation (EMA) [35], the methodology for maintaining the characters' emotional states is reactive in nature.

Versu [36] is a text-based interactive drama system that allows telling interactive stories using hand-authored episodes. Versu is strong autonomy since each character chooses their next action based on their own individual beliefs and desires, and its centralized experience manager rarely forces characters do anything; instead, it operates at a higher level by providing suggestions [36]. Versu assumes that human authors will reason about emotion and personality when scripting social practices; therefore, it does not rely on the world model to determine how and when emotions prompt actions.

There is also a wide range of systems that focus on external manifestations of emotion or personality, e.g., gestures or postures [37], [38], [39]. Although our models could be extended to include these manifestations, here, we only focus on the effects of personality and emotion on behavior, i.e., choosing between different actions.

In contrast, for strong story, the experience manager is given more control over the world and its NPCs to ensure achieving the author goals. The experience manager explores the space of all possible stories and provides the author with the highest degree of leverage over their narrative structure. This, however, comes at a high computational cost or at least higher than that of strong autonomy. This is particularly important, for instance, for educational and training purposes that have a clear set of pedagogical goals. *AUTHOR* [40] is on the opposite side of the spectrum as TALE-SPIN. In Author, characters may take actions in service of author goals, even if they contradict characters' own goals.

The Virtual Storyteller [41] uses a director agent that steps in and forbids NPC actions if they do not fit into the general plot structure. The Virtual Storyteller uses the OCC emotion model to determine a character's emotional state, map that state to action tendencies, and change a character's priority over their goals [42]. We utilize how actions affect character goals, and in doing so, our models do not rely on action tendencies and updating the importance of character goals.

Riedl and Young [43] enable authors to label operators with recommendations of which personality traits characters should have to perform those actions. These recommendations are not based on a specific personality model and relegate the responsibility to the authors to define them as they please. We believe that one advantage of our model of personality is that it does not ask authors to manually label actions when authoring a new story domain. Instead, similar to Khalpada and Garg [26], we use a set of personality features to decide which plan best fits the character's personality.

Bahamón and Young [44] define a mapping between personality traits and planning operators as a domain independent knowledge base. However, at the time of writing this document, they focus on one of the five factors of personality [28]: agreeableness. As other examples that focus on certain aspects of personality, Paradeda et al. [45] evaluate the effect of the level of *Assertiveness* in virtual agents on the participants' decision making and game experience, and Elgarf and Peters [46] consider *Extroversion* to investigate the process of matching the personality of the user with the virtual character through body language and its impacts on the likability of the character and the information recall of the story.

Our proposed models are not purely strong story, as they ensure that author goals are satisfied and, at the same time, all character actions can be explained in terms of character goals. For brevity, we will refer to them as being strong story since they are on the strong-story side of the spectrum. Our models can use all the leverage of a strong-story system, while still having the ability to reason about emotion and personality like a strong-autonomy system. In other words, they bring the improved believability that was previously mostly found in strong-autonomy systems into strong-story systems.

We build directly on previous strong-story narrative planners that equipped NPCs with goals [47] and beliefs [19], and we improve believability by giving those planners information on personality and emotion to work with. In doing so, we also strengthen the strong story nature of the planner by giving it more stories to explore. The more stories a planner can find, the more leverage it has to tell the story it needs.

B. Reactive Versus Proactive

The *Appraisal Theory* [48] is one of the most widely known and validated models of emotion. EMA [35] is one of the most notable computational models of emotion that adopts the appraisal theory to generate believable agents.

There are two main differences between story generation systems using our model of emotion and EMA, as well as other

Lazarus-based systems [49]. First, we choose to implement emotion types based on the OCC theory of emotion [27] rather than Lazarus's appraisal theory. Second, EMA is mainly reactive, and our system is proactive as well as reactive. Through the process described above, EMA determines the emotional state of an agent based on appraisals and enables agents to *react*, mainly through coping mechanisms. Another example of Lazarus-based systems is an emotion-driven artificial agent architecture that is based on rule-based systems [50]. These systems enable their agents to learn via emotion-based memory management. However, the rules that govern emotion elicitation are specific to the story domain.

With proactive reasoning, a strong-story system can explore the space of all possible stories and *foresee* many, if not all, sequences of events that will trigger different emotions for different characters. Using this information, the system can plan ahead to create specific emotional situations for both the player and the NPCs. That is a type of reasoning that EMA-based story generation systems do not attempt to do. Not only does our model determine what triggers an emotion and how characters react emotionally to events, but it also integrates reasoning about emotions into story generation. This proactive reasoning enables notable opportunities for story generation.

C. Other Notable Differences

As mentioned, our system is strong story rather than strong autonomy and proactive rather than reactive. The following outlines some other notable differences between our system and previous related work.

1) We do not overlook any of the five factors of personality, even though this may come at the cost of oversimplifying the original psychology model.
2) We rely on existing narrative structures to model emotion and personality in order to minimize the author burden and improve the models' reusability for various story domains.
3) Our models account for interactions between different characters and their expectations about each other, which makes it more effective in multiagent simulations.
4) Our main focus is the manifestation of character personality and emotion through their external behavior, rather than natural language dialog or physiology, such as facial expressions and gestures.

III. Narrative Planning

In this section, we will discuss the story domain that we will use in our examples throughout this article. We will also provide the narrative planning definitions used by our model of emotion and personality.

A. Example Story Domain

We use the following example throughout this article. We will refer to this example as *Tom's Tale*. *Tom* is sick and needs medicine. He has two coins and he wants to acquire the medicine while spending the least number of coins. He could either go to a nearby town and spend one coin to buy the medicine from a *Merchant* or he could go to a nearby forest and make it using herbs that grow there. Although he believes that he could do the latter, in reality, there are no herbs in forest that he could use to make the medicine.

Tom also knows that there is a *Bandit* in the forest that could steal all his coins. Tom can buy a sword from the merchant that prevents the bandit from robbing him. Having a sword also gives Tom the option to steal the medicine from the merchant. Both the bandit and the merchant want to have as many coins as they can.

B. Narrative Planning Problem

We build on what Helmert calls a *Multivalued Planning Task* [51]. A virtual world is represented by some number of *variables*, each of which is assigned a *value*. For example, Tom's location is a variable that could be assigned the value $Town$, $Forest$, and so on.[1] An *assignment* of a value u to a variable v is written $v = u$.

Definition 1: A narrative planning problem is defined as $\langle s_0, U, A, C, U_C \rangle$, where s_0 is the initial state, U is the author utility function, A is a set of actions, C is a set of characters, and U_C is a set of character utility functions.

Definition 2: A *proposition* follows the grammar

$$p \rightarrow True | False | v = u | b(c, p) | p \wedge p.$$

In other words, we permit five kinds of propositions: the constants $True$ and $False$, the assignment of a value to a variable, a belief proposition, and a conjunction of such propositions. The beliefs of a character are represented by modal propositions $believes(c, p)$ (or $b(c, p)$ for short), meaning character c believes proposition p, and the following applies to beliefs about conjunctions:

$$b(c, p \wedge q) \leftrightarrow b(c, p) \wedge b(c, q).$$

Belief propositions can be nested, e.g., $b(c_1, b(c_2, p))$ means character c_1 believes that character c_2 believes proposition p.

Definition 3: A *state* is a function that, for any proposition, returns True or False.[2]

Definition 4: The *initial state* is a state that represents the configuration of the world before any planning begins.

Definition 5: A *utility function* is a function that receives a state as input and returns a real number. For instance, the author's utility in Tom's Tale returns 1 if Tom has the medicine or the bandit has Tom's coins, and otherwise returns 0.

Definition 6: A *character* $c \in C$ is defined as a special constant that represents an agent with intentions and beliefs. The intentions of a character are defined in terms of a utility function

[1] Expressions like $at(Tom)$ and $at(Herbs)$ are two different variables (not functions). In our examples, we use notations, such as at, to make these variables more readable, but the planner considers each variable as a unique symbol.

[2] Note this model requires that every proposition, including belief propositions, can always be assigned True or False. This means that agents can have wrong beliefs, but they must commit to their beliefs. See [19] for full details. We have found this model sufficient for the kinds of interactive stories we want to tell without incurring the cost of full modal reasoning about which beliefs are possible and necessary.

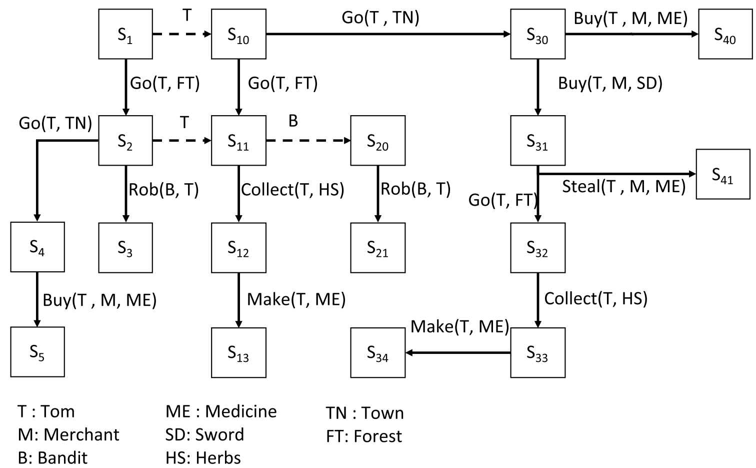

Fig. 1. Part of the state-space for Tom's tale.

$U(c, s)$. Each character intends to increase the value of their utility function. If in a state, Tom has both coins, his utility is 2, if he has both coins and the medicine, his utility is 4, and if he only has the medicine, his utility is 3.

Definition 7: An *action* a describes an event that can occur in the world using the following specifications:

1) $Pre(a)$: a proposition—the precondition of a—that must hold immediately before the action occurs;
2) *Eff(a):* a proposition—the effect of a—that becomes true immediately after a. The action effect is required to be deterministic;
3) $Par(a)$: a set of constants—parameters of a—that are involved in a;
4) $Con(a)$: a set of characters $\in C$—consenting characters of a—where $Con(a) \subseteq Par(a)$, that shows which characters are responsible for taking the action. This set includes the characters who must have a reason to take the action and not necessarily all characters affected by that action. For instance, in action *Rob*, the *robber* is a consenting character, but the *victim* is not;
5) $Obs(a, s)$: a set of characters $\in C$—observing characters of a—that shows the characters who observe the action when it occurs. $Obs(a, s)$ is a function of the current state and the parameters of the action.[3]

Preconditions cannot be contradictory (same for effects). For instance, $Pre(a)$ or *Eff(a)* of action a cannot be

$$at(Bandit) = Forest \wedge at(Bandit) = Town.$$

Definition 8: A *state space* is a graph whose nodes are states and whose directed edges represent actions. An edge $s \xrightarrow{a} s'$ exists if action a's precondition is satisfied in state s and applying a's effects would change the state to s'. Fig. 1 presents an example of a state space.

Definition 9: For some sequence of actions π in a state s, let $\alpha(\pi, s)$ denote the state *after* taking the actions in π. In Fig. 1, $\alpha(\langle Go(Tom, Forest)\rangle, s_1) = s_2$. α is only defined when, for every action $a \in \pi$, the precondition of a is satisfied in the state immediately before taking a, i.e., $s \models Pre(a)$.

Definition 10: For some character c in a state s, let $\beta(c, s)$ denote what c *believes* the state to be when it is actually s. In Fig. 1, $\beta(Tom, s_1) = s_{10}$. In s_1, there are no herbs in the forest, but in s_{10} there are, so Tom wrongly believes there are herbs in the forest.

After taking action a, the beliefs of every character may be updated as follows.[4]

1) $\forall c \in C : c \in Obs(a, s) \Rightarrow \beta(c, \alpha(a, s)) = \alpha(a, \beta(c, s))$.
2) $\forall c \in C : c \notin Obs(a, s) \Rightarrow \beta(c, \alpha(a, s)) = \beta(c, s)$.

It is possible that character c believes the precondition of action a is False and yet observes a. In that case, c's beliefs are first updated to believe the preconditions of a (to believe that action was possible contrary to their wrong beliefs) and then believe its effects. In other words, if $c \in Obs(a, s)$ and $\neg b(c, Pre(a))$, after observing a, $b(c, Eff(a) \wedge (\forall p : (Pre(a) \models p \wedge Eff(a) \not\models p) \Rightarrow p))$, where p is an assignment to a variable or a belief proposition. We refer to this as a *surprise* action, since the character is surprised by an action that they believed was not possible.

For instance, if Tom did not believe that the bandit is in the forest, after Tom observes action *Rob*, he first believes the precondition of *Rob*, i.e., the bandit is in the forest, and then its effect, i.e., bandit has his coins now.

C. Narrative Planning Solution

Various narrative planning frameworks differ in how they define explained actions. We use the following definition for explained actions.

Definition 11: In state s, an action a is *explained for* character $c \in Con(a)$ when there exists a sequence of actions π such that:

1) a is the first action in π;
2) $U(c, \alpha(\pi, \beta(c, s))) > U(c, s)$;

[3]Although methods for automatically determining action observers have been suggested by others [17], [52], [53], this model makes no particular commitment to how $Obs(a, s)$ is chosen.

[4]The details of updating character beliefs for an action whose effect models $b(c, p)$, e.g., to model deception, are not directly relevant to our models of emotion or personality. For full details, refer to [54].

3) every action after a in π is explained;
4) π does not contain a strict subsequence that also meets these four requirements.

In other words, an action makes sense for a character when that character can imagine a plan that: 1) starts with that action; 2) they believe will lead to a higher utility; 3) the plan makes sense for the other consenting characters; and 4) it does not contain unnecessary or redundant actions.

Definition 12: In state s, an action a is *explained* when, for all consenting characters $c \in Con(a)$, a is explained for c in s. In other words, an action is explained when it is explained for all the characters that need a reason to take it. Characters can have different reasons for taking an action. Tom can buy the medicine because he wants it, and the merchant will sell it because she wants coins. The merchant has no reason to give away the medicine, so Tom cannot expect her to.

Definition 13: A sequence of actions π is *explained* when, for all actions $a \in \pi$, a is explained in the state before a occurs. In other words, a sequence is explained when all its actions are explained.

Definition 14: A *solution* to a narrative planning problem is an explained sequence of actions that increases the author's utility and does not contain a strict subsequence that also meets these requirements.

Note that one character can expect another character to act; we call this *anticipation* [19]. A character should not only anticipate actions that help them increase their utility, e.g., expecting the merchant to consent to Buy, but also those that could decrease their utility, e.g., expecting the bandit to consent to Rob.

Definition 15: A sequence of actions π is *expected for* character c in state s when every action in π is explained and c's utility is changed as a result of π, i.e.

$$U(c, \alpha(\pi, \beta(c, s))) > U(c, s) \vee U(c, \alpha(\pi, \beta(c, s))) < U(c, s).$$

This criterion highlights the difference between this definition and the definition of an explained sequence of actions (see Definition 13). An expected sequence of actions for c does not necessarily lead to a higher utility for c. For instance, we cannot say that Rob is explained for Tom because Tom is not a consenting character. However, we say that Rob is expected for Tom because it could decrease his utility and (Tom believes that) it is explained for the action's consenting character, the bandit.

Indeed, characters can expect actions (often the actions of others) to decrease their utility. In keeping with the ideals of a strong-story system, characters can expect many sequences, not just one. Characters do not commit to a single expectation (what a BDI system might call an *intention*), but can expect any sequence that meets these requirements. This enables the planner to choose from a wide variety of believable stories when trying to meet the author's requirements.

In the next section, we will expand these definitions to incorporate emotions and personality. We will show how emotions are triggered as a consequence of actions and how characters distinguish between different plans based on their personality.

TABLE I
EMOTIONS, THEIR APPRAISAL, AND PLANNING TRIGGERS

Emotion	Trigger
Joy	Utility increases
Distress	Utility decreases
Hope	Expects a higher utility
Fear	Expects a lower utility
Satisfaction	Achieves the expected higher utility
FearsConfirmed	Achieves the expected lower utility
Disappointment	No longer expects the higher utility
Relief	No longer expects the lower utility
HappyFor	Own utility does not decrease Other's utility increases
Resentment	Own utility decreases Other's utility increases
Gloating	Own utility does not decrease Other's utility decreases
Pity	Own utility decreases Other's utility decreases

IV. EMOTIONS

The OCC model of emotion defines 22 different emotions [55]. Out of 22, 12 emotions are triggered based on the significance of events to goals, whereas the rest also consider the *standards* and *attitudes* of a character toward events and objects. Only the former set of emotions can be readily adapted into narrative planning without introducing a degree of domain dependence. Therefore, in this article, we will focus on 12 emotions presented in Table I.

A. Positive Emotions

In this section, we provide a formal definition of how each positive emotion is triggered and how the intensity of an emotion is calculated.[5]

1) *Joy Definition:* Joy is triggered for character c at state s after taking/observing action a if $U(c, s) > U(c, s')$, such that $\alpha(a, s') = s$.
Intensity: How much c's utility increases after a or $U(c, s) - U(c, s')$.
Example: Joy is triggered for Tom in state s_5 because his utility increases to 3.
2) *Hope Definition:* Character c feels Hope to achieve utility u as long as there is at least one expected plan π starting from state s, such that $hu = U(c, \alpha(\pi, \beta(c, s)))$ and $hu > U(c, s)$. We refer to hu as *hoped utility.*
Intensity: How much c's utility increases when it reaches hu or $U(c, s) - hu$.
Example: In state s_1, Tom hopes for utility values 4 (by making the medicine himself) or 3 (by buying the medicine).
3) *Satisfaction Definition:* Satisfaction is triggered for character c at state s if $U(c, s) = hu$, such as hu is the corresponding hoped utility. If a character is surprised by an action that increases their utility, they feel Joy but not Satisfaction.
Intensity: The intensity of the corresponding Hope.

[5]Although there is a one-to-one correspondence between emotion triggers in the OCC and planning, we defined emotion intensities here rather intuitively.

Example: Satisfaction triggers for Tom in s_5 for achieving his hoped utility of 3.

4) *Relief Definition:* Relief is triggered for character c at state s if c no longer fears utility fu—Fear is defined later—and $U(c, s) > fu$.
Intensity: The reciprocal of the intensity of the corresponding Fear.
Example: Relief is triggered for Tom at state s_{31} because Tom buys a sword and no longer expects to be robbed.

5) *HappyFor Definition:* Character c feels happy for character c' at state s after action a if for c, $c \in Con(a)$ or $U(c, s) > U(c, s')$, and for c', $U(c', s) > U(c', s')$, such that $\alpha(a, s') = s$.
Intensity: How much c''s utility increases or $U(c', s) - U(c', s')$.
Example: HappyFor is triggered for Tom at state s_5 because after buying the medicine, the merchant's utility is increased by 1.

6) *Gloating Definition:* Character c feels gloating towards character c' at state s after action a if for c, $c \in Con(a)$ or $U(c, s) > U(c, s')$, and for c', $U(c', s) < U(c', s')$, such that $\alpha(a, s') = s$.
Intensity: How much c''s utility decreases or $U(c', s') - U(c', s)$.
Example: Gloating is triggered for the bandit at state s_3 because the bandit's utility increases to 2 and Tom's utility decreases to 0.

B. Negative Emotions

The set of negative emotions are as follows.

1) *Distress Definition:* Distress is triggered for character c at state s after taking/observing action a if $U(c, s) < U(c, s')$, such that $\alpha(a, s') = s$.
Intensity: How much c's utility decreases after a or $U(c, s) - U(c, s')$.
Example, Distress is triggered for Tom in state s_3 because his utility reduces to 0.

2) *Fear Definition:* Character c fears that their utility could decrease to u as long as there is at least one expected plan π starting from state s, such that $fu = U(c, \alpha(\pi, \beta(c, s)))$ and $fu < U(c, s)$. We refer to fu as *feared utility.*
Intensity: How much c's utility decreases when it reaches fu or $fu - U(c, s)$.
Example: Tom fears his utility to decrease to 0 because he expects that the bandit could and would steal his coins.

3) *FearsConfirmed Definition:* FearsConfirmed is triggered for character c at state s if $U(c, s) = fu$, such as fu is the corresponding feared utility. If a character is surprised by an action that decreases their utility, they feel Distress but not FearsConfirmed.
Intensity: The intensity of the corresponding *Fear*.
Example: FearsConfirmed is triggered at s_3 when Tom is robbed as he feared he would be.

4) *Disappointment. Definition:* Disappointment is triggered for character c at state s if c no longer hopes for utility hu and $U(c, s) < hu$.
Intensity: The reciprocal of that of the corresponding Hope.
Example: Disappointment is triggered for Tom in state s_2 because he realizes there are no herbs in the forest.

5) *Resentment Definition:* Character c feels resentment for character c' at state s after action a if for c, $U(c, s) < U(c, s')$ and for c', $c' \in Con(a)$ or $U(c', s) > U(c', s')$, such that $\alpha(a, s') = s$.
Intensity: How much c's utility decreases or $U(c, s) - U(c, s')$.
Example: Resentment is triggered for Tom at state s_3 because the bandit's utility increases to 2 and Tom's utility decreases to 0.

6) *Pity Definition:* Character c feels pity for character c' at state s after action a if $U(c, s) < U(c, s')$ and $U(c', s) < U(c', s')$, such that $\alpha(a, s') = s$.
Intensity: How much c''s utility decreases or $U(c', s) - U(c', s')$.

C. Emotional Planning

Based on the expected emotions, we redefine explained actions as follows.

Definition 16: In state s, an action a is *explained for* character $c \in Con(a)$ when there exists a sequence of actions π such that:

1) a is the first action in π;
2) a positive emotion is triggered for c in $\alpha(\pi, \beta(c, s))$;
3) every action after a in π is explained;
4) π does not contain a strict subsequence that also meets these four requirements.

According to the previous definition (see Definition 11), criteria 2 states that an action is explained for character c if it increases c's utility, thus making them feel Joy or Satisfaction. Our definition of explained actions generalizes criteria 2 to include all other positive emotions. For instance, characters can now consent to actions in pursuit of friendship or rivalry to feel HappyFor or Gloating. Characters can also act in response to their fears (*expected* plans that could decrease their utility) to feel Relief. A simple example is when Tom decides to buy a sword. This is an explained action because, with the sword, he is relieved that the bandit can no longer rob him. His utility not only does not increase, but also decreases for using one of his coins. In short, the proposed model allows characters to act emotionally rather than just rationally.

V. Character Personality

Depending on the story domain, a narrative planner can find multiple valid plans for every character. Existing planning systems return the first valid plan, which is potentially the shortest, unless explicitly asked otherwise.

However, we believe that this choice must depend on characters' personality rather than being nondeterministic. We have already answered why characters choose to act—to feel positive emotions—and now we should address how they act—based on their personality. We select a set of features that describe character plans, independent of their domain-specific details, and then use those features to rank the plans based on different personalities.

A. Five-Factor Model

The FFM is a widely studied taxonomic personality model derived from a factor analysis of a large number of self and peer reports on personality adjectives [28]. The five factors considered by the FFM are Openness to experience, Conscientiousness, Extroversion, Agreeableness, and Neuroticism. An individual could *score* high or low on each of the five factors. We adapt the FFM into narrative planning by representing each factor by two planning features, one for each of its facets [56]. We hand-selected these features after studying multiple personality inventories [28], [56], [57] that include short descriptive phrases or sentences that correspond to each feature.[6]

There are some limitations to our model. First, we strive to achieve high domain independence to make it easy to apply the model to many story domains. In doing so, we intentionally limit ourselves to structures already provided by narrative planners, e.g., a planning step has preconditions, effects, and consenting characters [14], characters have goals, beliefs, and expectations [19]. Social conventions and relationships are example structures that are not already present in narrative planning, and thus, modeling the markers of Openness or Conscientiousness that address those concepts comes at the cost of impairing domain independence.

Moreover, we focus on expressing character personality through external actions, particularly via the choice between different actions, since domain actions are the building blocks of narrative planning. This additionally restricts modeling certain aspects of the FFM that correspond to internal thoughts that are addressed more frequently in other contexts such as theater or novels.

B. Plan Features

Given multiple plans that could achieve their goal, an agent should choose the one that best fits its personality. In order to do so, a plan is described by a set of features that can be automatically calculated across different story domains independent of domain-specific information. For instance, actions *Fight* or *Steal* could represent malice, or *Taking* or *Stashing* items could represent Greed. We update the description of each feature from our previous work [30] to define them in terms of character emotions. These features and their connection to FFM facets are presented in Table II and later fully described in this section.

We must note that features marked as (R) are negatively correlated with their corresponding aspect. For instance, for Conscientiousness in Table II, "# of actions in a plan (R)" means highly conscientious agents try to *minimize* the number of actions in their plans, but low conscientious agents *maximize* it and choose the longest plan. On the contrary, "# of actions with self as the consenting character" is positively correlated to Conscientiousness, so highly conscientious agents maximize the number of actions with themselves as consenting characters (and vice versa).

TABLE II
PROPOSED PERSONALITY PLAN FEATURES

F	Facet	Feat.	Description (High scores)
O	Openness	CPF	The intensity of Satisfaction
	Intellect	IPF	Average intensity of the Fear (R)
C	Orderliness	SEF	# of actions with self as a consenting character
	Industriousness	EPF	# of actions in a plan (R)
E	Enthusiasm	SPF	# of actions with non-self consenting characters
	Assertiveness	APF	# of actions with non-self non-consenting characters
A	Compassion	COF	Average intensity of HappyFor
	Politeness	PPF	Average intensity of Gloating (R)
N	Withdrawal	SRF	The intensity of Relief
	Volatility	NBF	# of times the character changes their mind

1) Creative Plan Feature (CPF): Exploratory creativity is the process of searching an area of the conceptual space governed by certain rules that determine the membership of concepts to the conceptual space, as well as their *value* [58]. In this context, a concept is a plan and the conceptual space is the space of all possible plans. Based on this definition, a creative character is capable of exploring this space to find the most valuable concept (the plan that maximizes their utility—feeling Satisfaction with the highest intensity).[7]

The CPF of plan π for character c is equal to the intensity of the Satisfaction that c expects to be triggered at the end of π. In Tom's Tale, with high Intellect, Tom makes the medicine himself because it maximizes his utility by not losing a coin.

2) Intelligent Plan Feature (IPF): Highly open individuals are intellectual; they want to solve complex problems and their plans rarely fail. The IPF shows how a character choose plans that are more likely to succeed.[8]

The IPF of plan π is calculated as the average intensity of Fear triggered by π. In short, the character takes actions that they do not expect could fail. In Tom's Tale, with high Openness, Tom buys the medicine because this plan is the least likely to fail.

3) Self-Efficacy Feature (SEF): This feature is meant to represent the self-confidence and self-efficacy of conscientious individuals. The plans are preferred that express independence and self-reliance. SEF of plan π for character c is calculated the number of actions a that $c \in Con(a)$. For instance, if Tom's Orderliness was low, he would wait for the merchant to come to him to sell the medicine rather than going to town himself.

4) Efficient Plan Feature (EPF): Conscientious individuals are industrious and focused and, thus, get things done quickly and efficiently. EPF of plan π is equal to the length (number of actions) of π. In Tom's Tale, if Tom is highly Industrious,

[6]Mentioned articles include evaluations of the correlation between these descriptors and the corresponding features.

[7]By saying that creative individuals tend to maximize their satisfaction, we do not mean that unimaginative individuals do not value Satisfaction; they simply are not creative enough to be able to think of plans that maximize their satisfaction.

[8]Here, the likelihood of success refers to the number of expected plans that could cause that character's plan to fail or could decrease their utility. For instance, a plan that requires going to the forest is less likely to succeed than one that requires going to the town because it is possible for Tom to get robbed in the forest.

he prefers to buy the medicine because that plan only has two steps.

5) Social Plan Feature (SPF): Since extroverts prefer to include others into their everyday lives, they tend to prefer actions that involve as many other characters as possible. SPF of plan π for character c is calculated as the number of consenting characters other than c in π. In Tom's Tale, with high Enthusiasm, Tom buys the medicine because it involves the merchant.

6) Assertive Plan Feature (APF): We represent the assertiveness of extroverts in how they include other characters in their plans whether they want it or not. APF of plan π for character c is calculated as the number of nonconsenting characters other than c in π ($c \in Par(a)$ but $c \notin Con(a)$).

An action may affect a nonconsenting character in a positive or negative way. One may choose to give an item to or attack another character where, in both cases, that character's consent is not needed by those actions. In Tom's Tale, an assertive Tom would choose to buy the sword and rob the merchant. We must note that Tom may not choose this plan if his Agreeableness is high.

7) Compassionate Plan Feature (COF): Highly agreeable individuals prefer actions that assist other characters along the way. COF of plan π for character c is calculated as the average intensity of the HappyFor that c expects to be triggered by π. In Tom's Tale, with high Compassion, Tom buys the sword and medicine from the Merchant.

8) Politeness Plan Feature (PPF): Agreeable individuals show their compassion for other people by avoiding to harm them in the process. PPF of plan π for character c is calculated as the average intensity of Gloating that c expects to be triggered by π. In Tom's Tale, a very polite Tom would not rob the merchant.

9) Stress Relief Feature (SRF): Neurotic individuals are prone to anxiety and try to take actions that help to remove their stressors and feel Relief. SRF of plan π for character c is calculated as the intensity of Relief that c expects to be triggered at the end of π. In Tom's Tale, with high Neuroticism, Tom prefers to buy a sword because it eliminates the threat of being robbed by the bandit.

10) Neurotic Behavior Feature (NBF): Highly neurotic individuals can be described as indecisive, self-doubting, or impulsive. One way to express such characteristics is through showing how often a character changes their mind and abandons their current plan. NBF of plan π is calculated as the number of strict subsequences of π that are valid plans.[9] For instance, $\langle Go(Tom, Town)$, $Go(Tom, Forest)$, $Go(Tom, Town)$,$Buy(Tom, Merchant, Medicine)\rangle$ has these two strict subsequences: $\langle Go(Tom, Forest)$, $Go(Tom, Town)$, $Buy(Tom, Merchant, Medicine)\rangle$ and $\langle Go(Tom, Town)$, $Buy(Tom, Merchant, Medicine)\rangle$

C. Preference Modeling

Now that we have defined and described our 12 features, we will show how a character chooses between a set of valid plans at any given state. Algorithm 1 returns the best plan for a character at a state based on their personality. Line 1 shows the inputs of the algorithm, all valid plans for character c at state s, as well as c's personality vector. The personality of the character is specified by five numbers in [0, 1] for each of the five factors (0.5 showing neutrality).

For each plan, we calculate the value of the 12 features in Table II (line 3). We then calculate the preference vector with five values for each of the five factors. Each value is a function of a personality factor and the features corresponding to that factor. The character's preference for a plan is represented by the plan's utility (U_i), which is calculated as the Euclidean norm of the preference vector (line 4).

Algorithm 1: $Preference(\Pi, P_c, c, s)$.

1: Let Π be the set of valid plans for character c at state s, and P_c be personality of character c with five values $p_\alpha, \alpha \in \{O, C, E, A, N\}$.
2: **for** each plan $\pi_i \in \Pi$ **do**
3: Calculate the set of feature values $\{f_{iO_1}, f_{iO_2}, f_{iC_1}, f_{iC_2}, \ldots, f_{iN_1}, f_{iN_2}\}$, representing two features for each factor as in Table II for plan π_i.
4: Let
$F_\alpha = p_\alpha \times \frac{f_{i\alpha_1} + f_{i\alpha_2}}{2}, \alpha \in \{O, C, E, A, N\}$
$U_i = \sqrt{\sum_{\alpha \in \{O,C,E,A,C\}} F_\alpha^2}$
return $\text{argmax}_{\pi_i \in \Pi} U_i$

VI. Evaluation

In this section, we evaluate our models of emotion and personality using multiple human subject studies. We will first evaluate our model of emotion and personality separately in multiple experiments and then investigate their combination in the final experiment. For all experiments, we implemented interactive stories using Twine 2. We then recruited participants from Amazon Mechanical Turk to read the stories and answer a questionnaire about them. We did not target any specific populations of AMT workers, and thus, no demographic data are available (other than all workers being 18 or above years old). We allowed each worker to participate in each experiment only once, but we did not monitor if the same workers participated in multiple experiments.

A. Evaluation of the Emotion Model

To evaluate the model of emotion, we claim that:

1) the set of stories generated by our model is a superset of stories generated by narrative planners that do not reason about emotions;
2) the emotions labeled by our model are similar to the emotions that human readers expect characters to experience;
3) human readers find the character behavior generated by our model more believable than those created by precedent narrative planners that do not reason about emotions.

[9] In order to calculate this feature for a slightly to highly neurotic character—scores of higher than 0.5, we relax the criterion of a valid plan, which constrains it to have no strict subsequences that follow the same criteria.

We support our first claim as follows. For all the stories generated by narrative planners that do not reason about emotions, characters only take actions that contribute to making them feel Joy, i.e., by increasing their utility (see Definition 11). Therefore, our model can generate all stories that are generated by narrative planners without emotions. In addition to those stories, there exists a set of stories in which characters take actions that could make them feel Relief, HappyFor, and Gloating. Since such actions do not necessarily increase the character's utility, this set of stories can be generated by our model but not narrative planners without emotions.

An example story is one where Tom goes to town, buys a sword, goes to the forest (realizes there are no herbs in the forest), goes back to town, and buys the medicine. This story can only be generated by our model since buying a sword makes Tom feel Relief. However, narrative planners without emotions do not generate this story since not only does "buying a sword" not increase Tom's utility, but actually decreases it.

These additional stories are believable because they follow our emotion model. We will empirically evaluate this through our second and third claims using Experiments 1 and 2 in the following sections. More specifically, in Experiment 2, we will show the believability of additional stories that include relief.

1) Experiment 1: Character Emotion Validation: In this section, we evaluate how accurately our model operationalizes six basic OCC emotions: Joy, Distress (Sadness), Hope, Fear, Disappointment, and Relief. We only considered these six emotions to avoid overwhelming the participant with a large number of options for each question. We did not include Satisfaction and FearsConfirmed since, in this example story, they were always triggered respectively when Joy and Distress were triggered.

In this experiment, we used the Tom's Tale story. We first provided a description of the story domain similar to Section III-A. We then presented the story one action at a time. Each action was a translation of the corresponding domain action using simple natural language templates. After specific sets of actions, we asked what the participant thinks Tom may feel at that moment, and participants could choose from one of the six emotions.

1) Results: For Experiment 1, among 70 total participants, only two participants chose not to answer all questions and their responses were removed. There were a total of seven questions that presented six emotions to the participant to choose from. Using Krippendorff's α [59], the inter-rater reliability was $\simeq 0.4$.[10] We then used the binomial exact test [60] to determine the correct answer to each question. For each question, if an answer was chosen significantly more times than we would expect to see by chance (i.e., $p < 0.05$ using the Binomial exact test), we say that the participants significantly agree on an answer and consider that as the correct answer. For five questions, the participants agreed on exactly one option ($p < 0.05$), and for two questions, the participants agreed on two options ($p < 0.05$—for both options). For those two questions, participants agreed that

[10]There are some Amazon Mechanical Turk workers who may be choosing options randomly as quick as possible to earn a larger sum of compensation in a shorter amount of time. Krippendorff's alpha ranges from -1 to 1, and while 0.4 represents some agreement, one reason that this agreement is low is that, in this experiment, we did not use any techniques to filter out those participants.

Fig. 2. Example of the Interface in Experiment 2.

Tom feels Relief and Sadness when he spends a coin to buy a sword (sad for losing a coin and relieved for having a sword), and Tom feels Joy and Relief when he makes it home with the medicine.

To calculate the accuracy of our model, the correct answer to each question was then compared to how our model answers that question. The accuracy of our model was 100% for the six considered emotions in the short story.

2) Experiment 2: Believability and Empathy in an Interactive Story: To show that the characters generated by our model are more believable, we generated a short text-based interactive narrative in which the participant played the role of the main character by choosing between the textual options available to them. The rules of the story are similar to Tom's Tale and the player's goal is to have medicine. The differences between this interactive story and Tom's Tale are as follows.

1) There are herbs in the forest and the player has to collect them first, give them to the merchant to make the medicine, and then buy the medicine for one coin.
2) The player has the option to buy the sword for one coin and, subsequently, sell it for one coin. It is possible for the player to buy the sword, go to the forest, come back to town, buy the medicine, sell the sword, and have one coin that they could give to an NPC.
3) We mention the forest bandit to the player, and the player has the option to first go to town and buy a sword. However, the bandit will never rob the player regardless of the sword.

The story also included two NPCs, *John* and *William*, one of which expressed emotions through text and facial expressions. The participant could view both characters' portraits and thoughts, which may have changed after certain player actions. Fig. 2 presents an example of these two characters. At different steps, the emotional character can express *Happy*, *Sad*, or *Scared* facial expressions and express their thoughts using emotion keywords, e.g., *hope*, *fear*, and so on, e.g., in Fig. 2, John expresses his fear that the bandit could rob him in the forest. For different participants, the emotional character is randomly chosen to be John or William or to be shown on the left or the right. The player has several opportunities to help either character or neither of them, e.g., they could give them a sword or a coin. We hypothesized that the expressions of the emotional character would cause the player to feel empathy and thus, help that character.

The player's goal is to buy the medicine and go back to the cottage. After satisfying this goal, we asked them a series of questions about the NPCs, e.g., whether they found each character to be not at all believable, somewhat believable, or very believable.

2) Results: For Experiment 2, among 70 participants, 15 did not finish the experiment and their incomplete data were discarded. Using the binomial exact test and Bonferroni correction for testing multiple hypotheses [61], the following results were obtained for the rest of the participants.

1) The players chose to buy a sword before going to the forest (34 out of 55—$p < 0.03$). This shows that "buying a sword" was a valid plan for a significant number of players, and therefore, it is also valid if NPCs choose to do it. By generalization, this supports our hypothesis that characters *may* take actions that make them feel Relief even at the cost of their utility.
2) The players chose to help the emotional character[11] (27 out of 55 —$p < 0.01$) by giving them a sword or one of their coins. This supports our model that characters may take actions that make them feel HappyFor even at the cost of their utility. The players also stated that they would have helped both characters if they could (45 out of 55—$p < 0.01$). This shows that characters generated by our model make players feel empathy towards them and significantly more so than the character without emotions. Moreover, these results also support that characters may take actions that make them feel HappyFor even at the cost of their utility.
3) The players agreed that the emotions and reactions of the emotional character were somewhat to very believable (51 out of 55—$p < 0.01$) and more so than the character that expressed no emotions (35 out of 55—$p < 0.03$).[12] These results support our hypothesis that our model of emotion can improve character believability compared to narrative planners that do not reason about emotions.

B. Evaluation of the Model of Personality

For the stories generated by our model of personality, we claim the following.

- *Hypothesis 1:* Human readers can perceive that a character's behavior in a story is demonstrating certain personality traits.
- *Hypothesis 2:* They can also recognize other stories in which the character is exhibiting the same personality traits.

We tested our above hypotheses in Experiment 3.

1) Experiment 3: Character Personality Perception and Recognition: We conducted Experiment 3 to evaluate our model of personality. In the first stage, for each participant, Tom's personality was chosen randomly to reflect high or low scores of a specific factor. More specifically, it was selected from the ten possible options: one where Tom has high Openness, one where he has low Openness, one where he is highly Conscientious, etc. For each option, Tom has either a high score (1) or low score (0) for one factor and average values (0.5) for the other four.

Subjects first read a brief description of the domain.[13] We then prompted the participants that Tom is considering four different plans to achieve his goals and showed four different stories that could unfold based on those plans. The plans were different from the stories that unfolded as a result of executing those plans. For instance, Tom could get arrested at the end of a story, but he would not plan for that to happen. After reading these four possible stories, we narrated which one actually happened, which demonstrates Tom's personality through his choice.

Subjects were then asked to say whether or not these statements applied to Tom using a five-point Likert scale. The following are some examples of the statements and which FFM inventory they were adapted from. We only presented the statements that corresponded to the selected factor out of the ten possible options.

1) Finds creative solutions to problems [56].
2) Gets things done quickly [56].
3) Feels comfortable around people [57].
4) Takes charge [56].
5) Avoids conflict [56].
6) Cannot be bothered with other's needs [56].
7) Is filled with doubts about things [56].

Since we only use existing planning features to simulate a simplified version of the Big Five, we selected the markers that best captured our simulated traits. For instance, we excluded markers such as "Feel comfortable with myself," "Rarely feel depressed," "Keep things tidy," "Laugh a lot," "Avoid philosophical discussions," or "Get deeply immersed in music." Such markers could not be conveyed through external actions or their inclusion came at the cost of increasing author burden.

In the second stage, subjects were shown four new stories and asked which one they thought would happen for Tom. These four stories were all different from the previous four. They included one that reflected a plan with high preference value for Tom, one with low preference value, the first story generated by the *Glaive* narrative planner (which does not reason about personality) [16], and a randomly chosen story that was not a duplicate.

1) Results: We generated 26 stories and collected results for 228 subjects. At least 40 subjects evaluated each factor (at least 20 for the high Openness, at least 20 for low Openness, etc.). All stories were the same for the participants viewing the same condition except for the random story in the second stage.

We claim that human readers can perceive that a character's behavior in a story is demonstrating certain personality traits. To support this claim, in the first stage of the experiment, we defined success as subjects reporting agree or strongly agree if the statement is positively correlated (or disagree or strongly disagree if it was negatively correlated) with Tom's score for the corresponding factor. We used a binomial exact test to detect if we observed more successes than we should expect to see by chance. The p-value and effect size (expressed as relative

[11] They had the option to help either character or help neither. Success is defined as choosing to help the emotional character out of the three total options they had.

[12] The latter refers to when players chose very believable for the emotional character and somewhat believable or not at all believable for the other character, or chose somewhat believable for the emotional character and not at all believable for the other.

[13] This domain was chosen different from Tom's Tale to reflect a larger variety of actions that reflect different personality factors.

TABLE III
RESULTS OF EXPERIMENT 3

	Hypothesis 1		Hypothesis 2	
	p-value	Effect Size	p-value	Effect Size
O	0.072	1.160	**0.026**	1.60
C	**0.016**	1.160	**0.001**	1.73
E	**0.024**	1.167	**0.014**	1.61
A	**0.048**	1.167	≺**0.001**	2.80
N	0.063	1.128	**0.002**	2.04

risk) for each factor are given in Table III. Significant p-values demonstrate that the majority of human readers agreed (or strongly agreed) with the statements that corresponded to the selected personality factor, therefore showing that they can perceive that the character's behavior corresponds to a certain personality factor. For instance, if the character was high conscientious, human readers significantly agreed with the statement that "[the main character] gets things done quickly." We rejected the null hypothesis at the $p < 0.05$ level for three factors (shown in bold) and at the $p < 0.1$ level for the other two.[14]

We also claim that human readers can recognize other stories in which the character is exhibiting the same personality traits. To support this claim, we show that in the second stage of the experiment, subjects chose a story for Tom that best expressed his personality according to our model. We defined success as a participant choosing the best matching story out of the four presented. The p-value and effect size for a binomial exact test for each factor are given in Table III. We rejected the null hypothesis at the $p < 0.05$ level for all factors. In other words, using our personality model to generate stories, human readers can recognize whether a character is acting according to their personality or otherwise, acting *out of character*—based on the reader's own perception of that character's personality.

Though many tests were significant, effect sizes were relatively low.[15] We prefer higher effects sizes as they imply a higher correlation between our personality model and human reader answers. We attribute some of this to the high noise collected from Mechanical Turk data.[16]

C. Evaluation of Emotion and Personality Combined

We claim that incorporating our models of personality and emotions into narrative planning results in generating more believable behaviors. We conducted Experiment 4 to support this claim. We used the Tom's Tale domain in Experiment 4. However, to evaluate personality, we needed to create two different situations (here referred to as acts), so that Tom's decisions in those two acts could be used to portray his personality. We extended the story to require Tom to go home after acquiring the medicine. He could go through the forest, which is the shorter but the riskier path. He also could pay the town guard a coin as a toll so that he would lower the bridge for Tom to go home.

[14] At $p < 0.1$, we say that results are only *marginally* significant. We emphasize significant results by showing them in bold.

[15] Effect sizes for the relative risk are 1.22, 1.86, and 3.00 for small, medium, and large, respectively.

[16] AMT data are notorious for their high amount of noise. The noise is mainly caused by AMT workers who try to complete as many assignments as possible within the shortest amount of time, e.g., by clicking the first available option at all times.

TABLE IV
CLM RESULTS (p-VALUES) IN EXPERIMENT 4

Question	Emotion	Personality	Both
Tom feels like a realistic lifelike character	0.252	0.422	0.532
Tom has a unique personality based on his actions	0.590	0.468	0.236
The story provides good descriptions of Tom's internal thoughts	0.655	$<$**0.001**	0.174
Tom's actions in act 1 were inconsistent to his actions in act 2	$<$**0.001**	0.215	0.235

1) Experiment 4: Evaluating the Space of All Stories: Based on the inclusion of personality or emotion in narrative planning, the space of all possible stories divides into four different sets. In Experiment 4, we generated all the stories of the Tom's Tale domain for all four sets of stories. There were a total of 15 stories comprising of five to eight sentences for the first act (acquiring the medicine) and three to five sentences for the second act (going home).

1) *PN:* Stories that only model personality. For a story to model personality, Tom's actions must be consistent over the two acts based on our personality model. More specifically, if there is at least one factor where Tom's actions reflect a high score for that factor in one act and a low score in the other that story is considered to lack a model of personality.
2) *NE:* Stories that only model emotion. For this set of stories, we added emotion keywords in the natural language templates that described the story. For instance, instead of "Tom plans to buy the medicine," we say "Tom hopes to buy the medicine." We also added extra sentences wherever necessary, e.g., to convey that Tom feels disappointment.
3) *PE:* Stories that model both emotion and personality. In these stories, we both use emotional descriptions and ensure Tom's actions are consistent over the two acts.
4) *NN:* Stories that model neither personality nor emotion.

There is a noticeable difference between stories with emotions and stories with personality. Although emotions affect character behaviors, e.g., taking actions to feel relief, they can also be used as external expressions, e.g., in the context of our experiment, as emotion keywords. This explicitly prompts the participant about the difference between stories with and without emotions. However, we only express personality externally through behavior, actions in narrative planning, and not visually or in the text. Participants would only implicitly perceive a difference between the stories with and without personality, and they would need to do so over the two acts.

After showing a description of the domain (similar to Section III-A), we randomly chose two stories from two different story sets. Participants first viewed the first act of each story followed by its second act and then a series of questions about that story. After participants read both stories, we then asked another set of questions that asked them to choose between

TABLE V
LINEAR REGRESSION RESULTS (p-VALUE) IN EXPERIMENT 4

	Personality	Emotion	Both
Question	p	p	p
Choose the story that Tom's actions were consistent in both acts.	**<0.001**	0.229	0.447
Choose the story that makes Tom feel more human like.	**0.0117**	**<0.001**	0.0942
Choose the story that you found more realistic.	**0.0253**	**0.0252**	0.861
Choose the story that you personally prefer to read.	0.288	**0.005**	0.957

the two stories. The questions are presented in the next section alongside their corresponding results.

1) Results: For Experiment 4, among 390 participants, 22 did not finish the experiment and their incomplete data were discarded. The first question for analysis was whether participants responded as predicted to the *dependent variables*, existence of personality or existence of emotion. To test this, participants responded to prompts related to these variables with a five-point Likert questions from strongly disagree to strongly agree. To better assess the ordinal differences between Likert-type ratings, we used a cumulative link model (CLM). Sometimes known as ordinal regression, CLMs are a special case of logistic regression, which assume that values are ordinal, with the added benefit of testing for interactions between multiple variables. The interaction between variables indicates whether their effect is additive or not. We generally hope to see no interaction between independent variables.

For each question, we additionally conducted pairwise Wilcoxon sum-rank tests to compare responses to each combination of conditions, using the Benjamini–Hochberg method of false discovery rate correction [62].

1) Participants who read stories with emotion gave higher Likert ratings to the prompt "The story provides good descriptions of Tom's internal thoughts" ($z = 5.370$, $p < 0.001$). This was not true of stories with only personality ($z = 0.447$, $p = 0.655$), and there was no interaction effect ($z = -1.359$, $p = 0.174$). Pairwise Wilcoxon tests indicated that all conditions with emotion had significantly higher ratings than conditions without emotion ($p < 0.001$), with no significant differences with or without personality.
2) Similarly, participants who read stories with personality gave lower ratings to the prompt "Tom's actions in act 1 were inconsistent to his actions in act 2." ($z = 5.834$, $p < 0.001$), with no effect of emotion ($z = 1.239$, $p = 0.215$) or interaction ($z = -1.188$, $p = 0.235$). Pairwise Wilcoxon tests indicated that, for this question, all conditions with personality had significantly lower ratings than conditions without personality ($p < 0.001$), with no significant differences with or without emotion.
3) Two additional prompts ("Tom feels like a realistic lifelike character" and "Tom has a unique personality based on his actions") had no significant differences by condition (all conditions and interactions $|z| < 1.185$, $p > 0.234$).

Table IV compares the effect of dependent variables, personality, emotion, and both against having neither feature. The effects for having both personality and emotion were not significant. This means that there is no effect where having both is significantly different than expected from the added effect of either independently. Since these tests compare all conditions, the results also show the significance of difference between having both personality and emotion over having either independently.

1) For "The story provides good descriptions of Tom's internal thoughts," having emotion and personality is significantly preferred over having only personality or having neither.
2) For "Tom's actions in act 1 were inconsistent to his actions in act 2," having emotion and personality is significantly preferred over having only emotion or having neither.

In sum, these results confirm that the personality and emotion conditions successfully influenced participant's interpretation of the character's consistency of behavior and emotional states, respectively. Since there were no interactions, personality and emotion contribute to preference additively and (seemingly) independently.

The second question for analysis was whether users prefer stories that have either personality, emotion, or both, versus stories without those features. To conduct this analysis, participants were simply asked which story they preferred on a variety of dimensions. As participants saw only two stories of the four potential conditions, and those stories were in different orders, we used a logistic regression to predict preference or dislike of a story based on condition (see Table V).

For each question, we calculated odds ratios of each condition relative to the no-personality no-emotion condition (NN) (see Table VI). Odds ratio can be thought of as the odds that an outcome will occur more in the comparison group than in the reference group. For instance, comparing NN to PE with an odds ratio of 1.50 means that the preferred story is 1.5 times more likely to be the PE story than the NN story.

1) Participants significantly preferred stories with personality features in response to the prompt "Choose the story that Tom's actions were consistent in both acts" ($p < 0.001$), with the highest odds ratio for personality and no effect of emotion ($p = 0.229$).
2) Both personality and emotion contributed to preference in response to the prompt "Choose the story that makes Tom feel more human like" ($p < 0.001$)—with a higher odds ratio for emotion—and "Choose the story that you found more realistic." ($p = 0.0117$).

Tables V and VI summarize the results. These results confirm that participants preferred stories with personality and emotion and found Tom more realistic and human like in those stories. Although having both was not significantly different than the

TABLE VI
LINEAR REGRESSION RESULTS (ODD'S RATIO) IN EXPERIMENT 4

	Personality		Emotion		Both	
Question	/174	OR	/191	Odds Ratio	/200	OR
Choose the story that Tom's actions were consistent in both acts.	111	3.716	73	1.304777	129	3.8320
Choose the story that makes Tom feel more human like.	78	1.760	114	3.207792	122	3.388
Choose the story that you found more realistic.	87	1.630	95	1.613782	121	2.497
Choose the story that you personally prefer to read.	78	1.261	103	1.816825	120	2.328

added effect of either independently, results show the highest odds ratio for having both personality and emotion.

VII. CONCLUSION

In this article, we focused on strong-story narrative planning to improve believable behavior generation. We built upon previous narrative planners that enabled their agents to have goals and beliefs and extended them with models of personality and emotion. Our models of personality and emotion are, respectively, inspired by the Big Five and OCC, which are two widely validated models in psychology. We drew from the concepts shared between those models and narrative planning to adapt them into their computational counterparts.

We investigated our proposed models in human subject studies that asked subjects to read or play a(n) (interactive) story and answer a few questions about them. Using our emotion model allows generating more stories than precedent planners, and we showed that those stories are found believable by human readers. Our subjects stated that they preferred to read the stories generated by our emotion model and found those characters more believable and sympathetic. Using our personality model, the subjects perceived that characters acted more consistently based on their personality. We also showed that combining both emotion and personality resulted in more believable stories than using either or none of the models.

There are some limitations to our models. We intentionally limited ourselves to structures that were already present in narrative planning to ensure a high degree of reusability and add the smallest amount of author burden. For personality, we focused on traits that were communicated through external actions and disregarded traits that were domain specific. Further modifications to narrative planning structures can enable adapting more personality traits and provide a deeper computational model of personality. For emotion, we only adapted 12 out of 22 emotion types. Although it is possible to extend our model to include the rest, it is necessary to define a model of social context that reasons about character standards.

Our proposed models are designed to be extensible. We plan to continue our research and expand or exchange certain components of models. We also hope that for other researchers of the community, these models provide foundations to build upon or insights to apply to their own work.

ACKNOWLEDGMENT

The authors would like to thank Rachelyn Farrell for her intellectual contribution to our model of belief.

REFERENCES

[1] M. O. Riedl and V. Bulitko, "Interactive narrative: An intelligent systems approach," *AI Mag.*, vol. 34, no. 1, pp. 67–67, 2013.
[2] C. Dede, "The evolution of constructivist learning environments: Immersion in distributed, virtual worlds," *Educ. Technol.*, vol. 35, no. 5, pp. 46–52, 1995.
[3] M. Roussou, "Learning by doing and learning through play: An exploration of interactivity in virtual environments for children," *Comput. Entertainment*, vol. 2, no. 1, 2004, Art. no. 10.
[4] J. P. Rowe, L. R. Shores, B. W. Mott, and J. C. Lester, "Integrating learning, problem solving, and engagement in narrative-centered learning environments," *Int. J. Artif. Intell. Educ.*, vol. 21, no. 1/2, pp. 115–133, 2011.
[5] J. M. Thomas and R. M. Young, "Annie: Automated generation of adaptive learner guidance for fun serious games," *IEEE Trans. Learn. Technol.*, vol. 3, no. 4, pp. 329–343, Oct.–Dec. 2010.
[6] A. Zook, S. Lee-Urban, M. O. Riedl, H. K. Holden, R. A. Sottilare, and K. W. Brawner, "Automated scenario generation: Toward tailored and optimized military training in virtual environments," in *Proc. Int. Conf. Found. Digit. Games*, 2012, pp. 164–171.
[7] E. T. Garcia, S. G. Ware, and L. J. Baker, "Measuring presence and performance in a virtual reality police use of force training simulation prototype," in *Proc. 32nd Int. Flairs Conf. Florida Artif. Intell. Res. Soc.*, 2019.
[8] R. Aylett, M. Vala, P. Sequeira, and A. Paiva, "FearNot!—An emergent narrative approach to virtual dramas for anti-bullying education," in *Proc. Int. Conf. Virtual Storytelling*, 2007, pp. 202–205.
[9] J. E. Dumas, N. Szilas, U. Richle, and T. Boggini, "Interactive simulations to help teenagers cope when a parent has a traumatic brain injury," *Comput. Entertainment*, vol. 8, no. 2, pp. 1–13, 2010.
[10] M. Zyda and J. Sheehan, *Modeling and Simulation: Linking Entertainment and Defense*. Washington, DC, USA: Nat. Acad. Press, 1997.
[11] K. Levine, "Narrative legos," in *Proc. Game Developers Conf.*, 2014.
[12] A. Shirvani and S. G. Ware, "Camelot: A modular customizable sandbox for visualizing interactive narratives," in *Proc. 12th Intell. Narrative Technol. Workshop 16th AAAI Int. Conf. Artif. Intell. Interact. Digit. Entertainment*, 2020.
[13] A. S. Rao and M. P. Georgeff, "BDI agents: From theory to practice," in *Proc. 1st Int. Conf. Multiagent Syst.*, 1995, pp. 312–319.
[14] M. O. Riedl and R. M. Young, "Narrative planning: Balancing plot and character," *J. Artif. Intell. Res.*, vol. 39, pp. 217–268, 2010.
[15] J. Teutenberg and J. Porteous, "Efficient intent-based narrative generation using multiple planning agents," in *Proc. Int. Conf. Auton. Agents Multiagent Syst.*, 2013, pp. 603–610.
[16] S. G. Ware and R. M. Young, "Glaive: A state-space narrative planner supporting intentionality and conflict," in *Proc. 10th Artif. Intell. Interact. Digit. Entertainment Conf.*, 2014, pp. 80–86.
[17] J. Teutenberg and J. Porteous, "Incorporating global and local knowledge in intentional narrative planning," in *Proc. Int. Conf. Auton. Agents Multiagent Syst.*, 2015, pp. 1539–1546.
[18] M. Eger and C. Martens, "Character beliefs in story generation," in *Proc. 13th Artif. Intell. Interact. Digit. Entertainment Conf.*, 2017, pp. 184–190.

[19] A. Shirvani, S. G. Ware, and R. Farrell, “A possible worlds model of belief for state-space narrative planning,” in *Proc. 13th Artif. Intell. Interact. Digit. Entertainment Conf.*, 2017, pp. 101–107.

[20] A. Shirvani, R. Farrell, and S. G. Ware, “Combining intentionality and belief: Revisiting believable character plans,” in *Proc. 14th Artif. Intell. Interact. Digit. Entertainment Conf.*, 2018, pp. 222–228.

[21] A. B. Loyall, “Believable agents: Building interactive personalities,” Ph.D. dissertation, Dept. Comput. Sci., Carnegie Mellon Univ., Pittsburgh, PA, USA, 1997.

[22] M. Mateas, “An Oz-centric review of interactive drama and believable agents,” in *Artificial Intelligence Today*. New York, NY, USA: Springer, 1999, pp. 297–328.

[23] P. Gomes, A. Paiva, C. Martinho, and A. Jhala, “Metrics for character believability in interactive narrative,” in *Proc. Int. Conf. Interact. Digit. Storytelling*, 2013, pp. 223–228.

[24] M. S. Lee and C. Heeter, “Cognitive intervention and reconciliation: NPC believability in single-player RPGs,” *Int. J. Role-Playing*, vol. 5, pp. 47–65, 2015.

[25] A. Bogdanovych, T. Trescak, and S. Simoff, “What makes virtual agents believable?,” *Connection Sci.*, vol. 28, no. 1, pp. 83–108, 2016.

[26] P. Khalpada and S. Garg, “Balancing consistency and plot structure in computational storytelling,” in *Proc. IEEE 10th Annu. Inf. Technol., Electron. Mobile Commun. Conf.*, 2019, pp. 0550–0555.

[27] A. Ortony, G. L. Clore, and A. Collins, *The Cognitive Structure of Emotions*. Cambridge, U.K.: Cambridge Univ. Press, 1990.

[28] R. R. McCrae, J. F. Gaines, and M. A. Wellington, “The five-factor model in fact and fiction,” in *Handbook of Psychology*, vol. 5, 2nd ed. Hoboken, NJ, USA: Wiley, 2012.

[29] A. Shirvani and S. Ware, “A formalization of emotional planning for strong-story systems,” in *Proc. AAAI Conf. Artif. Intell. Interact. Digit. Entertainment*, 2020, pp. 116–122.

[30] A. Shirvani and S. G. Ware, “A plan-based personality model for story characters,” in *Proc. 15th Artif. Intell. Interact. Digit. Entertainment Conf.*, 2019, pp. 188–194.

[31] R. Aylett, “Narrative in virtual environments-towards emergent narrative,” in *Proc. AAAI Fall Symp. Narrative Intell.*, 1999, pp. 83–86.

[32] J. Ryan, “Curating simulated storyworlds,” Ph.D. dissertation, Dept. Comput. Media, Univ. California Santa Cruz, Santa Cruz, CA, USA, 2018.

[33] J. R. Meehan, “TALE-SPIN, an interactive program that writes stories,” in *Proc. Int. Joint Conf. Artif. Intell.*, 1977, pp. 91–98.

[34] W. S. Reilly, “Believable social and emotional agents,” Ph.D. dissertation, Dept. Comput. Sci., Carnegie-Mellon Univ., Pittsburgh, PA, USA, 1996.

[35] S. C. Marsella and J. Gratch, “EMA: A process model of appraisal dynamics,” *Cogn. Syst. Res.*, vol. 10, no. 1, pp. 70–90, 2009.

[36] R. Evans and E. Short, “Versu—A simulationist storytelling system,” *IEEE Trans. Comput. Intell. AI Games*, vol. 6, no. 2, pp. 113–130, Jun. 2014.

[37] A. Normoyle, F. Liu, M. Kapadia, N. I. Badler, and S. Jörg, “The effect of posture and dynamics on the perception of emotion,” in *Proc. ACM Symp. Appl. Perception*, 2013, pp. 91–98.

[38] C. Ennis, L. Hoyet, A. Egges, and R. McDonnell, “Emotion capture: Emotionally expressive characters for games,” in *Proc. Motion Games*, 2013, pp. 53–60.

[39] M. Saberi, U. Bernardet, and S. DiPaola, “Model of personality-based, nonverbal behavior in affective virtual humanoid character,” in *Proc. ACM Int. Conf. Multimodal Interact.*, 2015, pp. 371–372.

[40] N. Dehn, “Story generation after tale-spin,” in *Proc. Int. Joint Conf. Artif. Intell.*, vol. 81, 1981, pp. 16–18.

[41] M. Theune, S. Rensen, R. op den Akker, D. Heylen, and A. Nijholt, “Emotional characters for automatic plot creation,” in *Proc. Int. Conf. Technol. Interact. Digit. Storytelling Entertainment*, 2004, pp. 95–100.

[42] A. B. Loyall and J. Bates, “Personality-rich believable agents that use language,” in *Proc. 1st Int. Conf. Auton. Agents*, 1997, pp. 106–113.

[43] M. O. Riedl and R. M. Young, “Story planning as exploratory creativity: Techniques for expanding the narrative search space,” *New Gener. Comput.*, vol. 24, no. 3, pp. 303–323, 2006.

[44] J. C. Bahamón and R. M. Young, “An empirical evaluation of a generative method for the expression of personality traits through action choice,” in *Proc. 13th AAAI Int. Conf. Artif. Intell. Interact. Digit. Entertainment*, 2017, pp. 144–150.

[45] R. Paradeda, M. J. Ferreira, R. Oliveira, C. Martinho, and A. Paiva, “The role of assertiveness in a storytelling game with persuasive robotic non-player characters,” in *Proc. Annu. Symp. Comput.-Human Interact. Play*, 2019, pp. 453–465.

[46] M. Elgarf and C. Peters, “Rock your story: Effects of adapting personality behavior through body movement on story recall,” in *Proc. 7th Int. Conf. Human-Agent Interact.*, 2019, pp. 241–243.

[47] M. O. Riedl and R. M. Young, “An intent-driven planner for multi-agent story generation,” in *Proc. 3rd Int. Joint Conf. Auton. Agents Multiagent Syst.*, 2004, pp. 186–193.

[48] R. S. Lazarus and R. S. Lazarus, *Emotion and Adaptation*. Oxford, U.K.: Oxford Univ. Press, 1991.

[49] M. Shvo, J. Buhmann, and M. Kapadia, “An interdependent model of personality, motivation, emotion, and mood for intelligent virtual agents,” in *Proc. 19th ACM Int. Conf. Intell. Virtual Agents*, 2019, pp. 65–72.

[50] V. Lungu, “Artificial emotion simulation model and agent architecture: Extended,” in *Advances in Intelligent Control Systems and Computer Science*. New York, NY, USA: Springer, 2013, pp. 207–221.

[51] M. Helmert, “The fast downward planning system,” *J. Artif. Intell. Res.*, vol. 26, pp. 191–246, 2006.

[52] D. B. Christian and R. M. Young, “Comparing cognitive and computational models of narrative structure,” in *Proc. AAAI Conf. Artif. Intell.*, 2004, pp. 385–390.

[53] H. T. Brinke, J. Linssen, and M. Theune, “Hide and sneak: Story generation with characters that perceive and assume,” in *Proc. AAAI Conf. Artif. Intell. Interact. Digit. Entertainment*, vol. 10, 2014, pp. 174–180.

[54] S. G. Ware and C. Siler, “Sabre: A narrative planner supporting intention and deep theory of mind,” in *Proc. 17th AAAI Int. Conf. Artif. Intell. Interact. Digit. Entertainment*, 2021, pp. 99–106.

[55] B. R. Steunebrink, M. Dastani, and J.-J. C. Meyer, “The OCC model revisited,” in *Proc. 4th Workshop Emotion Comput., Assoc. Adv. Artif. Intell.*, 2009, p. 62.

[56] C. G. DeYoung, L. C. Quilty, and J. B. Peterson, “Between facets and domains: 10 aspects of the big five,” *J. Pers. Social Psychol.*, vol. 93, no. 5, 2007, Art. no. 880.

[57] P. T. Costa and R. R. MacCrae, *Revised NEO Personality Inventory (NEO PI-R) and NEO Five-Factor Inventory (NEO-FFI): Professional Manual*. Pinellas Park, FL, USA: Psychol. Assessment Resour., Inc., 1992.

[58] M. A. Boden, *The Creative Mind: Myths and Mechanisms*. London, U.K.: Psychol. Press, 2004.

[59] K. Krippendorff, *Content Analysis: An Introduction to Its Methodology*. New York, NY, USA: SAGE, 2012.

[60] D. C. Howell, *Statistical Methods for Psychology*. Boston, MA, USA: Cengage Learning, 2012.

[61] S. Holm, “A simple sequentially rejective multiple test procedure,” *Scand. J. Statist.*, vol. 6, pp. 65–70, 1979.

[62] Y. Benjamini and Y. Hochberg, “Controlling the false discovery rate: A practical and powerful approach to multiple testing,” *J. Roy. Statist. Soc.: Ser. B (Methodol.)*, vol. 57, no. 1, pp. 289–300, 1995.