Representing Personality Traits as Conditionals

Richard Evans

Abstract. This paper compares two approaches to representing personality traits in synthetic agents. It proposes a set of goals that any computational implementation of personality should satisfy. It describes the personality trait system used in The Sims 3. Then an alternative system is described, in which traits are represented as conditionals relating world state to emotional state. It is shown that the conditionals model does a better job of satisfying the desiderata.

1 INTRODUCTION

The more highly evolved the species, the more one individual’s behaviour differs from another’s within that species [1, 2]. If we want to build evolved synthetic characters, they too must express individual personality.

Modelling individual personalities enables the following novel sort of player interaction: first the player makes some synthetic characters and chooses their personalities. (Perhaps, for example, he chooses an irascible old man, and his sweet forgiving daughter). Then he drops them into a particular social situation. (Perhaps, for example, he makes them homeless). Then he just sits back and watches the emergent drama unfold; the social situation they are in presents them with problems and their different personalities give them unique solutions to those personalities.

The Sims 3 has a model of individual personality that has allowed many players to experiment with exactly this sort of interaction [3]. One notable example is documented in the very popular Alice & Kev blog. Robin Burkinshaw [4] created two Sims with very different personalities: an irascible old man and his sweet forgiving daughter. He dropped them into a particularly challenging social situation: he gave them no money, no job, and made them homeless. Then he sat back, recording the events that unfolded autonomously:

“I have attempted to tell my experiences with the minimum of embellishment. Everything I describe in here is something that happened in the game. What’s more, a surprising amount of the interesting things in this story were generated by just letting go and watching the Sims’ free will and personality traits take over”.

In Burkinshaw’s emergent unfolding story, Kev (the irritable old homeless man) is always trying to find a place to stay the night. But whenever he finds people who are kind enough to have him, he ends up arguing with them. Eventually, he irritating them so much that they ask him to leave, and he is left sitting out on the streets again. It is a hard life, and he takes it out on his daughter. But she is remarkably sweet and always forgives him.

Part of what makes Burkinshaw’s blog so compelling is the heartfelt description of the homeless situation. But it is the combination of the social situation with the unique personalities that makes it dramatic. As Schopenhauer once wrote:

The revelation of the idea of man is accomplished chiefly by two means: by accurate drawings of significant characters, and by the invention of poignant situations in which they reveal themselves.” [5]

This paper outlines two different computational models of individual personality. It proposes a set of goals for any representation of personality. Then it sketches how The Sims 3 models personality, and evaluate how well it meets the desiderata. Next an alternative model of personality traits is proposed, in which each trait is modelled as a declarative conditional, relating world state to emotional state. It is shown that this alternative model does a better job of meeting the desiderata.

2 DESIDERATA FOR A COMPUTATIONAL MODEL OF PERSONALITY

Before evaluating different models of personality, we need a list of explicit requirements and goals that we can use to help us adjudicate.

The first requirement on any computational model of personality is that a personality be composed of atomic units, which can be reused in a variety of different personalities. Satisfying this compositional requirement is essential if we want to be able to generate a wide variety of personalities cheaply. As Chris Crawford puts it, we should “apply the ideas of vector analysis to the problem and look for a complete set of vectors that span the vector space of the problem” [15].

The second major requirement is that each personality trait has a distinct and obvious effect on autonomous behaviour. If you create a character that is a foul-mouthed extrovert, and leave him to his own devices, he should go out meeting new people and swearing at them. Further, the way he autonomously manifests his personality should be transparent to the player: the player shouldn’t have to take careful notes of every action the character does for many days, compile them in a spreadsheet, in order to notice that this character is 11.3% ruder. Manifestation of personality has to be obvious in individual behaviour without recourse to statistical patterns.

The third major requirement on our model of personality is that it explains how personality connects with emotion. We all know that different people are differently affected by the same external stimuli. Our model of personality must handle this.

The fourth requirement is a practical authoring requirement when building a large multi-agent system. If we are

1 RichardPrideauxEvans@gmail.com Thanks to Michael Mateas for feedback on a previous draft. Thanks also to Matt Brown, Ray Maaza, Peter Ingebretson, Tom Wang, Eric Holmberg-Weidler and Matt Goss for many illuminating discussions on these issues.
going to build a world with a large number of personality-traits, we will have to minimize the amount of authoring needed when adding a new trait. We must minimize the amount of code and data that needs to be touched when adding a new trait. If there are $n$ affordances and $m$ personality traits, we need an authoring approach which requires considerably less content than $n \times m$.

The remaining goals are nice to have, but not as important as the preceding.

The fifth goal is that our computational mechanism for generating personalities is sufficiently compositional that we can generate an indefinite number of personalities. A system that can generate a large, but finitely bounded, set of personalities is not as rich as a recursive generative system.

Personality traits seem to have varying degrees of resolution and specificity. For example: some people are aggressive (tout court), others become aggressive if humiliated.

Our sixth goal is that our model of traits should allow us to specify traits at different levels of specificity.

Some traits are incompatible with others. For example: a character cannot both have a good sense of humour, and also be completely humourless. Now trait incompatible such as these could be hand-authored, or (preferably), the incompatibility between traits could be a derived fact, entailed by the description of the traits themselves. Our seventh goal is that incompatibility between traits should be automatically derivable in the model (rather than having to be painstakingly hand-coded).

One’s past affects one’s personality. Different people narrate early life-events differently, and their interpretation of those events determines their understanding of their current possibilities of action. Our final goal is that the system can express the way personal narratives explain personality.

To summarize, the computational model of personality should satisfy the following goals:

1. Each personality must be decomposable into atomic units which can be reused in other personalities
2. Personality trait must affect autonomous action. When we leave the characters alone, we want them to autonomously behave in-character
3. Each personality trait must affect emotion
4. Requirements for authoring: minimize the number of places we need to touch in order to add a new trait
5. The model should be able to make an indefinite number of personalities
6. It should provide for the idea that some traits are fine-grained refinements of others
7. The model should provide the means for automatically deriving which pairs of traits are incompatible
8. The model should make it possible for personal narratives to explain personality

3 HOW THE SIMS 3 REPRESENTS PERSONALITY

The Sims 3 has a model of individual personality based on simple traits. Each Sim can have up to five traits from a pool of 80. This means there are 80-choose-5 (240 million) distinct personalities.

Personality traits affected autonomous behaviour in three main ways. (1) For each trait there was a unique motive associated with it. For example: a mean-spirited Sim was given an extra desire to undermine the self-worth of other people. This extra motive affected his autonomous behaviour, so he behaved in character. Actions that were specially suited to that personality trait were tagged as satisfying the corresponding motive. (2) When a Sim was interacting with objects, traits affected emotional state via a large number of ad-hoc if-thens scattered throughout the code. Naturally, these were rather difficult to find, and difficult to maintain. (3) When a Sim was interacting with other Sims, traits affected emotional state via a set of if-then rules expressed in a simple declarative language (horn- clauses with a small fixed set of free variables). For example:

- If my interlocutor makes a joke, then find it amusing.
- If my interlocutor makes a joke, but I have no sense of humour, find it boring.
- If my interlocutor makes a joke, but I have no sense of humour, but we are good friends, then find it friendly.

Trait-specific conditionals would override the more general-purpose conditionals, so the person would respond in character.

The implementation of traits in The Sims 3 involved a variety of approaches, (1) and (3) were strongly data-driven approaches, whereas (2) involved a lot of procedural code. One of the motivations for the traits-as- conditionals approach, which this paper proposes in the next section, is precisely to replace this heterogeneous hodgepodge with a uniform representation where all the data is in one place.

Traits in The Sims 3 satisfied some of our core goals. Personality traits affected autonomy via a distinct motive associated with each trait. Further, different personalities did have different emotional responses to the same stimulus: Sims who disliked children would get irritable in the presence of children, while family-oriented Sims would enjoy their company.

The trait system was a major step forward from previous versions of The Sims, and many reviewers noticed how the trait system in The Sims 3 made the characters richer: “The Sims themselves are now powered by much more sophisticated psychological systems than found in earlier games... Traits are designed to reflect how people describe themselves in the real world and are so eerily portrayed in their behavior that The Sims 3 feels like an anthroplogy study with teeth” [19].

But a number of our other goals were not satisfied. To start with, adding a new personality trait meant making a wide number of different types of changes before it was manifest in behaviour: you needed to add an associated motive, define a variety of trait-specific social interactions, define how that personality-trait responded to social interactions initiated by others. If you wanted that personality-trait to have different emotional responses, you had to sprinkle the code with if- statements. Each of these aspects of trait manifestation was expressed in a different representation, so there was a substantial authoring burden when adding a new trait.

Although you could make a large number of personality types in The Sims 3, the number was finitely bounded. There was no concept of describing personality traits in a language with recursive structure, allowing an infinite number of possible traits.

Further, the personality traits were simple atomic objects (elements in an enumeration). This meant there could be no explanation in the model of why one trait was incompatible
with another, so incompatibility between traits had to be authored by hand. This was time-consuming and error-prone.

Because traits were atomic objects, there could be no understanding of why personal events could determine personality. Intuitively, if a character suffered a traumatic early event involving a dog, this would explain a subsequent fear of dogs. But in The Sims 3, the personality trait of fearing-dogs had no constituent structure – it had no understanding that fearing-dogs involved dogs, so there was no way to connect the event (being traumatized by a dog) with the trait (fearing dogs).

4 AN ALTERNATIVE: REPRESENTING PERSONALITY TRAITS AS CONDITIONALS

This paper describes a richer model of personality traits than that used in The Sims 3. Instead of a trait being a simple atomic object, whose effects are scattered through the code and data, now a personality trait is represented by a declarative conditional specifying the condition under which the character has an emotional state. For example, Jealousy could be represented as:

\[
\text{If somebody contradicts something I say } \rightarrow \text{Angry(Contradiction)}
\]

\[
\text{If somebody proclaims that I am a traitor } \rightarrow \text{Shame}
\]

\[
\text{If somebody proclaims that I am a traitor } \rightarrow \text{Shame}
\]

The first specifies the condition under which the argumentative person becomes annoyed. The second describes the conditions under which that anger dissipates: by proving somebody else wrong.

Being Vengeful can also be represented as a pair of conditionals:

\[
\begin{align*}
\text{If } x \text{ harms me } & \rightarrow \text{Vengeance(x)} \\
\text{Vengeance(x) and I take revenge on } x & \rightarrow \text{Vengeance(x) dissipates}
\end{align*}
\]

(Note that this example requires emotional state with constituent structure: the Vengeance emotion is directed towards a specific individual, x. In my implementation, this structure is encoded naturally in Exclusion Logic [10]).

5 AN AGENT ARCHITECTURE WHICH SUPPORTS TRAIT-CONDITIONALS

In The Sims 3, actions are tagged directly with the trait-motives that they satisfy. This involves a tight coupling between the set of actions and the set of traits:

The action promises that if you perform it, it will satisfy the motive in question. A large number of autonomy bugs came from places where this promise is broken: places where an action claimed to satisfy a motive, but in fact did not do so for one reason or another, because of failure or unforeseen conflict. In these cases, the Sim is unmasked: continually repeating the same action sequence over and over again, falsely believing that this time he will get the satisfaction that has so far eluded him. (Another set of problems from the opposite direction stem from the fact that the advertised motives specify what typically should happen when the action is performed. But if this particular instance of the action in this particular situation would also satisfy another motive, this fact will be entirely lost on the Sim. Sims will never, in other words, be opportunistic because satisfied motives are assigned to action-types at design-time, rather than assigned differently to different action-instances at run-time).

To support traits-as-conditionals, we need a different agent architecture – one which involves a looser coupling between actions and traits. In this alternative agent-architecture, actions specify declarative post-conditions. They specify what will be true when the action has been performed – not what goals will be satisfied when the action has been performed. These post-conditions in turn make true the antecedents of the trait-conditionals:

---

\[^3\] Isn’t it redundant to say that the conditional is declarative? Aren’t all conditionals declarative? No – some conditionals are deontic: they specify what should happen under certain circumstances, not what will happen under certain circumstances. I previously considered an alternative model of personality traits based on these deontic conditionals [11]. This is why I stress the declarative aspect of the conditionals used here – to distinguish them from deontic conditionals.
Instead of actions directly tagged with motivating factors, there are two extra levels of indirection: actions are tagged with postconditions, and those postconditions make the antecedents of the conditionals true, which in turn activate the emotional response. This involves a looser coupling between the actions and the emotions:

![Figure 2. Looser Coupling between Actions and Emotions](image)

In this alternative architecture, the trait conditionals are performing double-duty:
- They are used by the planner to decide what action to perform: the agent calculates the emotional state he would be in if he performed the action. The emotional state is intrinsically attractive or unattractive.
- They are used by the simulator to update the emotional state of the agent. When he performs an action, the postconditions are added to the database. This in turn makes the conditional’s antecedent true, which in turn updates the agent’s emotional state.

6 EXAMPLES OF TRAITS REPRESENTED AS CONDITIONALS

The conditional model can naturally express a wide variety of personality characteristics. For example, it can express the so-called Big Five traits (a statistical agglomeration of a very large number of traits). For each of the big five, there is a pair representing positive and negative values of the trait:

Open to Experience:
- If I receive a new experience → Excitement

Closed to Experience:
- If I receive a new experience → Anxiety

Conscientious:
- If I do a job properly → Satisfaction
- If I do a job badly → Shame

Careless:
- If I do a job badly → Satisfaction

Extroverted:
- If I meet a new person → Excitement

Shy:
- If I meet a new person → Anxiety

Agreeable:
- If I am friendly to somebody → Satisfaction

Disagreeable:
- If I am mean to somebody → Satisfaction

Neuroticism comes in a variety of flavours. We give one example - neuroticism with respect to bodily contact:
- If I am touched by somebody → Repulsion

Well-adjusted (the opposite of neurotic):
- Again, we only give one example: being well-adjusted with respect to bodily contact with someone I trust:
- If I am touched by somebody who I trust → Warmth

This model can also accommodate the more specific personality traits of The Sims 3. For example:
- Childish: Performing childish action → Enjoyment
- Commitment issues: being in committed state → Anxiety
- Couch-Potato: Exercise → Anxiety; Eating → Enjoyment
- Coward: Doing brave action → Anxiety
- Dislikes children: Being around children → Irritation
- Excitable: Unremarkable event occurs → Irritation
- Family-Oriented: Giving nurturance/help to children → Enjoyment
- Flirty: Receiving attention from attractive men → Enjoyment

7 TRAITS-AS-CONDITIONALS SATISFY THE DESIDERATA ABOVE

1) Decomposition. This model satisfies the decomposability requirement, just as The Sims 3 does: a personality is a bundle of independent personality traits. But now each trait is represented by a conditional with constituent structure, rather than by an atomic unit.

2) Autonomy. In the traits-as-conditionals model, traits directly affect emotion, and indirectly affect autonomy. The trait conditional specifies an emotional state on the right-hand-side. This emotional state is intrinsically motivating: he either wants to be in it or wants to avoid it. So the conditional indirectly tells him what he should do.
(3) Personality affecting emotion. The conditional does double-duty in this architecture. It is used by the planner to decide what to do, and it is also used by the simulator to update emotional states: when the conditional fires, the emotional state is updated directly.

(4) Minimize the authoring requirements to add a new trait. In The Sims 3, traits were linked directly to action. If there were \( n \) actions and \( m \) traits, there was a sparse matrix with \( n \times m \) entries. In the conditional model proposed here, by contrast, there is a small finite intermediary between the world state and the traits: a list of emotional states. So adding a new trait is significantly less burdensome in this conditional model than in The Sims 3.

(5) Supporting an indefinite number of personalities. One significant advantage of the traits-as-conditional model is that it can express an indefinite number of traits. A trait is just a conditional, expressed as a horn-clause. There are just as many possible personality traits as there are possible horn clauses.

(6) Supporting various levels of granularity. Traits-as-conditional make it very natural to express traits at varying levels of specificity. We can define a generally mean-spirited character as:

If I am mean to somebody \( \rightarrow \) Enjoyment

We can define somebody who is mean-spirited to women by adding an extra conjunct on the left-hand side:

If I am mean to somebody female \( \rightarrow \) Enjoyment

We can keep adding conjuncts on the left-hand side, without end, to make more and more specific conditionals.

(7) Supporting the idea that some traits are incompatible. The traits-as-conditional model also provides an explanation of why certain traits are incompatible. Two traits are incompatible if the left-hand-side of one is entailed by the left-hand-side of another, but the emotional state on the right-hand side of one is different from the emotional state of the other. For example, Good-Sense-of-Humour could be characterized as:

If somebody tells a joke \( \rightarrow \) Amused

No-Sense-of-Humour could be described as:

If somebody tells a joke \( \rightarrow \) Bored

The constituent structure of the conditionals makes it possible for the machine to automatically detect which traits are incompatible, rather than (as was the case in The Sims 3) having to hand-author all incompatibility-pairs by hand.

(8) Supporting the idea that past history can explain personality. Finally, because the traits-as-conditional approach treats a trait as a declarative sentence with structure, it can naturally accommodate the idea that personal narratives explain traits. For example: it is easy to see how, after being traumatized by a dog in infancy, to add a conditional:

If I see a dog \( \rightarrow \) Anxious

If traits are conditionals, particular traumatic or transformative moments could be turned dynamically into traits that have been generated on the fly by the situation:

If I am in a situation which has aspect F, and I am having a traumatic / transformative experience, then add a trait conditional: If the situation satisfies F \( \rightarrow \) Anxiety / Enjoyment

This last suggestion is largely programmatic. I have brushed over the considerable issue of how the agent chooses which aspects of the traumatic situation merited the anxious response. If the agent was traumatised when standing in front of a barking dog on a sunny Tuesday, which trait conditional does he add:

If I see a dog \( \rightarrow \) Anxious

If it’s sunny \( \rightarrow \) Anxious

If it’s a Tuesday \( \rightarrow \) Anxious

Nevertheless, acknowledging that this architecture does not directly answer this question - if we do separately find a good answer to it, then the traits-as-conditional architecture is well placed to support the ability to learn traits on the fly based on past experiences.

8 RELATED WORK

Many games, RPGs in particular [6], have used individual personality traits. But the personality traits that are chosen in these games merely affect the stats of his avatar - not the autonomous behaviour of all the NPCs. What is distinctive about the approaches described in this paper is that personality deeply affects autonomous behaviour. This is what allows autonomous improvisation.

Some previous systems [8] had a model where traits affected autonomy. But in these early systems, each agent had the same set of personality aspects. The only thing that differed was the numeric value of each aspect: “The Universe program uses a trait-based personality model. Each story world character is represented by a person frame which stores information about that character such as the character’s name, stereotypes, traits, interpersonal relationships with other characters, and the character’s history. Traits, such as intelligence, moodiness, and promiscuity whose values range in integral value from 0 to 10, are continuous dimensions and the degree to which a character manifests a trait is stored as an integer value. Traits such as intelligence and moodiness and promiscuity have ranges between 0 and 10. Traits such as guile, self-confidence, and niceness have ranges between -10 and 10 where a negative value indicates that that character has the opposite of the trait”.

In Universe, the personality differences are quantitative differences (the difference between having a 5/10 and a 8/10) as opposed to the qualitative difference you get if a trait is modelled by having an element that would otherwise not be there at all. The Sims 1 & 2 similarly had a personality model based on a
small number of quantitative differences [12], rather than a large pool of qualitatively distinct elements.

The approach described in this paper is clearly related to cognitive appraisal theory [13]. Cognitive appraisal theory is a psychological theory that explains why people have the emotions they do, and how the same stimulus can elicit different emotional responses in different people. The explanation of an emotional response involves two types of judgement: the primary appraisal is the agent’s judgement whether the outcome is in line with her desires and goals. The secondary appraisal is her judgement whether she is able to affect the outcome, or to what extent she is helpless. In the traits-as-conditionals approach described here, the fact that the situation causes the emotion is unexplained: it is just taken as given that this sort of person will respond to this sort of situation with this sort of emotional response. The cognitive appraisal theory is a deeper theory in that it attempts to unpack why this conditional is true by appealing to primary and secondary appraisals. The cognitive appraisal theory is more cognitive than my trait-conditionals in that it attempts to explain the emotion in terms of the agent’s judgment about discrepancies between what is and what should be the case: it is the discrepancy between the agent’s judgment of how the situation has turned out and how it should have turned out which explains the emotional upset.

Cognitive appraisal theory has been used to update the emotional state of the agent after action has been performed. But it has not been used to determine action-selection. In the traits-as-conditionals approach described in this paper, a trait-conditionals performs double-duty. It is both used to update the emotional state based on what did happen, and also used to anticipate the emotional consequences of what might happen in order to decide what the agent should do next.

There are a number of other systems that use personality to influence action-selection [16, 17]. In these systems, personality affects action-selection in the following direct way: the action is tagged directly with the personality-trait or emotional state that it satisfies. For example: the action of eating chocolate is tagged with the choc-aholic personality-type. The model described here, by contrast, is much more truly simulationist in that, instead of specifying the consequences of the type of action directly in terms of emotional state or personality state, we are specifying the consequences of that particular action in terms of world-consequence, and then, as a separate step, we compute what the emotional update of that consequence is in the current context. For example: the consequence of eating chocolate is that chocolate is consumed. The personality-trait of being a choc-aholic means that consuming chocolate is particularly pleasurable. In this particular case, the consequence is the same, but the extra level of indirection gives us the ability to be sensitive to the specificities of the situation. E.g. the consequences of moving a pawn forward in a particular chess situation depend on the precise state of the board.

The traits-as-conditionals approach proposed here is based on the personality model developed by Walter Mischel [7]. Mischel was a situationist and interactionist who developed a powerful critique of the big-five trait model, and eventually produced a constructive alternative based on situation-sensitive conditionals. But one major different is that the conditionals Mischel considered were deontic conditionals, relating world-state to the action the agent should do – rather than conditionals relating world-state to emotional state, as proposed here.

9 IMPLEMENTATION
The traits-as-conditionals approach described here has been implemented in a multi-agent simulation. In one scenario, two agents are playing tic-tac-toe. They both want to win, but one of the agents has a personality trait of being a bad loser: losing is particularly upsetting for him. The other has a trait of being sensitive to the other’s feelings: seeing that the other is feeling upset means that she also feels upset. In this situation, when the sympathetic player is about to win, she will anticipate that her winning will upset the other, and sees that him being upset will also upset herself. So she deliberately avoids winning, and aims for a draw, to spare his feelings.

Initial results suggest that the authoring burden is significantly lighter when specifying traits as conditionals. This is precisely because the conditional does double-duty in determining both emotional effects and action-selection.

10 CONCLUSIONS
This paper has contrasted two ways of implementing autonomous personality traits in synthetic characters: the trait model in The Sims 3, and the traits-as-conditionals approach. This paper proposed a set of requirements and goals that any implementation should satisfy. It has been argued that, although The Sims 3 does a reasonable job of satisfying these goals, the traits-as-conditionals approach does a better job:

<table>
<thead>
<tr>
<th>Requirement/Goal</th>
<th>The Sims 3</th>
<th>Traits-as-Conditionals</th>
</tr>
</thead>
<tbody>
<tr>
<td>Personality decomposable into traits</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Personality affects autonomous action</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Personality affects emotion</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Minimal authoring for adding new trait</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Indefinite number of personalities</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Some traits are</td>
<td>No</td>
<td>Yes</td>
</tr>
</tbody>
</table>
Perhaps the major advantage of the traits-as-conditionals approach is that, because one conditional does double-duty in determining both emotional update and action-selection, the authoring burden is lighter. This consideration becomes increasingly important as we scale up from academic proof-of-concept implementations to industrial-size implementations, with hundreds of personality traits and tens of hundreds of different types of action.

REFERENCES
