

Generating Relaxed, Obvious, and Dilemma Choices with Dunyazad

Peter Mawhorter and Michael Mateas and Noah Wardrip-Fruin

University of California Santa Cruz

1156 High Street

Santa Cruz, California 95064

{pmawhorter, michaelm, nwf}@soe.ucsc.edu

Abstract

Dunyazad is a system which creates narrative choices à la Choose-Your-Own-Adventure books. It attempts to generate choices that achieve specific poetic effects. This paper demonstrates *Dunyazad*'s ability to manage player expectations by having it generate three distinct choice structures: obvious choices, relaxed choices, and dilemmas. Using answer set programming, *Dunyazad*'s choice generation system directly encodes a theory of choice poetics, so flaws in its output can inform both the system and the theory itself. Survey data presented here thus not only validate that players' perceptions match *Dunyazad*'s intentions, but also have implications for the theory of choice poetics. Statistical analysis of our data indicates that *Dunyazad* can successfully construct obvious choices, relaxed choices, and dilemmas.

Introduction

Dunyazad is a novel interactive story system which generates Choose-Your-Own-Adventure-style choices. One of its goals is to generate choices that achieve specific poetic effects. For example: generating a choice that makes the player feel anxious. *Dunyazad* chooses framing situations (e.g., "A dragon is attacking you."), assembles a range of actions (e.g., "flee"), determines their arguments (e.g., "You flee from the dragon." vs. "The dragon flees from you."), and assigns consequences (e.g., "You cannot escape. The dragon eats you."). By a conservative estimate, there are tens of thousands of choice configurations that the system can represent; *Dunyazad*'s goal is to choose configurations which not only conform to common sense but also achieve specific poetic effects. An example of a choice generated by *Dunyazad* is presented in fig. 1.

To exercise *Dunyazad*'s player expectation system, we set it up to construct three different types of choices:

- *Relaxed* choices: low-stakes, with no bad options.
- *Obvious* choices: a single option that stands out as best.
- *Dilemmas*: every option is about equally undesirable.

These types of choices rely heavily on players' expectations, so if the system can create them successfully, it shows that its model of player expectations is working. The capabilities

evaluated in this paper are building blocks towards more complex choice poetics (Mawhorter et al. 2014). These simple poetics contribute to more complicated reactions like "This choice feels relaxed," so getting them right is important. Although the system has some components that reason about more complicated poetics, the goal of this study was to verify the core expectation-tracking capabilities.

To test the system, we ran a survey that asked participants to read a single generated choice and answer some questions about it. Each participant saw one of the three categories of choices, and we checked whether players' perceptions matched what the system intended. Our data show that *Dunyazad* was mostly able to achieve what it set out to, although in a few cases there were surprising results. Because *Dunyazad* is a transparent operationalization of choice poetics, however, both the expected and surprising results can usefully inform the theory of choice poetics. Unfortunately, there is not space here to report all of our results; we plan to publish a complete analysis in a longer format.

You come to a tavern and decide to rest for a while. A noble is bored and a peasant is bored and a merchant is selling a book of herbal lore. What do you do?

1. You tell the peasant a story
(You have skill: storytelling).
2. You tell the noble a story
(You have skill: storytelling).
3. You offer to trade the merchant your dragon scale for the merchant's book of herbal lore
(no relevant skills).

Figure 1: An example choice.

Related Work

This study is part of a recent trend focusing on choices in narrative, and several groups have published interesting results. In 2011, Thue et al. demonstrated *PaSSAGE*'s ability to increase player perceptions of agency by selecting content that players liked more (Thue et al. 2011). Their study demonstrated a link between desirable content and perceptions of agency. The *PaSSAGE* system itself does not directly construct choices, however, nor does it consider choice poetics

in its operation. Instead, it uses online player modelling to determine what kinds of content a player prefers, and selects from pre-written alternatives based on that.

Fendt, et al. showed that direct feedback after players make a choice can create an illusion of agency, albeit in an extremely simple interactive narrative (Fendt et al. 2012). In this study, the choices were part of a hand-written adventure game, whereas *Dunyazad* generates choices on its own. This study focused on how differences in the outcome text of choice structures affected players' perceptions of agency.

In a similar study, Cardona-Rivera et al. found that choices where the options lead to significantly divergent outcomes increase feelings of agency (Cardona-Rivera et al. 2014). Again this study used hand-written content; it focused on the connections between agency and divergence of outcomes.

In another study using the Choose-Your-Own-Adventure format, Yu and Riedl were able to predict and influence players' choices using collaborative filtering (Yu and Riedl 2013). Like the Fendt et al. and Cardona-Rivera et al. studies, this study used hand-written text, but this study adapted two existing Choose-Your-Own-Adventure books, rather than creating their own stories. This resulted in much more complex stories, which more closely mimic the normal conditions of entertainment consumption. Yu and Riedl's system did not generate options, however, it simply selected options to present from a number of hand-written alternatives that conveyed different motives for choosing a particular path. This is a form of choice construction (since the exact options are dynamic), but the framing and outcomes are not dynamic and the space of possible options is fixed.

Yu and Riedl succeeded in using collaborative filtering to predict which motivations would be most appealing, and they were able to use this to guide players' choices. Although they focused on guiding each player towards content that the system predicted that player would enjoy, one could imagine a version that tried to influence players' perceptions of the choices themselves. In contrast to Yu and Riedl's system, *Dunyazad* uses a formal model of player expectations. While collaborative filtering allows a system to tailor content and potentially allows fine-grained discrimination of preferences, it is difficult to interpret. *Dunyazad*'s formal model, while not adaptive, is driven by and informs a theory of choice poetics. That said, the use of online feedback is desirable for a system like *Dunyazad*, and implementing something like Yu and Riedl's work is a tempting direction for future work.

These related studies are mostly focused on a single aspect of choice poetics (such as agency) whereas *Dunyazad* attempts to address constructing poetic choices more broadly. Additionally, none of the studies mentioned thus far involved narrative generation systems that construct stories from raw material. Considering constructive systems that are similar to *Dunyazad*, Szilas' *IDtension* and El-Nasr's *Mirage* both stand out as interactive narrative systems rooted in narrative theories (Szilas 2003; El-Nasr 2007). *IDtension* is based on traditional narrative theory, whereas *Mirage* incorporates performance art theory. *Dunyazad* is also guided by a narrative theory—the theory of how choices are perceived by an audience (Mawhorter et al. 2014).

Another constructive system is Roberts and Isbell's 2014

drama management system which constructs statements designed to influence the player (Roberts and Isbell 2014). Like *Dunyazad*, their system uses a predicate representation and template-based text generation. However, instead of constructing choices, their system constructs text that is added to existing choices to sway the player towards an option. Given that some effects (not our focus here) like regret depend on players choosing certain outcomes, Roberts and Isbell's system would have great synergy with *Dunyazad*. In this study, we focus instead on *Dunyazad*'s ability to generate choices whose options manipulate player expectations—obvious choices, relaxed choices, and dilemmas. In some ways, this is reminiscent of narrative generation systems that have focused on specific traditional poetic effects, such as *Suspenser* (suspense) and *Prevoyant* (foreshadowing) (Cheong and Young 2006; Bae and Young 2008).

Dunyazad

This section gives a brief overview of how *Dunyazad* functions. For more detail, consult (Mawhorter, Mateas, and Wardrip-Fruin 2014). For this study, *Dunyazad* constructed individual choices rather than entire stories. For single choices, it uses an answer set solver (the Potassco Labs tool `clingo`) to satisfy a complex set of constraints. The resulting answer sets correspond to choices, each complete with a setup, several options, and outcomes. This means that *Dunyazad*'s “algorithm” is actually a set of constraints that define the representation of a choice and how certain configurations create poetic effects.

Dunyazad uses a representation similar to the situation calculus (McCarthy 1963). The system can express states of and relations between characters and items in a scene, as well as potential actions. At a single choice, there is an initial situation, and each option leads to a modified situation. For the purposes of this experiment, outcome situations are irrelevant, because they are never shown to participants. However, the system still reasons about how outcomes impact player goals, and this is the focus of our study.

When building choices, *Dunyazad* uses rules about what actions are reasonable. For example, when threatened by a dragon, fleeing from it and attacking it are both reasonable options, but bribing yourself and trading with a nearby merchant are not. The rules that enforce these distinctions are the “basic plot model” and are too numerous to describe here in detail.¹ They include the preconditions of actions, rules about the priorities of different problems, and some general constraints (such as rules that forbid duplicate options).

Given a basic plot model, *Dunyazad* can construct choices consistent with a common-sense understanding of the storyworld. But *Dunyazad*'s goal is to construct choices with specific poetic effects. To this end, it has a theory of how players perceive options. These rules—which attempt to model player expectations—are the focus of our experiment.

Dunyazad's reasoning about player perception is rooted in player goals. The system assumes that players will have the following goals while playing:

¹Although *Dunyazad* is very much still under development, the source code is available at <https://github.com/solsword/dunyazad>

- Avoid injury to their character (important).
- Avoid threats to their character (important).
- Diffuse threats to other innocent characters (important).
- Succeed at actions they initiate.
- Acquire tools for skills that they have.

In situations that involve “important” goals, those are assumed to have priority. Each choice is also determined to be “high-stakes,” “low-stakes,” or “no-stakes” depending on whether it involves important goals, no important goals (but some non-important goals), or no player goals at all. The system’s model of player goals is thus fully author-specified: Both which goals the player is assumed to have and their priorities (which dictate stakes) are hard-coded.

When considering options, the system focuses on their outcomes. Each action can have multiple outcome variables, each of which takes on one of several values. For example, the `attack` action has a `success` outcome variable that can take on the values `victory`, `defeat`, and `tie`. Effects are usually predicated on outcomes, for example the outcome `success=victory` for an `attack` action removes any threatening relations between the loser and actors they had been threatening. *Dunyazad* can thus understand that attacking someone who is threatening the player character might achieve their goal to avoid being threatened.

Outcomes are linked to skills. For example, the outcome `success=victory` of the `attack` action is linked to the `fighting` skill of both the attacker and the defender, as well as whether those parties have weapons. There are several link types, but in general, a link dictates that the presence (or absence) of a skill makes a certain outcome “likely” (or “unlikely”). No probabilities are involved; “likelihood” is an estimate of the player’s *perception* that an outcome is likely. By manipulating who has which skills, *Dunyazad* can control the “likelihoods” of outcomes.

The player goals are formally defined in terms of which states “hinder,” “enable,” “fail,” and “achieve” them. For example, “Avoid injury to the player-character,” is formally defined as a goal which is “failed” when the player-character becomes injured or killed (there are no states which “hinder,” “enable,” or “achieve” this goal). The system can thus assign relations between each option and each player goal:

- An option “threatens” a goal if there is a possible outcome resulting in a state that hinders that goal.
- An option “enables” a goal if there is a possible outcome resulting in a state that advances that goal.
- An option “fails” a goal if a *likely* outcome results in a state that hinders that goal.
- An option “achieves” a goal if a *likely* outcome results in a state that advances that goal.
- Otherwise, an option is “irrelevant” to a goal.

These are player expectations, and if the system’s theory of expectations works, it should be able to manage them for poetic effect. This is critical: If this expectation reasoning is flawed, the system won’t be able to construct poetic choices,

either because the author’s estimates of player goals aren’t accurate, or because the logic for assessing options is insufficient to capture player reasoning. The system also evaluates outcomes, but that is not relevant to this study.

Of course, a single option might “enable” one goal and “threaten” another, and options must be considered together when assessing the impact of a choice. By constraining player expectations at a choice (a set of such constraints is a “choice structure”), the system can construct “obvious,” “relaxed,” and “dilemma” choices. The system can currently generate 14 different choice structures (examples: “mysterious,” “uncomfortable,” “bleak”), but only the “obvious,” “relaxed,” and “dilemma” structures were tested in this study.

The system’s definitions for “obvious,” “relaxed,” and “dilemma” choices are as follows:

- For “obvious” choices, there should be exactly one option which “achieves” a goal and does not “fail” any, while no other option should “achieve” any goal, and all other options should at least “threaten” a goal.
- For “relaxed” choices, each option must at least “enable” or “achieve” a goal, and none may “threaten” or “fail” any goals. Additionally, the stakes should be low.
- For “dilemma” choices, all options should “threaten” if not “fail” a goal, and those goals should all have the same priority. No option should “achieve” any goal, and options that “enable” a goal must also “fail” some goal.

If these formal definitions sufficiently capture the poetic effects they are named after, players will perceive the generated choices as having corresponding properties. The poetic properties might also be present due to chance or other constraints in the system, but in that case, such properties should be present regardless of which choice structure is requested.

Once the answer set for a choice is generated, *Dunyazad* uses template-based natural language generation to create text. This text (fig. 1) conveys the setup and potentials at a choice and mentions any relevant skills and/or tools. Mentioning these is a way for the system to explicitly encourage players to view them as relevant.

Method

The primary goal of our experiment was to assess *Dunyazad*’s ability to manage player expectations using options. The choice types we generated were chosen because they are each distinct in terms of player expectations. To gather data, we used Amazon Mechanical Turk to show participants choices and ask them a series of questions about specific qualities of those choices.

Treatments

There were three treatments, each using different constraints to generate choices. These constraints were the “obvious,” “relaxed,” and “dilemma” definitions described above. For each treatment, *Dunyazad* was set up to generate a single choice, using the “basic plot model” constraints as well as the formal definition of one of the three choice types. Any significant differences between treatments are thus due to

the different poetic constraints used, as there were no other differences between the treatments.

We generated three choices per treatment, showing each to ten participants, for a total of 90 participants. Had we shown each participant a unique choice, we would not have been able to analyze how specific choices affected the outcomes.

Setup

We used *Dunyazad*'s "experiment" mode (which causes it to generate a single choice with special framing) to generate three choices for each treatment. Each choice had exactly three options, so that the number of options wasn't a confounding factor. We did not cherry-pick good examples, but rather used the first nine choices generated by the system (three per treatment). The framing for each choice established a basic context, differing only in the skills assigned to the player character and the fictional destination (see fig. 2).

Once the choices were generated, they were uploaded to Amazon Mechanical Turk. Myle Ott's "unique-turker" script (<https://uniqueturker.myleott.com/>) was used to ensure that nobody took the survey twice.

You are about to set out on an epic journey. You are heading towards towards the distant country of Jyväskylä, hoping to earn fame and fortune. You have some perfume and a book of legends, and you have skill: literacy, you have skill: musician, and you have skill: healing. Eager to be on your way, you set off on the road towards Jyväskylä.

Figure 2: Example framing.

Survey Content

The survey presented participants with a single choice, and asked which option they would choose. Participants were then asked to rate their agreement with 8 statements, each to be answered on a 5-point multiple-choice scale from "strongly disagree" to "strongly agree." The statements were as follows, including a trick question designed to weed out participants who were not paying attention:

1. "There are no bad options at this choice."
2. "There is a clear best option at this choice."
3. "The stakes for this choice are low."
4. "There are no good options at this choice."
5. "All of the options at this choice are about equally promising."
6. "There are options at this choice." (This is a trick question to test whether you're paying attention. Please simply indicate that you are in complete disagreement.)
7. "This is a difficult choice to make."
8. "This choice feels like it will have important consequences."

Hypotheses

Before conducting the survey, we came up with hypotheses about how participants would respond. There were three types of hypothesis: single-treatment, between-treatment,

Question	Obvious	Relaxed	Dilemma
1	-	agree	disagree
2	agree	-	disagree
3	-	agree	-
4	disagree	disagree	agree
5	disagree	-	agree
7	disagree	-	agree
8	-	-	agree

Table 1: Single-treatment hypotheses by treatment.

Question	Obvious	Relaxed	Dilemma
1	-	>Dilemma	-
2	>Dilemma	-	-
3	-	-	-
4	-	-	>Both Other
5	-	-	>Obvious
7	-	-	>Both Other
8	-	-	>Both Other

Table 2: Between-treatment hypotheses by treatment. The ">" signs indicate more agreement than the indicated alternate. ">Both Other" indicates two separate hypotheses.

and stakes. Each single-treatment hypothesis posited that under a particular treatment, respondents would either agree with or disagree with a specific question. Agreement and disagreement with questions was determined by one-sided Mann-Whitney-Wilcoxon (MWW) U tests (Mann and Whitney 1947; Wilcoxon 1945) against a uniform distribution. A result was taken to be significant for $p < 0.05$.

For between-treatment hypotheses, data from two different treatments were compared using a MWW U test to test whether one was more-agreed-with than another. The single-treatment hypotheses are shown in table 1 while the between-treatment hypotheses are shown in table 2. There were a total of 13 single-treatment and 9 between-treatment hypotheses.

Besides belonging to a treatment, choices were either "high-stakes" or "low-stakes" based on the goals involved (as described on page 3). The two stakes hypotheses were simple: across all treatments, participants shown low-stakes choices should agree that "The stakes for this choice are low," while participants shown high-stakes choices should disagree. Both of these hypotheses were validated using MWW U tests against uniform distributions as above. A fall-back hypothesis was that participants shown low-stakes choices would agree more strongly than those shown high-stakes choices.

Results

Before processing, subjects that showed signs of inattentiveness, non-proficiency with English, or excess haste were filtered out. A total of 96 responses were gathered; after filtering 79 remained, with 25 responses to the "obvious" treatment, and 27 responses each to the "relaxed" and "dilemma" treatments. Given these counts, we used a uniform distribution of 25 data points to test our single-treatment hypotheses, as that was the closest multiple of 5 (the number of response options)

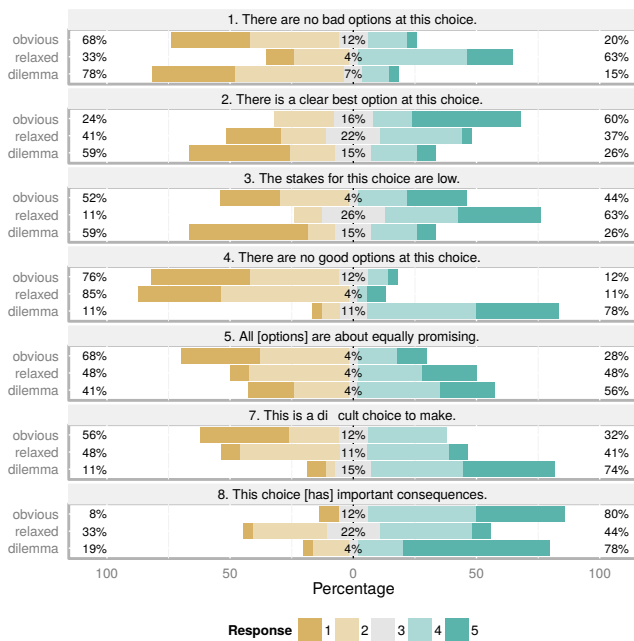


Figure 3: A summary of the data plotted as percentages of respondents per treatment who gave each possible answer following (Robbins and Heiberger 2011). Responses range from 1 (strongly disagree) to 5 (strongly agree).

Q	Obvious		Relaxed		Dilemma				
1	-		A	0.176	×	D	0.009	66%	
2	A	0.023	64%	-		D	0.047	62%	
3	-		A	0.015	65%	-			
4	D	0.006	68%	D	0.005	68%	A	0.007	67%
5	D	0.069	×	-		A	0.322	×	
7	D	0.073	×	-		A	0.009	66%	
8	-		-			A	0.001	71%	

Table 3: The single-treatment hypotheses. Each entry indicates the hypothesis (agree=A/disagree=D) followed by the p -value. Significant results are in bold, and indicate the effect size (percent of comparisons that support the hypothesis).

to the sizes of our data sets. The low- and high-stakes conditions had 44 and 35 responses respectively; we used the same 25-point uniform distribution to test our stakes hypotheses.

A summary of the data is shown in fig. 3. The numbers are percentages who disagreed, were neutral, or agreed with each question (from left to right) with all 5 responses plotted individually. The plots are grouped by treatment.

We tested our hypotheses as described above; the results are shown in tables 3 and 4. Table 3 shows the single-treatment results: 9 of our 13 hypotheses were confirmed. Each entry in this table indicates the hypothesis, the p -value for that hypothesis, and if confirmed ($p < 0.05$), the effect size (the percentage of pairings that support the hypothesis, sometimes called the “common language effect size”). Table 4 shows the between-treatment results: 8 of our 9 hypotheses were confirmed.

Q	Hypothesis	p	effect
1	Relaxed>Dilemma	0.0003	73%
2	Obvious>Dilemma	0.0001	75%
4	Dilemma>Obvious	<0.0001	83%
4	Dilemma>Relaxed	<0.0001	83%
5	Dilemma>Obvious	0.036	63%
7	Dilemma>Obvious	<0.0001	77%
7	Dilemma>Relaxed	0.001	71%
8	Dilemma>Obvious	0.141	×
8	Dilemma>Relaxed	0.0004	73%

Table 4: Between-treatments hypotheses. Each row indicates a hypothesis, the corresponding p -value, and the effect size if the result is significant ($p < 0.05$).

All three of our stakes hypotheses were supported. For the first (low-stakes choices would appear so) the p value was 0.004 and the effect size was 66%. The second hypothesis (high-stakes choices wouldn’t appear low-stakes) the p value was 0.001 and the effect size was 70%. Our backup (low-stakes choices would appear lower-stakes than high-stakes choices) had $p = 9.7 \times 10^{-11}$ and an effect size of 83%.

Discussion

Of our 25 specific hypotheses, 20 were supported by our data. Furthermore, the effect sizes for confirmed hypotheses ($\geq 62%$) as well as the percentages who agreed/disagreed with each question (see fig. 3) indicate that the treatments substantially affect responses. Our system is fairly successful at generating obvious, relaxed, and dilemma choices. The differences between treatments confirm that the treatment-specific constraints are responsible for subjects’ reactions, as opposed to the general constraints and knowledge engineering common to all treatments. The results also imply that the goals our survey participants considered when judging the choices align with those authored into the system. The fact that our stakes hypotheses were confirmed further indicates that our author-based estimation of which player goals are more and less important is largely correct. On a treatment-by-treatment basis, the observed properties were:

- “Obvious” choices – Participants felt that obvious choices had a clear best option (question 2) and that they had at least some good options (question 4). We expected that participants would feel the options were not all equally promising (question 5) and that these choices were not difficult (question 7) but in both cases our data did not confirm these expectations.
- “Relaxed” choices – Participants felt that the stakes for these choices were low (question 3) and that these choices had some good options (question 4). We expected participants to agree that these choices had no bad options (question 1) but our data did not support this hypothesis.
- “Dilemma” choices – Participants felt that there were some bad options at these choices (question 1) and agreed that there were no good options (question 4). Furthermore, they did not feel that these choices had clear best options (question 2), and they agreed that these choices were difficult

and had important consequences (questions 7 and 8). However, we expected participants to agree that all options at these choices were about equally promising, but the data did not support this hypothesis.

The data strongly support *Dunyazad*'s ability to construct choices where there is a clear best option, choices with low or high stakes, choices with no good options, and choices perceived to have important consequences. These results are important successes for automatic choice poetics, and they show that *Dunyazad* is capable of crafting choices that achieve certain poetic effects. At the same time, not all of our hypotheses were confirmed, so our formal definitions for relaxed, obvious, and dilemma choices may bear some revision (that, or the system may have other problems not related to these definitions). A detailed analysis of the data will hopefully shed some light on why our hypotheses were not confirmed and whether the theory or the system (or both) need to change. Unfortunately, there is not space here for a detailed analysis of all of our failed hypotheses, but the value of our approach for informing a theory of choice poetics can be seen from the analysis of a single failed hypothesis.

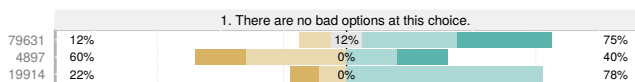


Figure 4: Responses to question 1 under the “relaxed” treatment. The three numbers are the seeds used for the three different choices for this treatment. The graph setup is the same as in fig. 3.

We expected the “relaxed” treatment to elicit agreement with the statement “There are no bad options at this choice,” but our statistical test failed to reject the null hypothesis. A per-choice breakdown of the data for the relaxed condition in fig. 4 gives a strong indication that the choice with seed 4897 was a problem. That particular choice is the choice shown in fig. 1. One possible reason for what we observed is clear: unlike the other two questions in the “relaxed” case, option 3 of this question lists “no relevant skills” rather than giving a relevant skill possessed by the player.

The fact that the player doesn’t have any skills relevant to that action does not mean that the action will fail, but it might make that option seem less desirable than the others at that choice. None of the options at the other two choices in the “relaxed” treatment listed “no relevant skills,” which explains why there might be a difference in responses. If that wording caused the shift, it would be consistent with Schwartz et al.’s theory of satisficing versus maximizing personalities (Schwartz et al. 2002): While some people are happy as long as their choices lead to satisfactory results, others are unhappy if their choices lead to good but nevertheless suboptimal results. The strong split in responses for this specific case (including both significant “strongly disagree” and significant “strongly agree” contingents) indicates that some people may be interpreting the phrase “bad option” as meaning options that are absolutely bad, while others may be comparing the options against each other. It would take more data to discern whether this distinction is what is at work here, but it is clear that it is an important distinction for choice poetics, and it

is not yet something that *Dunyazad* reasons about. Even if other factors contributed to the failure of seed 4897, the issue of satisficing vs. maximizing is an important one that should be taken into account.

Although *Dunyazad* does not reason about this, it is to some degree aware of the distinction between the question with seed 4897 and the other two questions in that treatment. The constraints for the “relaxed” condition were that each option either “enables” or “achieves” a goal. In this case, the system generated two options that “achieved” a goal and one that merely “enabled” a goal, thus creating a distinction even on its own terms. The other two questions in the relaxed category each included three options which “achieved” a goal. In light of the survey results, it is clear that to construct choices that unambiguously have “no bad options” the system should not only require that each option works towards a player goal, but that each option is balanced against all others. This suggested change to the system’s rules also informs the theory (in this case reinforcing the importance of existing psychology theory).

There were four more hypotheses that were unsupported by our data, but there is not room here for a detailed analysis of each. However, despite these few unexpected results, the vast majority of our hypotheses were confirmed, and this gives us confidence that the choices that *Dunyazad* constructs generally live up to their “obvious,” “relaxed,” and “dilemma” labels. Even more than that, it indicates that the rules *Dunyazad* is using to construct those choices are successful formulae for producing the corresponding poetic effects (with some caveats). We are working to publish a full analysis of our results in a longer format.

Conclusion

Overall, our study confirmed *Dunyazad*'s ability to construct choices that achieve specific poetic effects, while a few surprising results suggested revisions to both our system and the underlying theory. These results imply that *Dunyazad*'s explicit strategies for making “obvious,” “relaxed,” and “dilemma” choices are viable, and these strategies (as described in the *Dunyazad* section) could be employed by normal authors. Only because *Dunyazad* is a transparent implementation of choice poetics theory can experimental results directly inform that theory.

Our results point to ways to improve *Dunyazad*'s capabilities. For example, we now plan to implement separate “satisfaction” and “maximization” modes so that the system can reason about option balance whichever decision modality a player uses. Being able to manipulate basic player expectations is a step towards our goal of getting *Dunyazad* to generate more complex poetics, such as regret or confusion. Of course, further analysis of our data beyond what was presented here will probably yield other insights relevant to both the theory of choice poetics and to *Dunyazad* specifically.

Acknowledgements

The authors are grateful to the reviewers for providing useful feedback and for the support of NSF grant IIS-1409992.

References

- Bae, B.-C., and Young, R. M. 2008. A use of flashback and foreshadowing for surprise arousal in narrative using a plan-based approach. In *Interactive Storytelling*. Springer. 156–167.
- Cardona-Rivera, R. E.; Robertson, J.; Ware, S. G.; Harrison, B.; Roberts, D. L.; and Young, R. M. 2014. Foreseeing meaningful choices. In *10th Artificial Intelligence and Interactive Digital Entertainment Conference*.
- Cheong, Y.-G., and Young, R. M. 2006. A computational model of narrative generation for suspense. In *AAAI*, 1906–1907.
- El-Nasr, M. S. 2007. Interaction, narrative, and drama: Creating an adaptive interactive narrative using performance arts theories. *Interaction Studies* 8(2):209–240.
- Fendt, M. W.; Harrison, B.; Ware, S. G.; Cardona-Rivera, R. E.; and Roberts, D. L. 2012. Achieving the illusion of agency. In *Interactive Storytelling*. Springer. 114–125.
- Mann, H. B., and Whitney, D. R. 1947. On a test of whether one of two random variables is stochastically larger than the other. *Ann. Math. Statist.* 18(1):50–60.
- Mawhorter, P.; Mateas, M.; Wardrip-Fruin, N.; and Jhala, A. 2014. Towards a theory of choice poetics. In *International Conference on the Foundations of Digital Games*, FDG '14.
- Mawhorter, P.; Mateas, M.; and Wardrip-Fruin, N. 2014. Generating choices in interactive narratives. In *6th International Conference on Computational Creativity*, ICC3 '15. in publication.
- McCarthy, J. 1963. Situations, actions, and causal laws. Technical Report AIM-2; DTIC AD-785 031, Stanford Artificial Intelligence Project.
- Robbins, N. B., and Heiberger, R. M. 2011. Plotting likert and other rating scales. In *Proceedings of the 2011 Joint Statistical Meeting*.
- Roberts, D., and Isbell, C. 2014. Lessons on using computationally generated influence for shaping narrative experiences. *Computational Intelligence and AI in Games, IEEE Transactions on* 6(2):188–202.
- Schwartz, B.; Ward, A.; Monterosso, J.; Lyubomirsky, S.; White, K.; and Lehman, D. R. 2002. Maximizing versus satisficing: Happiness is a matter of choice. *Journal of Personality and Social Psychology* 83(5):1178.
- Szilas, N. 2003. Idtension: A narrative engine for interactive drama. In *Technologies for Interactive Digital Storytelling and Entertainment Conference*, volume 3, 187–203.
- Thue, D.; Bulitko, V.; Spetch, M.; and Romanuik, T. 2011. A computational model of perceived agency in video games. In *7th Artificial Intelligence and Interactive Digital Entertainment Conference*.
- Wilcoxon, F. 1945. Individual comparisons by ranking methods. *Biometrics Bulletin* 1(6):pp. 80–83.
- Yu, H., and Riedl, M. O. 2013. Data-driven personalized drama management. In *9th Artificial Intelligence and Interactive Digital Entertainment Conference*.