Friend, Collaborator, Student, Manager: How Design of AI-Driven Game Level Editor Affects Creators

Matthew Guzdial, Nicholas Liao, Jonathan Chen, Shao-Yu Chen, Shukan Shah, Vishwa Shah, Joshua Reno, Gillian Smith, and Mark O. Riedl

CHI '19: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems

AIM

Two mixed methods studies:

Guiding the development of the tool.

Examining how the tool impacted the behaviour of practising game designers.

Background

Primarily, the AI is used as a tool, not a partner.

Some studies focused on interfaces and frameworks to allow the co-creation of AI and Human designers.

ex.) Sketch-RNN, Drawing Apprentice, and DuetDraw

Morai Maker

Figure 1: Screenshot of the final level editor.

Study1

They used three AI approaches and focused on a primarily quantitative user study.

Three AI Agents based on:

Markov Chain looked at 2 x 2 grid of level content.

Bays Net looked at 16 x 16 grid of level content.

LSTM looked at the entire level.

Study1

Study method

(1) Which of the two agents was the most fun?

(2) Which of the two agents was the most frustrating?

(3) Which of the two agents was the most challenging?

(4) Which of the two agents most aided your design?

(5) Which of the two agents lead to the most surprising and valuable ideas?

(6) Which of the two agents would you most want to use again?

Study1

results

Table 1: A table comparing the ratio of first rankings for the three comparisons and the p-value of the Wilcoxon rank-sum test, testing if the two ranking distributions differed significantly.

	Most Fun		Most Frustrating		Most Challenging		Most Aided		Most Creative		Reuse	
Pair of AI	ratio	p	ratio	p	ratio	p	ratio	p	ratio	p	ratio	p
Bayes-LSTM	15:13	0.6029	11:17	0.1142	9:19	**0.0083**	17:11	0.1142	19:9	**0.0083**	17:11	0.1142
Bayes-Markov	12:16	0.1469	15:13	0.6029	11:17	0.1142	14:14	1	10:18	**0.0349**	11:17	0.1142
LSTM-Markov	11:17	0.1142	15:13	0.6029	16:12	0.2937	12:16	0.2937	13:15	0.6029	13:15	0.6029

Table 2: A table comparing the Spearman's correlation between the different agent ranking questions. Each cell contains the Spearman's rho and p values for the ranking across the 84 participants.

	Fun		Frustrating		Challenging		Aided		Creative		Reuse	
	rho	p	rho	p	rho	p	rho	p	rho	p	rho	p
Fun	-		-0.74	**<2.2e-16**	-0.10	0.2194	0.79	**<2.2e-16**	0.76	**<2.2e-16**	0.88	**<2.2e-16**
Frustrating	-0.74	**<2.2e-16**	-		0.21	**0.0053**	-0.81	**<2.2e-16**	-0.64	**<2.2e-16**	-0.71	**<2.2e-16**
Challenging	-0.10	0.2194	0.21	**0.0053**	-		-0.21	**0.0053**	-0.10	0.2194	-0.07	0.3572
Aided	0.79	**<2.2e-16**	-0.81	**<2.2e-16**	-0.21	**0.0053**	-		0.74	**<2.2e-16**	0.81	**<2.2e-16**
Creative	0.76	**<2.2e-16**	-0.64	**<2.2e-16**	-0.10	0.2194	0.74	**<2.2e-16**	-		0.83	**<2.2e-16**

Study1

output levels

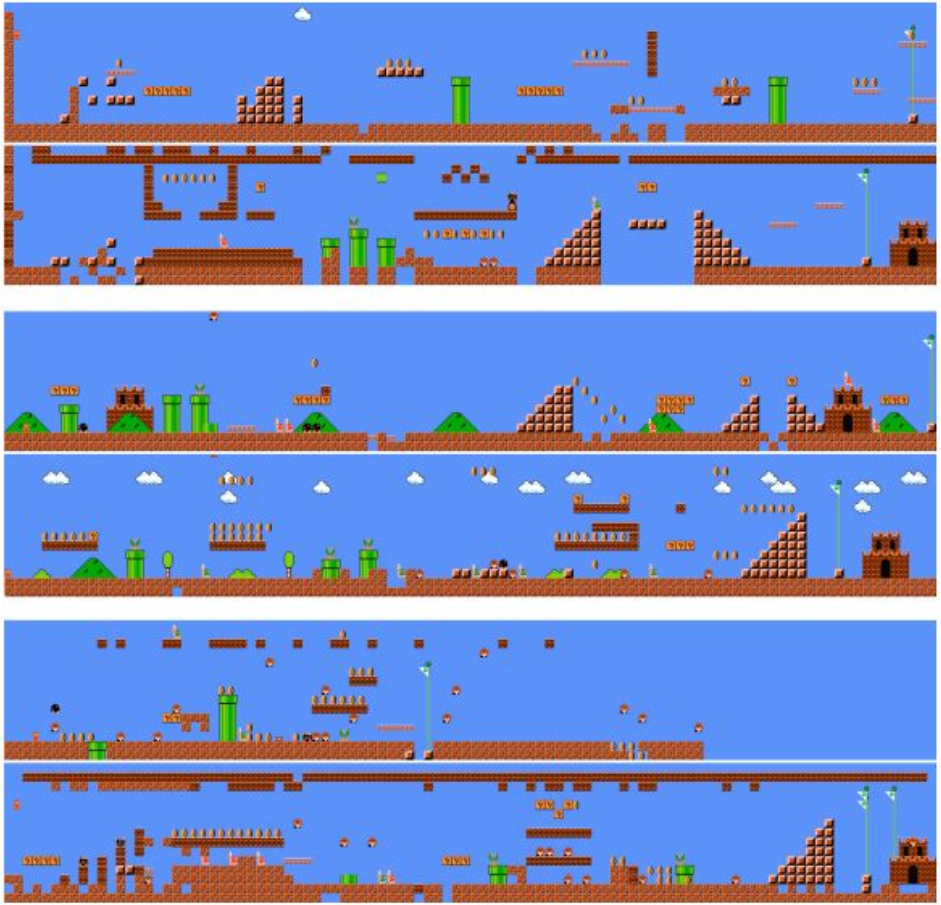

Figure 2: Examples of six final levels from our study, each pair of levels from a specific co-creative agent: Markov Chain (top), Bayes Net (middle), and LSTM (bottom).

Study1

results discussion

No one static agent could meet all of their expectations for an AI partner.

Participants' designs are far from the typical Super Mario Bros. structure.

Participants lacked a clear understanding of their AI partners, but they were willing to invent an explanation for how these partners behaved.

Study2

RQ1: By leveraging active learning to adapt the AI partner to a user, can our tool better serve the needs of level designers?

RQ2: Can Explainable AI allow users to better understand the AI, and therefore to better utilize the tool?

RQ3: Will our overall changes to the tool lead to beneficial experiences for the designers?

Study2

Changes to Morai Maker

The interaction is a semi-Markov Decision Process with concurrent actions.
The final agent trained on the interactions with the 91 participants, using the “Reuse" ranking (1 or -1) as the final reward.
A small negative reward (-0.1) if the human deletes an addition made by the AI partner.
A small positive reward (+0.1) if the human keeps an addition made by the AI partner.
Replace the "Options" button with a "Remove" button.
Removed the "Run" button.

Study2

Method

(1) Did you prefer the agent's behaviour in the first or second session? (a) First (b) Second.
(2) Would you prefer to use this tool with or without the AI partner? (a) With (b) Without (c) No preference.
(3) Did you feel that the agent was collaborating with you? (a) Yes (b) No
(4) Did you feel that the agent was adapting to you? (a) Yes (b) No
(5) If you asked for explanations, did you find that they improved your experience? (a) Yes (b) No

Study2

Quantitative results and Qualitative results

Table 3: Ratio of the answers to the survey's question in the second study.

	First	Second
Most Fun	5	9
Most Frustrating	8	6
Most Aided	5	9
Most Creative	5	9
Preference	6	8
	Yes	No
Collaborating	7	7
Adapting	9	5

Study2

Roles and User Adapting Analysis

Friend: Participants viewed interaction with the AI as primarily a fun activity.

Collaborator: Participants wanted an equal design partner.

Student: Participant seemed to expect the AI to follow their specific design beliefs or instructions.

Manager: Participant seemed to view the AI as giving instructions to them or judging their design

Conclusion

Users varied widely in their expectations and expected role of the AI agent.
Users demonstrated a willingness to adapt their behaviour to the agent.
Overall viewed the tool as having potential value in their design practice.

Thank you for your attention!