Teachable Characters: User Studies, Design Principles, and Learning Performance

Andrea L. Thomaz and Cynthia Breazeal

MIT Media Lab, Cambridge, MA 02139, USA,
alockerd@media.mit.edu

Abstract. Teachable characters can enhance entertainment technology by providing new interactions, becoming more competent at game play, and simply being fun to teach. We argue that is important to understand how human players try to teach virtual agents in order to design agents that learn effectively from this instruction. We present results of an initial user study where people teach a virtual agent a novel task within a reinforcement-based learning framework. Analysis yields lessons of how human players approach the task of teaching a virtual agent: 1) they want to direct the agent's attention; 2) they communicate both instrumental and motivational intentions; 3) they tailor their instruction to their understanding of the agent; and 4) they use negative communication as both feedback and as a suggestion for the next action. Informed by these findings we modify the agent's learning algorithm and show improvements to the learning interaction in a second set of user studies. This work informs the design of real-time learning agents that better match human teaching behavior in order to learn more effectively and be more enjoyable to teach.

1 Introduction

The development of interactive characters that learn from experience continues to be an exciting area of research. In particular, teachable characters, where the human player can shape their behavior, have been successfully incorporated into a number of computer games. In *Black & White*, for instance, human players shape their characters by leading them with different leashes to be "nice" or "naughty" [1]. In *NERO*, virtual armies can be incrementally taught new battle skills from human crafted training exercises [2]. Animal training paradigms are also popular. In *Dogz*, for instance, people can teach their virtual canines new tricks through reward (i.e., treats) or punishment (i.e., spray bottle) [3], whereas Blumberg *et al.* present a learning architecture for a virtual dog that can be taught via clicker training [4]. In sum, teachable characters can enhance game play by either introducing the opportunity for new interactions, becoming more competent at a game, or by simply being fun to teach.

Most of the above examples are Reinforcement Learning (RL) based approaches. RL has certain desirable qualities, in particular the general strategy of exploring and learning from experience. Although the theory of reinforcement learning was originally formulated for systems to learn on-line, independent of human participation, a number of prior works have explored having a human contribute in a supervisory role [4, 8–10]. This models the human input as indistinguishable from any other feedback coming

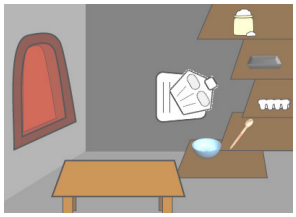


Fig. 1. *Sophie's Kitchen*: oven, table, shelf, and five baking objects.

from the environment. *But is this a good assumption?* We posit that the human is a *special* part of the environment that needs to be more fully understood, supported, and leveraged in the learning process. To do this properly, we must understand the human's contribution: *how* do they teach, *what* are they trying to communicate to the learner?

The contribution of this paper is two-fold. First, we present a systematic study and analysis of human behavior when teaching a virtual robot agent within a reinforcement-based learning framework. Our system, *Sophie's Kitchen*, is a computer game that allows an agent to be trained interactively to bake a cake through sending the agent feedback messages. We use this platform to study people's interactions, and we report four characteristics of how humans approach explicitly teaching a game agent. 1) People want to direct the agent's attention and guide the exploration. 2) Players communicate both instrumental and motivational intents. 3) Transparency behaviors that reveal the internal state of the agent can improve the human's teaching. 4) The human's negative feedback is both feedback and a suggestion to reverse the action if possible.

Second, we have incorporated these findings into specific modifications of the agent's interface and learning algorithm. We had over 200 people play the *Sophie's Kitchen* game in a second set of experiments, showing that our modifications significantly improve the agent's learning performance. This work contributes to the design of real-time learning agents that are better matched to human teaching behavior, to learn more effectively and be more fun to teach.

2 Experimental Platform: Sophie's Kitchen

We implemented a Java-based simulation platform, *Sophie's Kitchen*. The scenario is a kitchen (Fig. 1), where the agent (Sophie) learns to bake a cake. The kitchen creates a sufficiently complex domain with on the order of 10,000 states, and 2-7 actions available from each state. The task is hierarchical with subgoals that can be completed out of order. A successful task completion requires 30-35 steps. We use Q-learning [6] to investigate how people interactively teach an agent. Importantly, we believe these lessons and modifications can be applied to the general class of reinforcement-based learning approaches, and not just Q-learning in particular.

The kitchen has `Flour`, `Eggs`, and a `Spoon` each with a single object state. The `Bowl` has five states: `empty`, `flour`, `eggs`, `unstirred`, `stirred`, and the `Tray` has three states: `empty`, `batter`, `baked`. There are four locations: `Shelf`, `Table`, `Oven`, `Agent`. The locations are arranged in a ring, thus the agent can `GO left` or `right`.

She can PICK-UP any object in her current location; she can PUT-DOWN any object in her possession; and she can USE any object in her possession on any object in her current location. Each action advances the world state (e.g., executing PICK-UP <Flour> changes the state such that the Flour is in location Agent). The agent can hold one object at a time. USEing an ingredient on the Bowl puts that ingredient in it; using the Spoon on the unstirred Bowl transitions its state to stirred, and so on.

In the initial state, the agent faces all the objects on the Shelf. A successful task completion includes putting flour and eggs in the bowl, stirring the ingredients using the spoon, putting the batter into the tray, and putting the tray in the oven. The goal state has a positive reward ($r = 1$), and some end states are so-called *disaster* states since they are unrecoverable (e.g., putting the eggs in the oven). These result in a negative reward ($r = -1$), the termination of the current trial, and a transition to the initial state.

A central feature of *Sophie's Kitchen* is the interactive reward interface. Using the mouse, a human trainer can—at any point—award a scalar reward signal, $r \in [-1, 1]$. The user receives visual feedback enabling them to tune the reward before sending it, and choosing and sending the reward does not halt the progress of the agent, which runs asynchronously to the reward interface. After the initial experiment, additional elements (guidance, motivation) were added to the interaction, these are detailed in Section 4.

3 Experiments

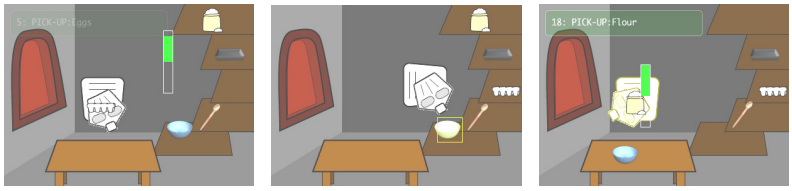
We conducted two experiments with *Sophie's Kitchen* where participants played a computer game. Their goal was to get the virtual robot, Sophie, to learn how to bake a cake. Participants were told they could not tell Sophie what actions to do, nor could they do any actions directly. They were only able to send messages with the mouse to help Sophie learn the task. The experiments differed in messages available and in the behavior of the agent. The experiments are introduced briefly and detailed in Section 4.

3.1 Experiment 1: How do humans teach?

The purpose of the first experiment was to understand, when given a single reward channel, how do people use it to teach the agent? We solicited participation from the campus community and obtained 18 volunteers. In this experiment, people have direct control of the reward signal that the agent uses in the Q-Learning algorithm. The system maintains an activity log of each of the following: state transitions, actions, human rewards, reward aboutness (described in section 4.1), disasters, and goals. Players were only able to give Sophie feedback messages with the mouse and were given the following explanation of the messages: Drag the mouse up for a green box, a positive message; and down for red/negative (Figure 2(a) shows a positive feedback message). Lift the mouse button to send the message, and Sohpie sees the message color and size.

3.2 Experiment 2: Can we improve this interaction?

After the first experiment, we made four algorithm and interface modifications (Guidance, Gaze, Motivation, and Undo), making the exploration process more accessible



(a) Feedback. (b) Guidance. (c) Motivation.

Fig. 2. The embellished communication channel. 2(a): feedback is given by left-clicking and dragging the mouse up to make a green box (positive) and down for red (negative). 2(b): guidance is given by right-clicking on an object, selecting it with the yellow square. 2(c): a motivation message is given by doing a feedback message on top of Sophie.

and guidable to the human partner, as well as making the behavior of the character more expressive, natural and understandable (transparent) to the human. To evaluate the benefits of these modifications, we deployed *Sophie's Kitchen* on the World Wide Web, and collected data from over 200 players.

4 Results

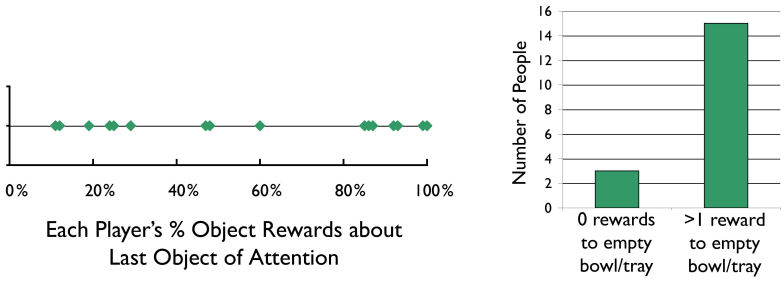
We present four sets of results. In each case a finding in the first experiment led to one of the modifications mentioned above. This modification is explained in detail, and then the second experiment sheds light on its effect on the learning interaction.

4.1 Teaching is more than Simple Feedback

One of the major findings of the first experiment was that, perhaps not surprisingly, people have multiple communication intentions beyond the simple feedback that the Q-Learning algorithm expects. In particular, people want to guide the agent and they want a generic encouragement/discouragement channel of communication.

Guidance: Finding from Experiment 1: In the first experiment the interface let the user assign their feedback to a particular object (object specific rewards). The hypothesis was that people would prefer to indicate what their feedback was 'about'. Object specific rewards are used only to learn about the human trainer's communicative intent; the learning algorithm treats all rewards *equally* and in the traditional sense of pertaining to the whole state. Even though the instructions clearly stated that all communication and rewards were *feedback* messages, we saw that many people assumed that object specific rewards were future-directed messages or guidance for the agent. This is evident from both interviews with participants and from the correlation of object/action pertinence and rewards given.

If people were using the object rewards in a traditional RL sense, they should be feedback and pertain to the agent's last action. In Figure 3(a), there is a mark for each



(a) (b)

Fig. 3. Fig. 3(a) shows each player's percentage of object rewards that were about the most recent object used; many were rarely feedback oriented. Fig 3(b), shows the number of people that gave (0 vs. 1 or more) rewards to the bowl or tray empty on the shelf, which is assumed to be guidance.

player indicating their percentage of object rewards that were about the last object used. A substantial number of people had object rewards that were rarely correlated to the last object of attention, thus not feedback oriented.

We suspect these people's rewards actually pertain to the future, indicating what they want (or do not want) the agent to use next. To test how many people were using the object rewards as guidance, we consider an example case. When the agent is facing the shelf, a guidance reward could be administered (about what object to pick up). A reward given to either the empty bowl or empty tray on the shelf should *only* be a guidance reward since these are not be part of any desired sequence in the cake baking task. In Fig.3(b), we see that 15 of the 18 people had a non-zero number of rewards to the empty bowl or tray on the shelf. Thus we conclude that many participants tried using the reward channel to guide the agent's behavior to particular objects, giving rewards for actions the agent was *about to do* in addition to traditional feedback. While delayed rewards have been discussed in the RL literature [7], these *anticipatory* rewards observed from everyday human trainers will require new tools and attention in learning algorithms in order for the agent to use them as the human partner intended.

Modifications to the Learning Agent to add Guidance: This behavior suggests that people want to speak to the action selection of the algorithm to influence the exploration. To accomplish this, we added a guidance communication channel. Clicking the right mouse button draws a yellow square. When the yellow square is administered on an object, this is a guidance message to the agent, the content of which is the object. Figure 2(b) shows the player guiding Sophie to the bowl. Note, the left mouse button allows the player to give feedback in the same way as the first experiment.

An RL algorithm can be described as continually looping through the following: `select-action`, `take-action`, `sense-reward`, `update-values`. Our modified Q-Learning algorithm adds a phase where the agent registers guidance communication to bias action selection. Thus, the process becomes: `sense-guidance`, `select-action`, `take-action`, `sense-reward`, `update-values`. The agent waits for guidance mes-

Table 1. 1-tailed t-test, effects of guidance on learning performance. (F: failures, G: first success).

Measure	Mean no guide	Mean guide	chg	t(26)	p
# trials	28.52	14.6	49%	2.68	$<.01$
# actions	816.44	368	55%	2.91	$<.01$
# F	18.89	11.8	38%	2.61	$<.01$
# F before G	18.7	11	41%	2.82	$<.01$
# states	124.44	62.7	50%	5.64	$<.001$

sages during the `sense-guidance` step (a short delay allows the teacher time to administer guidance). Upon receiving a guidance message the agent saves the object as the `guidance-object`. In the `select-action` step, the default behavior (a standard approach) chooses randomly between the set of actions with the highest Q-values, within a bound β. If any guidance was received, the agent will *instead* choose randomly between the set of actions that have the `guidance-object` as their object.

Evaluation of Guidance: In the second experiment we evaluate the effects of this guidance feature by analyzing training sessions with human subjects in two conditions: the `no guidance` condition has feedback only; the `guidance` condition has both guidance and feedback available. The comparison is summarized in Table 1. The `guidance` condition shows improvements in how long it takes to learn the task. The number of training trials needed to learn the task was 49% less; and the number actions needed to learn the task was 55% less. In the `guidance` condition the number of unique states visited was 50% less, thus the task was learned more efficiently and presumably without visiting as many non-useful states. And finally the `guidance` condition provided a more successful training experience. The number of trials ending in failure was 38% less, and the number of failed trials before the first successful trial was 41% less.

Motivation: Findings from Experiment 1: For many people, a large majority of rewards given were positive, the mean percentage of positive rewards for all the players was 69.8%. We thought this may be due to the agent improving and behaving more correctly over time (soliciting more positive rewards); however, data from the first quarter of training shows that well before the agent is behaving correctly, the majority of participants show a positive bias. Fig. 4 shows reward histograms for each participant's first quarter of training; the number of negative rewards on the left and positive rewards on the right. Most participants have a much larger bar on the right. A plausible hypothesis is that people are falling into a natural teaching interaction with the agent, treating it as a social entity that needs encouragement. Some people specifically mentioned in the interview that they felt positive feedback would be better for learning.

Modification to the Learning Agent to add Motivation: Due to this bias, a second embellishment to the communication channel adds a dedicated motivational input. This is done by considering a reward motivational if it is administered *on* Sophie. For visual feedback the agent is shaded yellow to let the user know that a subsequent reward will be motivational. Figure 2(c) shows a positive motivational message to Sophie. Instructions

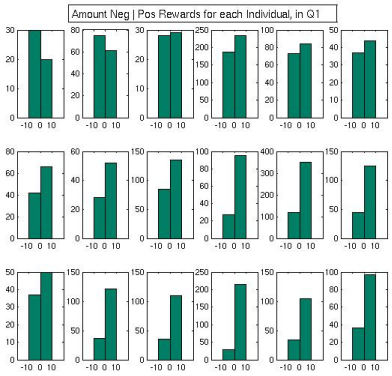


Fig. 4. Histograms of rewards for each individual in the first quarter of their session. The left column is negative rewards and the right is positive rewards.

indicate this communication channel is for general feedback about the task (e.g. ”Doing good Sophie!” or ”Doing bad!”) as opposed to feedback about a particular action.

Evaluation of Motivation: In the second experiment, players that had the motivation signal had a significantly more balanced feedback valance than the players that did not have it. Players that did not have a motivational channel had a mean ratio (*#positive*/*#negative*) of 2.07; whereas those with the motivational channel had a mean ratio of 1.688. This is a significant effect, $t(96) = -2.02, p = .022$. Thus, we conclude that motivation is a separate intention that was folded into the positive feedback in the initial study. Future work is to understand how an agent can utilize this signal in a different way to improve the learning interaction.

4.2 The Human Tries to Maximize Their Impact

In human learning, teachers direct a learner’s attention, structure experiences, support attempts, and regulate complexity. The learner contributes by revealing their internal state to help guide the teaching process. This *collaborative* aspect of teaching and learning has been stressed in prior work [11], and the findings in this study support this notion of *partnership*. When everyday users train a machine learning agent, we see them adjust their training as the interaction proceeds, reacting to the behavior of the learner.

Findings from Experiment 1: We expected people would habituate to the activity and that feedback would decrease over the training, informed by related work [9]. However, we see an increasing trend in the rewards-to-actions ratio over the first three quarters of training. Fig. 5 shows data for the first three quarters of training, each graph has one bar for each player indicating the ratio of rewards to actions. By the third graph more bars are approaching or surpassing a ratio of 1. One explanation for this trend is a shift in mental model; as people realize the impact of their feedback they adjusted their training to fit this model of the learner. This finds anecdotal support in the interview responses. Many users reported that once they concluded that their feedback was helping they subsequently gave more rewards.

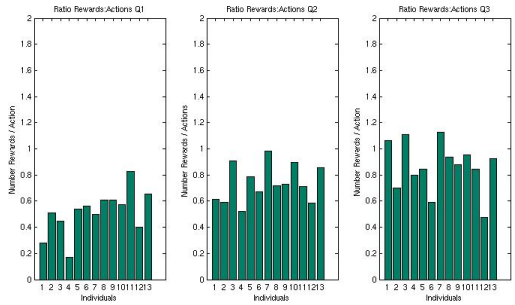


Fig. 5. The ratio of rewards to actions over the first three quarters of the training sessions shows an increasing trend.

Additionally, we had a second hypothesis about the unbalanced use of feedback (Fig. 4), namely that the agent did not have enough of a behavioral response to negative rewards. A typical RL agent does not have an instantaneous reaction to either positive or negative rewards, but particularly in the case of negative rewards, this could be interpreted as the agent 'ignoring' the feedback. The user may stop using them when they feel the agent is not paying attention to them.

These are concrete examples of the human trainer's propensity to learn from the agent how to best impact the process. This presents a huge opportunity for an interactive learning agent to *improve its own learning environment* by communicating more of its internal state to the teacher, making the learning process more transparent. This led us to make the following two modifications: 1) adding a gaze behavior to increase the transparency and foster more timely and relevant guidance, and 2) adding an `UNDO` behavior to create a more immediate response to negative feedback.

Gaze Behavior: Modification to the Learning Agent: We explored gaze as a means of making the learning process more transparent to the human. Gaze requires that the agent have a physical/graphical embodiment that can be understood by the human as having a forward heading. In general, gaze precedes an action and communicates something about the action that is going to follow.

We extended *Sophie's Kitchen* to add gaze. Recall our interactive Q-Learning loop: `sense-guidance`, `select-action`, `take-action`, `sense-reward`, `update-values`. In the `sense-guidance` phase, the learning agent now finds the set of actions, A, with the highest Q-values, within a bound β. For every action, a, in A, the agent gazes for 1 second at the `object` of a (if it has one). This communicates a level of uncertainty through the amount of gazing that precedes an action. It introduces an additional delay (proportional to uncertainty) prior to `select-action`, soliciting and providing the opportunity for guidance. We expect this transparency to improve the teacher's model of the learner, creating a more understandable interaction for the human and a better learning environment for the agent.

Evaluation of Gaze Behavior: Our hypothesis is that the gazing behavior helps the human understand when the agent did (and did not) need their guidance instruc-

Table 2. 1-tailed t-test. Gaze helped players give more guidance if needed and less if not.

Measure	Mean gaze	Mean no gaze	t(51)	p
% guidance when $<= 3$ choices	.79	.85	-2.22	.01
% guidance when $>= 3$ choices	.48	.36	1.96	.03

tion. We evaluate this in the second experiment, studying players that used the feedback and guidance *without gaze* versus those that had the feedback and guidance *with gaze* (summarized in Table 2). Note that players without the gaze behavior had ample opportunity to administer guidance messages; however, the time that the agent waits is uniform throughout the interaction. We found that the players in the gaze condition gave less guidance than the no gaze condition when the agent had three or less action choices (uncertainty low), $t(51) = -2.22, p = .015$. And conversely they gave more guidance then the no gaze condition when the agent had three or more action choices (uncertainty high), $t(51) = 1.96, p = .027$. Thus, when the agent uses the gaze behavior to indicate which actions it is considering, the human trainers do a better job matching their instruction to the needs of the agent throughout the training session.

Undo Behavior: Modification to the Learning Agent: In the standard Q-Learning framework, the effects of feedback on the policy are not seen until the state from which the action was made is revisited (which can take a long time). In many cases being too responsive to any one reward would be detrimental to the exploration needed to learn; however, we argue that when this signal is from a benevolent human teacher the agent should be more responsive to negative feedback. We expect just-in-time error correction more closely resembles a natural human teaching interaction and will be more understandable for the human.

We modified the Sophie agent to respond to negative feedback with an `UNDO` behavior (a natural correlate or opposite action) when possible. A negative reward is handled in both the `update-values` step and the subsequent `select-action` step. In the `select-action` step immediately following negative feedback, the action selection mechanism chooses the action that reverses the last action if possible. The proper `UNDO` behavior is represented within each primitive action (e.g. `GO <direction>` returns `GO <-direction>`; `PICK-UP <object>` returns `PUT-DOWN <object>`; etc.).

Evaluation of Undo Behavior: We found the `UNDO` response to negative feedback from the human trainer significantly improves the learning performance of the agent in a number of ways. In the second experiment, we compare players that had the `UNDO` behavior to those that did not. Table 3 summarizes these results. The agent visits 13% fewer states using 12% fewer action with the `UNDO` behavior, thus the learning process is more efficient. The agent takes 21% fewer actions before its first success. The `UNDO` behavior helps the agent avoid failures: The total number of failed trials was 37% less, and the number of failed trials before the first successful trial was 40% less.

Table 3. 1-tailed t-test, effects of UNDO on learning performance. (F = failures, G = first success).

Measure	Mean without undo	Mean with undo	chg	t(96)	p
# states	48.3	42	13%	-2.26	=.01
# F	6.94	4.37	37%	-3.76	<.001
# F before G	6.4	3.87	40%	-3.7	<.001
# actions to G	208.86	164.93	21%	-2.25	=.01
# actions	255.68	224.2	12%	-1.32	=.095

5 Discussion

This work emphasizes the *interactive* elements in teaching. There are inherently two sides to an interaction, in this case the human teacher and the machine learner. Our approach aims to enhance machine learning from both perspectives: modifying the algorithm to build a better learning agent, and modifying the interaction techniques to provide a better experience for the human teacher. Understanding how humans want to teach is an important part of this process.

Our studies show that *people want to guide the agent*. This makes intuitive sense, and techniques like 'luring' the agent into particular behaviors have been explored [4, 1]. In the past this choice has been inspired by animal learning, but our work formalizes this inspiration, grounding it in human behavior with a game character. In *Sophie's Kitchen* we observed people's desire to guide the character to an object of attention, even when explicitly told that only feedback messages were supported. In doing this people meant to bias the action selection mechanism of the RL algorithm. When we allow this, introducing a separate interaction channel for attention direction and modifying the action selection mechanism of the algorithm, we see a significant improvement in the agent's learning performance.

We also see that players treat the agent as a social entity and want a *motivational channel of communication* to encourage it. This is seen despite the fact that the learning agent is not particularly human-like. One can assume this effect will only be more prominent with agents designed to be socially and emotionally appealing. We argue that to build successful agents that learn from people, attention of the research community should focus on understanding and supporting the psychology and social expectations of the human teacher. It remains future work to explore how this motivational channel of communication should influence the learning algorithm in a different way than ordinary feedback. Our hypothesis is that this communication is intended to influence the internal motivations, drives and goals of the agent.

Another important question of this research is how the learning agent's behavior can shape the behavior of the teacher. In a social learning interaction both learner and teacher influence the performance of the tutorial dyad. While this observation seems straightforward in the human literature, little attention has been paid to the communication between human teacher and artificial agent in the machine learning literature. Particularly, we believe that the transparency of the learner's internal process is paramount to the success of the tutorial dialog. However, a fine balance must be struck between engulfing the human teacher with all pertinent information, and leaving them in the

dark. This work offers a concrete example that *transparent behavior can improve the agent's learning environment*. Specifically, when the learning agent uses gaze to reveal its potential next actions, people were significantly better at providing more guidance when it was needed and less when it was not. Thus the agent, through its own behavior, was able to shape the human's input to be more appropriate.

Additionally these transparency behaviors boost the realism and believability of the character, thereby making it more engaging for the human. The creation of believable characters that people find emotionally appealing and engaging has long been a challenge [14, 15]. Autonomy complicates this goal further, since the agent has to continually make action choices that are useful as well as believable. Blumberg *et al.* have some of the most extensive work in this domain [16, 17] within a dog learning context. Thus another challenge for teachable characters is to be appropriately responsive to the human's instruction. In this work we have studied one aspect of such responsiveness, informed by our initial study. Negative feedback from a human teacher can be treated as both feedback and suggestion to *reverse the action if possible*. With this strategy the agent's learning is improved in both speed and efficiency.

We chose to use Q-Learning for this work because it is widely understood. Thus affording the transfer of these lessons and modifications to any reinforcement-based approach. We have shown significant improvements in an RL domain showing that learning in a situated interaction with a human partner can help overcome some well recognized problems of RL. Furthermore, these improvements in performance will contribute to higher quality interactive characters and more fun and enjoyable game play.

In sum, our empirically informed modifications indicate ways that an interactive game agent can be designed to learn better and faster from a human teacher across several dimensions. Our studies show the modified agent is able to learn tasks using fewer executed actions over fewer trials. Our modifications also led to a more efficient exploration strategy with less time in irrelevant states. A learning process, as such, that is seen as less random and more sensible will lead to more understandable and more believable game characters. The guidance and undo modifications illustrate avenues to a more responsive and natural interactive character. Our modifications also led to fewer failed trials and less time to the first successful trial. This is a particularly important improvement for interactive characters in that it implies a less frustrating experience, which in turn creates a more fun and engaging interaction.

6 Conclusion

Dynamic and engaging teachable characters that learn from human players would usher a new genre of entertainment technology, and we posit that such agents should support how people naturally approach teaching. Accordingly, this paper describes an experimental platform and a series of studies that reveal lessons about how people interactively teach graphical AI characters via reward and punishment.

Given these findings, we made modifications to our learning agent and interface to improve the interaction. Our modifications include: an embellished communication channel that distinguishes guidance, feedback, and motivational intents; transparency behaviors that reveal aspects of the learning process to the human; and a more natu-

ral reaction to negative feedback. A second set of studies show that these empirically informed modifications improve several learning dimensions including the speed of learning, the efficiency of exploration, the human's ability to understand the agent's learning process, and a significant drop in the number of failed trials.

These empirical results inform and ground the design of teachable characters. We believe these lessons and modifications are not specific only to the particular algorithm and agent used in these studies, and that this work broadly contributes to the creation of fun and engaging teachable characters that learn in real-time from human players.

References

1. Evans, R.: Varieties of learning. In Rabin, S., ed.: AI Game Programming Wisdom. Charles River Media, Hingham, MA (2002) 567–578
2. Stanley, K.O., Bryant, B.D., Miikkulainen, R.: Evolving neural network agents in the nero video game. In: Proceedings of IEEE 2005 Symposium on Computational Intelligence and Games (CIG'05). (2005)
3. Stern, A., Frank, A., Resner, B.: Virtual petz (video session): a hybrid approach to creating autonomous, lifelike dogz and catz. In: AGENTS '98: Proceedings of the second international conference on Autonomous agents, New York, NY, USA, ACM Press (1998) 334–335
4. Blumberg, B., Downie, M., Ivanov, Y., Berlin, M., Johnson, M., Tomlinson, B.: Integrated learning for interactive synthetic characters. In: Proceedings of the ACM SIGGRAPH. (2002)
5. Sutton, R.S., Barto, A.G. In: Reinforcement learning: An introduction. MIT Press, Cambridge, MA (1998)
6. Watkins, C., Dayan, P.: Q-learning. Machine Learning **8**(3) (1992) 279–292
7. Kaelbling, L.P., Littman, M.L., Moore, A.P.: Reinforcement learning: A survey. Journal of Artificial Intelligence Research **4** (1996) 237–285
8. Kaplan, F., Oudeyer, P.Y., Kubinyi, E., Miklosi, A.: Robotic clicker training. Robotics and Autonomous Systems **38(3-4)** (2002) 197–206
9. Isbell, C., Shelton, C., Kearns, M., Singh, S., Stone, P.: Cobot: A social reinforcement learning agent. 5th Intern. Conf. on Autonomous Agents (2001)
10. Kuhlmann, G., Stone, P., Mooney, R.J., Shavlik, J.W.: Guiding a reinforcement learner with natural language advice: Initial results in robocup soccer. In: Proceedings of the AAAI-2004 Workshop on Supervisory Control of Learning and Adaptive Systems, San Jose, CA (2004)
11. Breazeal, C., Brooks, A., Gray, J., Hoffman, G., Lieberman, J., Lee, H., Lockerd, A., Mulanda, D.: Tutelage and collaboration for humanoid robots. International Journal of Humanoid Robotics **1**(2) (2004)
12. Breazeal, C., Kidd, C., Thomaz, A.L., Hoffman, G., Berlin, M.: Effects of nonverbal communication on efficiency and robustness in human-robot teamwork. In: Proceedings of the IROS. (2005)
13. Cassell, J., Vilhjalmsson, H.H., Bickmore, T.: Beat: the behavior expression animation toolkit. In: SIGGRAPH '01: Proceedings of the 28th annual conference on Computer graphics and interactive techniques, New York, NY, USA, ACM Press (2001) 477–486
14. Thomas, F., Johnson, O.: Disney Animation: The Illusion of Life. Abbeville Press, New York (1981)
15. Bates, J.: The role of emotion in believable agents. Communications of the ACM **37**(7) (1997) 122–125
16. Blumberg, B.: Old tricks, new dogs: ethology and interactive creatures. PhD thesis, Massachusetts Institute of Technology (1997)
17. Tomlinson, B., Blumberg, B.: Social synthetic characters. Computer Graphics **26**(2) (2002)