An Experimental Ethics Approach to Robot Ethics Education

Tom Williams
Colorado School of Mines
Golden, CO 80402

Qin Zhu
Colorado School of Mines
Golden, CO 80402

Daniel Grollman
Plus One Robotics
Boulder, CO 80301

Abstract

We propose an *experimental ethics*-based curricular module for an undergraduate course on Robot Ethics. The proposed module aims to teach students how human subjects research methods can be used to investigate potential ethical concerns arising in human-robot interaction, by engaging those students in real experimental ethics research. In this paper we describe the proposed curricular module, describe our implementation of that module within a Robot Ethics course offered at a medium-sized engineering university, and statistically evaluate the effectiveness of the proposed curricular module in achieving desired learning objectives. While our results do not provide clear evidence of a quantifiable benefit to undergraduate achievement of the described learning objectives, we note that the module did provide additional learning opportunities for graduate students in the course, as they helped to supervise, analyze, and write up the results of this undergraduate-performed research experiment.

Introduction

Computer Science educators are increasingly acknowledging that it is insufficient for the Computer Science curriculum to entirely focus on technical issues relating to computing and programming. Rather, for Computer Science students to be effective practitioners after graduation, their education must include key concepts from the arts, social sciences, and humanities, so that students not only have the technical knowledge necessary to implement and analyze computational systems, but also have the knowledge from those other fields necessary to decide, on ethical grounds, whether they *should* implement those systems, and if so, how they should go about designing and evaluating the effectiveness of those systems. This is especially true in the fields of Artificial Intelligence and Human-Robot Interaction, in which myriad ethical concerns have captured the public's attention, and in which human interactivity necessitates design and evaluation techniques not otherwise taught in the Computer Science curriculum.

Appropriate teaching of these skills is made especially challenging due to the interdisciplinary nature of the subject matter. AI Ethics education must clearly cover key concepts from ethics and moral philosophy as well as the use of those concepts to analyze applications of AI and Robotics. But moreover, such courses may also cover computational approaches to moral decision making; psychological theories of moral decision making and blame; and methods for experimentally investigating ethical issues. This requires instructors with broad interdisciplinary backgrounds, and pedagogical methods that draw on these disparate disciplines.

In this work, we explore pedagogical techniques aimed at improving students achievement of learning objectives that sit at this confluence of disciplines. Specifically, we present and analyze the efficacy of a *Robot Ethics* course module in which students participate *as experimenters* in experimental robot ethics research, which requires them to simultaneously learn methodological approaches to the study of experimental robot ethics, and then use those methods to engage with key theoretical concepts from robot ethics (in our case, robots' normative influence). Moreover, in this work we not only analyze the efficacy of this module, but additionally interrogate how this efficacy depends on the *role* that each student played in the research process.

This educational research effort thus involves two nested levels of experimentation: a randomized controlled experiment in which the research participants were undergraduate students enrolled in a Robot Ethics class and the researchers were the course staff, and a second randomized controlled experiment in which the research participants were undergraduate students sampled from across the university, and the researchers were both the Robot Ethics students and their instructors. In this paper, we will focus on the first of these experiments (the AI Education research effort), leaving many of the details of the second experiment (the experimental Robot Ethics research effort) to publication elsewhere.

As described in this paper, our analysis yielded mixed results with respect to the efficacy of this experimental course module. While we did not find any evidence that participating in experimental ethics research was any more effective than merely listening to a traditional lecture about the research effort and its goals *in general*, we did find that for participants that participated in the research effort, the *role* they played in the research effort may indeed have affected their learning of the concepts explored in that research. Moreover, as we will describe, the module provided additional learning opportunities to the small number of graduate students in the course, who supervised undergraduate students, and directly contributed to the data analysis and writing of the scientific paper submission resulting from the proposed curricular module.

Background

AI Ethics Education

Explicit discussions of AI or "expert systems" in computer ethics education literature and textbooks can be traced back to the 1990s, although most of these discussions are often speculative reflections about broader "macro" and social impacts of AI on humans, cultures, and societies. For instance, Forester and Morrison (1994) hold a humanistic and speculative view toward the employment of AI in the society and their major concerns include: (1) whether AI is a proper goal as most AI projects are funded by the military; (2) it is a technocratic idea to employ AI in public administration, legal practice, and social governance; (3) introducing AI to developing countries is another techno-fix that attempts to remedy the symptoms without addressing the causes; and (4) AI degrades the human condition.

Two curriculum design approaches have been developed to teach ethics in AI and robotics: (1) standalone courses (these courses can be offered in either computer science or philosophy); and (2) ethics modules in AI and robotics courses (Burton et al. 2017). Many of these curriculum development approaches are not so much different from traditional applied ethics approaches to teaching ethics of technology and engineering. For instance, in order to understand and discuss AI ethics issues, Burton et al. (2017) suggest that it is necessary for students to be familiar with three major ethical theories (deontology, consequentialism or utilitarianism, and virtue ethics) as tools for ethical decision-making. Students are then invited to practice on how to use the three ethical theories to analyze specific AI ethical situations and formulate possible courses of actions. Burton et al. (2017) point out that teaching students the three ethical theories and their applications in specific cases studies can be achieved in either one module in a technical course or a semester-long, standalone AI ethics course (with additional readings and case studies).

An increasing number of universities such as Harvard, Stanford, and the University of Kentucky have started to offer AI and robot ethics in their computer science curriculum. Educators at these institutions have been experimenting with innovative ethics pedagogies. Burton, Goldsmith, and Mattei (2015; 2018) have explored the use of science fiction as a pedagogical tool for teaching AI ethics. They argue that using science fiction as a pedagogy has at least two strengths compared to traditional pedagogies such as lectures: (1) the futuristic settings of science fictions enable students to detach or "decontextualize" from political preconceptions; and (2) science fictions have been proven to be appealing and popular to students (Burton, Goldsmith, and Mattei 2015). Science fictions provide students with a "safe" environment for "discussing and reasoning about difficult and emotionally charged issues without making the discussion personal" (Burton, Goldsmith, and Mattei 2018).

The same argument can be made for other similar media-based pedagogies such as using movies to teach AI and robot ethics. However, from the perspective of moral psychology, there is a gap between ethical reasoning (e.g., knowing what is good vs. bad and why) and ethical action (e.g., someone is committed to do good) (Rest et al. 1999). Effective professional ethics education requires future professionals to relate their moral learning experience to their own everyday personal and professional experience (or how they actually do things) (Martin 2000). For computer science students, it is crucial to reflect on how their moral learning experience is relevant to their everyday, practical experience, empathizing with potential users and their needs, and reflecting on the (powerful) role of their expertise in shaping the society. As such, it is critical to consider how real-world examples and hands-on experiences with realistic AI and robotics technologies may help to fill this gap.

Carnegie Mellons "Artificial Intelligence Methods for Social Good" course, for example, goes beyond the traditional instructional approach that teaches theories of AI and robot ethics through classroom lectures alone. Students in the course instead acquire practical experience through research projects that employ AI methods to address pressing social issues in fields such as healthcare, social welfare, security and privacy, and environmental sustainability (Hsu 2018). Such hands-on experience may help students develop sensitivity to the normative influence of technology on humans and the society. Arguably, this model of teaching computer science students to perform or experimentally investigate ethics in a technical class has some advantages. To some extent, it helps make visible the values that are embedded in the design of AI and robotic technologies. Furthermore, it creates a mindset among students that technological development involves value choices.

As science and engineering education curricula are already packed, integrating ethics modules into technical courses is often more realistic for faculty. Furey and Martin (2018) have shared their experience with integrating a module about the ethics of algorithm development for autonomous vehicles into a semester-long AI course. They argue that there are certain advantages with such modular approach to AI ethics education: (1) modules are easily integrated into technical courses; and (2) students can connect the specific AI ideas they are learning in class to their ethical implications. In their module, the classical Trolley Problem and the utilitarian ethical framework were introduced to students and employed as tools for evaluating the benefits and costs of algorithm development. Nevertheless, one major concern with integrating ethics modules in technical classes is that they may create an impression that ethics is added or supplementary to technology.

Research-Based Pedagogy

Education researchers have established numerous benefits of undergraduate involvement in research outside of the classroom (Kardash 2000; Landrum and Nelsen 2002; Strayhorn 2010; Hathaway, Nagda, and Gregerman 2002; Laursen et al. 2010), prompting researchers to explore the integration of laboratory research into undergraduate curriculum itself, in the form of laboratory research modules (Lopatto 2010). These efforts have had great success, finding that involvement of undergraduates in the research process as part of classroom learning notably increases interest in science and graduate studies (Harrison et al. 2011), improves specific course topic mastery, and benefits students' general critical thinking skills (Harrison et al. 2011). While much of this work has focused on integration of laboratory research modules into upper-level classes, more recent work has sought to expose students to research even earlier by integrating laboratory research modules into lower-level courses as well (Harrison et al. 2011; Coker 2017), as one of the largest barriers to student involvement in research is awareness (Wayment and Dickson 2008).

Based on the success of these educational research efforts, it may be valuable to build on those approaches by integrating laboratory research modules into AI and Robot ethics classes. We believe this will be successful for a number of reasons. First, we would expect students to incur the same benefits as have been observed with previous laboratory research modules (including improved mastery of course content, interest in graduate study, and improved critical thinking skills). Second, we believe this may be a unique opportunity to expose students from a variety of fields to research opportunities. Not only are AI and Robot Ethics courses appealing to students from a wide variety of disciplines[1], but the legwork of experimental ethics research can easily be performed by undergraduates from disparate backgrounds, without requiring them to have deep knowledge of the concepts being explored (assuming that experiments themselves are designed by instructors rather than students). Finally, it helps provide some quantitative content to an otherwise qualitative course, which may help build interest in the course's other subject matter for those engineering students who are otherwise reticent to engage with the course content.

Experimental Ethics Curriculum Design

In this section, we describe the design and implementation of a research-based experimental ethics module into a primarily undergraduate robot ethics class at a medium-sized engineering university.

Our proposed curricular module was designed to achieve the following interdisciplinary learning objectives:

LO1: Normative Influence of Technology – Students should understand how technologies (like robots) can exert influence on human behaviors due to their perception as moral and social agents, and further understand how this influence can carry over into human-human relationships.

LO2: Experimental Ethics – Students should understand how human-subject experimentation can be used to explore the ethical implications of technology (in this case, hypothesized normative influence of decisions made during robot interaction design.)

LO3: Ethical Research Conduct – Students should understand ethical concerns that can arise in the design and conduction of Experimental Ethics experiments, and how those concerns should be addressed.

To achieve these learning objectives, our curricular design followed a multi-stage process.

Phase Zero: Research Ethics Certification

Before the Experimental Ethics Module is introduced in class, all students complete the CITI (Collaborative Institutional Training Initiative) Social-Behavioral-Educational basic course; a three-hour online course comprised of reading passages interleaved with short quizzes, designed to teach social, behavioral, and educational researchers the basics of ethical research conduct, including risk assessment, informed consent, research concerns with specialized populations, and so forth (Braunschweiger and Hansen 2010).

This phase serves two purposes. First, it lays the groundwork for fulfillment of Learning Objective 3 by introducing and assessing key tenets of ethical research conduct. Second, it provides students with the certification needed to participate in conducting IRB-approved human-subjects research.

Phase One: Classroom Lecture

Next, students are introduced to key curricular concepts through a 45-minute classroom lecture[2]. This lecture begins with an introduction to and motivation of experimental moral philosophy, experimental moral psychology, and ethics-oriented empirical studies of human-robot interaction. The lecture then introduces an ethical concern surrounding human-robot interaction design, and the design of a human subject experiment intended to assess the validity of that concern, including the experiment's hypotheses, design, and procedure, and how data collected through this experiment can be statistically analyzed post-experiment to test the experiment's hypotheses.

This phase serves two purposes, laying the groundwork for fulfillment of Learning Objectives 1 and 2, by introducing and motivating experimental ethics and the core robot ethics topic underlying the presented experiment (in our case, normative influence of technology).

Phase Two: Hands-On Training

Immediately after receiving this lecture, students travel to a human-robot interaction laboratory configured for human-subject experimentation and receive hands-on training for conducting the experiment. As described later on, our particular instantiation of the proposed curriculum involved training students either in the *experimenter* role (in which

[1] In our own Robot Ethics course, our students this year came from Applied Math and Statistics, Computer Science, Electrical Engineering, Engineering Physics, Mechanical Engineering, and Petroleum Engineering.

[2] Because the proposed module is intended to be highly flexible, we will for the moment leave the definition of these concepts vague; in the *Implementation* section below, we will go on to describe the specific concepts investigated in our implementation of the proposed curricular module.

students learned how to guide participants through consent procedures, brief participants on their experimental task, and debrief participants on experimental motivations upon study completion) or in the *wizard* role (in which students learned how to teleoperate the robot used in the experiment).

This phase serves to reinforce the key concepts necessary to achieve all three learning objectives.

Phase Three: Research Participation

Finally, students leverage their training to assist in conducting the proposed experiment, with each student helping to run three participants through the experiment over the course of several weeks.

This phase serves to further reinforce the key concepts necessary to achieve all three learning objectives, thus incorporating real-world praxis into the learning process.

Assessments

Once all phases of the proposed curriculum are completed, student learning is assessed through an in-class quiz in which students must recall details of the experiment related to all three learning objectives, including research hypotheses, metrics, experimental design, and consent procedures.

Implementation

While the proposed curricular module is designed to be sufficiently flexible to teach a wide range of core Robot Ethics concepts, our specific implementation of this curricular module focused specifically on teaching the concepts necessary to achieve Learning Objective 1, i.e., Normative Influence of Technology. In this section we will briefly summarize this concept, and the experiment students were involved with on that topic.

With the increase of internet of things (IoT) technologies, voice interfaces are being added to a wide array of home and work appliances, including refrigerators, microwaves, and even faucets (Faucet 2019). Despite several decades of research into mixed initiative dialogue (Allen, Guinn, and Horvitz 1999; Horvitz 1999) and turn taking (Cassell, Torres, and Prevost 1999; Traum and Rickel 2002), the dominant paradigm in consumer-grade voice interaction is to use platform-specific *wakewords*, such as “Alexa”, “Okay Google”, or “Hey Siri”, to prevent false positives in speech recognition and improve user privacy. However, there has been significant public concern expressed in the mass media that wakeword-driven interactions may encourage technology-directed language that is terse and direct, and that if children become accustomed to addressing machines in this manner, this behavior could carry over into their interactions with other humans, leading to impolite human-directed behavior (Gordon 2018; Truong 2016). As consumer-facing interactive robots begin to be deployed into the wild, they will also likely require wakeword-based interaction, potentially with greater risk of these feared effects.

Human networks of social and moral norms are well known to be dynamic and malleable (Gino 2015), with norms defined, communicated, and enforced by community members (and the technologies with which they interact) (Verbeek 2011). As we have recently argued in our own work, social robots wield unique influence over these norms due to their unique sociotechnical niche, defined by their joint status as perceived community members and as technological tools (Jackson and Wililams 2019). This influence, which social robots may wield both through direct persuasion and implicit social pressure (Briggs and Scheutz 2014; Kennedy, Baxter, and Belpaeme 2014; Jackson and Williams 2019; Winkle et al. 2019), may be especially strong among language capable robots, due to greater levels of perceived social and moral agency, leading to greater influence on users’ systems of social and moral norms, including sociocultural norms such as norms of politeness.

Figure 1: The SoftBank Pepper robot used in our laboratory research module.

In our implementation of the proposed curricular module, our in-class lecture and class-run experiment investigated this concern: the experiment examined the effect a designer’s *choice* of wakeword might have on both robot- and human-directed politeness. In the experiment, participants interacted with a SoftBank Pepper robot (Fig. 1) and a human confederate in a restaurant scenario. When interacting with the robot, participants were required, depending on their experimental condition, to use either a traditional wakeword (e.g., “Hey Pepper”) or a polite wakeword (e.g., “Excuse me, Pepper”). To investigate normative influence of technology, the experiment collected linguistic statistics surrounding participants’ use of politeness cues in their language towards both the robot and the human confederate throughout the experiment.

Experimental Evaluation

The proposed curriculum was evaluated through a randomized controlled experiment performed through a Robot Ethics class at the Colorado School of Mines in Spring 2019. This course was crosslisted so as to be offered to a wide variety of students, with students able to take the course either for upper-level Computer Science or Humanities credit, or for graduate credit, with differing writing and programming requirements depending on the type of credit earned. The assessment measures were used to evaluate efficacy of curriculum. Data was collected from 32 undergraduate students in this course (25 Male, 6 Female, 1 NA/Other).

Procedure

At the beginning of the semester, IRB exemption was acquired for both our educational research and the human-subjects experiment described in the previous section. Once students had completed Phase Zero of the curricular module, an amendment was approved adding all students in the class to the exempted protocol.

All undergraduate students in the class were then assigned to one of three roles. Nine students were assigned to participate in the experiment as an *experimenter*, guiding experimental participants through consent procedures and debriefing them at the end of the experiment. Ten students participated in the experiment as a *wizard*, controlling the robot's movements behind the scenes. Finally, thirteen students did not participate in the experiment, serving as a control condition. We originally intended to have eleven students in each of these three groups; however, after training had taken place, four students originally assigned to experimenter or wizard roles ultimately could not participate, e.g. due to overly restrictive schedules (Many of our undergraduates have occasional six-class semesters due to heavy course requirements and strict course sequencing). External to these three experimental conditions, Graduate students enrolled in the course were assigned to participate as *confederates*; actors who carried out the requests of participants during the experiment.

Phases One and Two were then carried out in an interleaved fashion: all students in the class first attended the first portion of the classroom lecture. Then, students in the wizard role and the control condition stayed in class for the second portion of the lecture while students in the experimenter role visited the lab, where graduate students led them through their version of phase two. Once this was complete, students in the experimenter role returned to class for the second portion of the lecture, while students in the wizard role visited the lab for their version of phase two, and students in the control condition left class early.

Finally, in the weeks following this lecture and training, the experimental ethics experiment (Phase Three) was run by the students in the class. Initial pilots of the experiment revealed that students in the wizard condition had difficulty accurately teleoperating the robot under the time pressures of live experimentation, and thus the experiment underwent minor revision. First, the graduate students in the class modified the robot used during the experiment (SoftBank's Pepper) to perform the experiment autonomously, without teleoperation. This allowed participants previously designated as wizards to instead serve as confederates; a role which required very little additional training. The graduate students in the class then, instead of serving as confederates, merely served as supervisors, attending experiment sessions to supervise undergraduate experimenters make sure that things ran smoothly. This modification thus yielded three new undergraduate groups: nine experimenters, ten confederates, and thirteen non-participating controls. From this point on, the experiment was run as expected, with each undergraduate in an experimenter or confederate role helping to run three experiment sessions each.

Once the experiment was completed, the assessment quiz was administered as an ungraded pop quiz. For the purposes of this educational research, we selected from among the responses given by students on the quiz a number of "key" concepts, and scored each quiz based on the number of key concepts recalled in each category (research hypotheses, research metrics, experimental design characteristics, and consent procedure requirements). Coding and scoring was performed blind with respect to students' experimental group membership. All students in the class then provided voluntary informed consent for the course staff to use their scores on this ungraded quiz for this educational research.

After this quiz was administered, students in the control group, who had not been required to participate in the running of the experiment, were given an annotation homework assignment, so that they could have the opportunity to be involved in the research effort and to prevent differences in workload across students, while allowing them to still function as a control group for the purposes of this educational research.

Hypotheses

In evaluating our proposed curriculum, we sought to test the following research hypotheses, each of which is associated with one of our key Learning Objectives.

H1: Normative Influence of Technology

(a) Students participating in the experimental ethics module will achieve better understanding of how technologies (like robots) can exert influence on human behaviors due to their perception as moral and social agents, and better understand how this influence can carry over into human-human relationships.

(b) This learning advantage will be especially true of students who participate in the Experimenter role (due to their repeated debriefing of experimental participants on the true focus of the research).

H2: Experimental Ethics

(a) Students participating in the experimental ethics module will achieve better understanding of how human-subject experimentation can be used to explore the ethical implications of technology (in this case, hypothesized normative influence of decisions made during robot interaction design.)

(b) This learning advantage will be especially true of students who participate in the Experimenter role (due to their repeated debriefing of experimental participants on the true focus of the research).

H3: Ethical Research Conduct

(a) Students participating in the experimental ethics module will achieve better understanding of ethical concerns that can arise in the design and conduction of Experimental Ethics experiments, and how those concerns should be addressed.

(b) This learning advantage will be especially true of students who participate in the Experimenter role (due to their repeated explanation and application of consent procedures to experimental participants).

Measures

To test Hypothesis 1, we assessed whether the mean number of research hypotheses and metrics recalled during the in-class quiz differed according to students' experimental role.

To test Hypothesis 2, we assessed whether the mean number of experimental design characteristics recalled during the in-class quiz differed according to students' experimental role.

To test Hypothesis 3, we assessed whether the mean number of consent procedure requirements recalled during the in-class quiz differed according to students' experimental role.

Results

We analyzed our results using the JASP (JASP Team 2016) software package for Bayesian statistical analysis. A Bayesian Analysis of Variance with Bayes factor analysis (Morey, Rouder, and Jamil 2015) was performed to assess the affect of experimental condition on mean number of key items recalled by students in each of the following categories (1) research hypotheses, (2) research metrics, (3) experimental design characteristics, and (4) consent procedure requirements.

As shown in Table 1, our results presented weak evidence *against* any impact of experimental condition on recall of experimental metrics (BF=0.328[3]), consent procedures (BF=0.346), or elements of experimental design (BF=0.332), and inconclusive evidence with respect to effect of experimental condition on recall of research hypotheses (BF=0.884).

Post-hoc analysis of this inconclusive result revealed weak evidence against a difference between non-participating students and confederate students (BF=0.385), inconclusive evidence regarding a difference between non-participating students and experimenter students (BF=0.980), and moderate evidence in favor of a difference between confederate students and experimenter students (BF=6.606), with students in the experimenter role overall recalling more key hypotheses (μ=2.56, SD=0.73) than students in the confederate role (μ=1.7, SD=0.48), as shown in Fig. 2.

Taking this data together, our results suggest that the type of experimental participation had no effect on students' recall of information regarding experimental procedure, but did have an effect on their recall of the experimental hypotheses, thus partially supporting Hypothesis 1 and refuting Hypotheses 2 and 3. specifically, students who participated in the experiment as an experimenter recalled more details of the experiment's goals than did students who participated as a confederate.

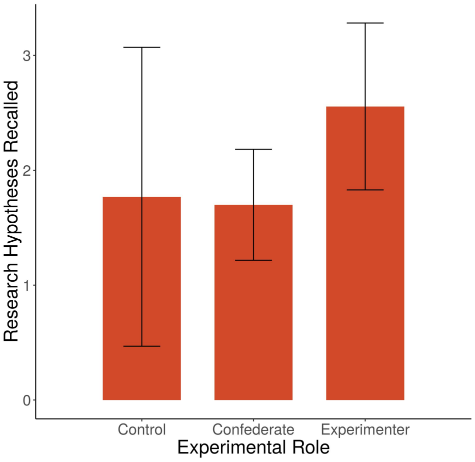

Figure 2: Observed differences in effect of experimental condition on recall of research hypotheses.

Discussion

In this section we will discussed practical lessons learned while implementing the proposed curriculum.

Scheduling

One of the major hurdles faced while implementing the proposed curriculum was student scheduling. While scheduling is typically fairly straightforward for normal experimental ethics experiments, this was not the case for this experiment due to the large number of experimenters, and due to the fact that most of the experimenters were undergraduate students with heavy courseloads.

Because each student was required to help run three experimental participants, and because each experimental session required two undergraduate students to run the experiment, the course staff needed to collect from every student in the class a schedule of times at which they would be able to run participants, and compare those schedules to identify times at which pairs of students were free, while making sure to assign students to timeslots for which few other students were available. After each participant was run through the experiment, the course staff then needed to take note of which students had run that participant; once a student had finished all three of their required experimental slots, the course staff often then needed to consider who might be able to fill in for future slots that that now-finished student had been scheduled to run (in the case that earlier assigned slots went unfilled).

This entire procedure was a major effort on the part of the course staff. If the proposed curriculum were to be used again, the course staff would need to ensure that a

[3] A Bayes Factor of 0.328 indicates that the ratio of probabilities between the two models is 0.328 times larger when measured using the posterior rather than the prior, indicating that the data observed was 0.328 times more likely if there were a difference between the groups than if there were not; or conversely, that it the data was $1/0.328 = 3.048$ times more likely to have been observed if there was *not* a difference between groups than if there were. A Bayes Factor with a value greater than 3.0 provides moderate evidence in support of the hypothesis in question: in this case, the null hypothesis.

Measure	Control	Confederate	Experimenter
Experimental Metrics	1.62 (1.12)	2.0 (0.82)	1.56 (0.53)
Consent Procedures	2.0 (1.35)	1.80 (0.63)	2.44 (1.13)
Elements of Experimental Design	0.62 (0.51)	0.60 (0.70)	0.33 (0.5)
Research Hypotheses	1.77 (1.3)	1.7 (0.48)	2.56 (0.73)

Table 1: **Assessment Results.** Each cell contains the mean (standard deviation) of key items in each category (rows) recalled by students in each condition (columns).

software-based scheduling solution were used. Due to the unique scheduling needs involved for the curricular module, a unique software solution would likely be required.

Sample Sizes

The statistical power of both the experimental ethics experiment and the educational research experiment were limited due to small sample sizes. The small size of the course meant that only a small number of students fell into each experimental category. Moreover, this meant that because we only required each student to run three experimental participants, the total number of datapoints collected in the experimental ethics experiment was also smaller than desired.

These issues could be ameliorated by increasing the number of sessions required of each student, or by accepting a larger number of students into the course. However, both of these solutions would have exacerbated the scheduling concerns described above, and may have significantly added to the cost of the experimental ethics experiment.

Graduate Students

While the implementation and analysis of our curricular module focused on the undergraduate students who served as experimenters or confederates during the experimental ethics experiment, the module may actually have been more effective for the graduate students in the class. While the graduate students were only responsible for supervising the undergraduates, and did not themselves interact with any experimental participants, they ended up participating much more deeply in the experiment, for a number of reasons.

First, because a small number of graduate students were responsible for supervising all experimental sessions, each graduate student ended up observing a much larger number of sessions than did each undergraduate student. Second, graduate students participated not just by following a script, but rather by observing and watching out for adverse events, to identify and correct for experimental problems, leading to a different level of engagement during experiments.

Third, because the graduate students in the course were so deeply involved with the management and refinement of the experimental ethics experiment, all students were given (and took) the opportunity to become authors on the scientific paper that resulted, each contributing writing, figures, data analysis, and/or supplemental videos, allowing for much deeper engagement with the material and a much more substantial involvement in the research process.

Unfortunately, this creates a paradox from an education research perspective: by involving graduate students in the scientific process and in paper writing, they likely made greater learning gains than they would have if they had not been involved in this way; but because the students are publicly named co-authors, their grades (e.g., quiz scores) cannot be analyzed without partial de-anonymization.

Conclusion

We conducted a nested educational research experiment in which students achieved learning objectives by taking an active role in an experimental ethics laboratory experiment. Our results suggest that the proposed curriculum yielded no learning gains with respect to traditional lecture-based curriculum, but for students who participated in the proposed curriculum, retention of the research hypotheses explored through the experimental ethics research experiment was greater for students who participated as an experimenter rather than as a confederate.

We do not believe these learning gains were sufficient to justify the added overhead imposed by this curriculum. However, it may be value to reexamine incorporation of this curriculum after addressing some of the process-based lessons learned while conducting the research. Moreover, it may be worth incorporating the curriculum for the benefits provided to graduate students in the course. Another option would be to explore the incorporation of experiment design and piloting (Coker 2017): in many human-robot interaction courses, for example, students are required to design and pilot a human-robot interaction experiment. Because these experiments are piloted rather than run with real participants, they have no associated cost; and because they are designed by students themselves, they lead to deeper engagement with the research questions under investigation. However, this means that they also are not publishable, and thus do not give the opportunity for research authorship. However, it may be worth examining the use of this pilot-oriented research curriculum in AI and Robot Ethics classes for their other benefits listed above.

Acknowledgments

This work was supported in part by a grant from the National Science Foundation (IIS-1849348) and in part by Colorado School of Mines 2018 Tech Fee CS-01. The authors would like to extend thanks to the graduate students of Colorado School of Mines' 2019 Robot Ethics class for their assistance in supervising the experiments described in this paper.

References

Allen, J. E.; Guinn, C. I.; and Horvitz, E. 1999. Mixed-initiative interaction. *IEEE Intelligent Systems and their Ap-*

plications 14(5):14–23.

Braunschweiger, P., and Hansen, K. 2010. Collaborative institutional training initiative (citi). *J Clin Res Best Pract* 6:1–6.

Briggs, G., and Scheutz, M. 2014. How robots can affect human behavior: Investigating the effects of robotic displays of protest and distress. *International Journal of Social Robotics* 6(3):343–355.

Burton, E.; Goldsmith, J.; Koenig, S.; Kuipers, B.; Mattei, N.; and Walsh, T. 2017. Ethical considerations in artificial intelligence courses. *AI magazine* 38(2):22–34.

Burton, E.; Goldsmith, J.; and Mattei, N. 2015. Teaching ai ethics using science fiction. In *Workshops at the Twenty-Ninth AAAI Conference on Artificial Intelligence*.

Burton, E.; Goldsmith, J.; and Mattei, N. 2018. How to teach computer ethics through science fiction. *Communications of the ACM* 61(8):54–64.

Cassell, J.; Torres, O. E.; and Prevost, S. 1999. Turn taking versus discourse structure. In *Machine conversations*. Springer. 143–153.

Coker, J. S. 2017. Student-designed experiments: A pedagogical design for introductory science labs. *Journal of College Science Teaching* 46(5):14.

Faucet, D. 2019. Voice faucet. `https://www.deltafaucet.com/Voice`. Accessed: 2019-06-26.

Forester, T., and Morrison, P. 1994. *Computer ethics: cautionary tales and ethical dilemmas in computing*. Mit Press.

Furey, H., and Martin, F. 2018. Introducing ethical thinking about autonomous vehicles into an ai course. In *Thirty-Second AAAI Conference on Artificial Intelligence*.

Gino, F. 2015. Understanding ordinary unethical behavior: Why people who value morality act immorally. *Current opinion in behavioral sciences* 3:107–111.

Gordon, K. 2018. Alexa and the age of casual rudeness. `https://www.theatlantic.com/family/archive/2018/04/alexa-manners-smart-speakers-command/558653/`. Accessed: '19-06-26.

Harrison, M.; Dunbar, D.; Ratmansky, L.; Boyd, K.; and Lopatto, D. 2011. Classroom-based science research at the introductory level: changes in career choices and attitude. *CBELife Sciences Education* 10(3):279–286.

Hathaway, R. S.; Nagda, B. A.; and Gregerman, S. R. 2002. The relationship of undergraduate research participation to graduate and professional education pursuit: An empirical study. *Jour. of College Student Development* 43(5):614–631.

Horvitz, E. 1999. Principles of mixed-initiative user interfaces. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*, 159–166. ACM.

Hsu, Y.-C. 2018. *Designing Interactive Systems for Community Citizen Science*. Ph.D. Dissertation, Ph. D. Dissertation. Carnegie Mellon University, Pittsburgh, PA.

Jackson, R. B., and Wililams, T. 2019. On perceived social and moral agency in natural language capable robots. In *2019 HRI Workshop on The Dark Side of Human-Robot Interaction*.

Jackson, R. B., and Williams, T. 2019. Language-capable robots may inadvertently weaken human moral norms. In *2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 401–410. IEEE.

JASP Team. 2016. Jasp. *Version 0.8. 0.0. software*.

Kardash, C. M. 2000. Evaluation of undergraduate research experience: Perceptions of undergraduate interns and their faculty mentors. *Journal of educational psychology* 92(1):191.

Kennedy, J.; Baxter, P.; and Belpaeme, T. 2014. Children comply with a robot's indirect requests. In *Proceedings of the 2014 ACM/IEEE international conference on Human-robot interaction*, 198–199. ACM.

Landrum, R. E., and Nelsen, L. R. 2002. The undergraduate research assistantship: An analysis of the benefits. *Teaching of Psychology* 29(1):15–19.

Laursen, S.; Hunter, A.-B.; Seymour, E.; Thiry, H.; and Melton, G. 2010. *Undergraduate research in the sciences: Engaging students in real science*. John Wiley & Sons.

Lopatto, D. 2010. Science in solution. *Tucson, AZ: Research Corporation for Science Advancement*.

Martin, M. W. 2000. *Meaningful work: Rethinking professional ethics*. Practical and Professional Eth.

Morey, R. D.; Rouder, J. N.; and Jamil, T. 2015. Bayesfactor: Computation of bayes factors for common designs. *R package version 0.9* 9:2014.

Rest, J.; Narvaez, D.; Bebeau, M.; and Thoma, S. 1999. A neo-kohlbergian approach: The dit and schema theory. *Educational Psychology Review* 11(4):291–324.

Strayhorn, T. L. 2010. Undergraduate research participation and stem graduate degree aspirations among students of color. *New Directions for Institutional Research* 2010(148):85–93.

Traum, D., and Rickel, J. 2002. Embodied agents for multi-party dialogue in immersive virtual worlds. In *Proceedings of the first international joint conference on Autonomous agents and multiagent systems: part 2*, 766–773. ACM.

Truong, A. 2016. Parents are worried the amazon echo is conditioning their kids to be rude. `https://qz.com/701521/parents-are-worried-the-amazon-echo-is-conditioning-their-kids-to-be-rude/`. Accessed: 2019-06-26.

Verbeek, P.-P. 2011. *Moralizing technology: Understanding and designing the morality of things*. University of Chicago Press.

Wayment, H. A., and Dickson, K. L. 2008. Increasing student participation in undergraduate research benefits students, faculty, and department. *Teaching of Psychology* 35(3):194–197.

Winkle, K.; Lemaignan, S.; Caleb-Solly, P.; Leonards, U.; Turton, A.; and Bremner, P. 2019. Effective persuasion strategies for socially assistive robots. In *2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 277–285. IEEE.