Psychological Measurement of Technology Ethics Education using the REGAIN Empirical Framework

Emily Foster-Hanson
Department of Psychology
Swarthmore College
Swarthmore, USA
fosterhanson@swarthmore.edu

Kerem Oktar
Department of Psychology
Princeton University
Princeton, USA
oktar@princeton.edu

Tania Lombrozo
Department of Psychology
Princeton University
Princeton, USA
lombrozo@princeton.edu

Steven Kelts
Center for Information Technology Policy
Princeton University
Princeton, USA
kelts@princeton.edu

2025 IEEE International Symposium on Ethics in Engineering, Science, and Technology (ETHICS) | 979-8-3315-3228-4/25/$31.00 ©2025 IEEE | DOI: 10.1109/ETHICS65148.2025.11098444

***Abstract*— Ethics coursework in higher education offers a key opportunity to shift the ethical culture of technology design and development. It could improve anticipation of potential tech harms, increase use of reasoning to address harms proactively, and change how students weigh values against other competing goals within the complex systems of tech companies. Yet, of the empirical measurements that have been developed to assess the effects of tech ethics coursework, most focus only on measuring the quality of students' abstract reasoning, not their ability to foresee problems or their intended strategies to address them in complex environments. Here we draw on evidence from the human psychology of belief and behavior change to develop a new framework for measuring the effects of tech ethics coursework. Our REGAIN framework assesses how students Reason about ethical decisions, Evaluate their own ethical decision-making, prioritize ethical Goals and values, become Aware of ethical dilemmas, acquire ethically-relevant Information, and perceive social Norms around ethical behavior. We describe the psychological research informing how we operationalize these constructs in the current framework, and we report a study[1] using this framework to measure the effects of a course on tech ethics at a research institution in the United States. Though we cannot draw conclusions about causation, our data suggests that students who completed a tech ethics course showed higher moral awareness of potential tech harms compared to a control condition. Tech ethics students also differed in their reasoning strategies and metacognitive judgements, and they reported stronger intentions to seek diverse perspectives and prioritize society's goals more than their own goals as developers.**

Keywords*— *ethics education, empirical measurement, cognitive mechanisms

I. Introduction

Over the next five years, 42% of companies expect to prioritize jobs using AI and big data [1]. How is higher education preparing our future workforce to engage ethically with these increasingly complex systems? The past decades of growth in the web, social media, and novel information and communication modes have generally seen these technologies exacerbate existing social problems and cause new ones. Researchers, policy-makers, and consumers have leveled criticisms that new tech entrenches racial and other biases [2], weakens data privacy protections [3], [4], undermines user autonomy [5], and more. The inventors of these technologies have often been trained in computer science departments at American universities. Yet despite efforts to build accreditation standards requiring coursework to sensitize computing students to the social effects of their work (ACM 2023; IEEE 2023), no standard is required of all departments.

Importantly, university courses offer a unique opportunity to shift ethical culture: students buy in (choose these classes themselves), commit to a whole semester (allowing extended interventions), and may be more receptive to behavioral interventions because they have less deeply ingrained habits and are in a state of transition [6]. Yet evidence as to whether and which strategies are effective is scant: While recent research across both industry and education has aimed to elicit more ethical behavior[7], [8], [9], [10], [11], [12], such research rarely empirically examines the efficacy of instruction (see below for a few notable exceptions). Moreover, students in engineering may even become less concerned about ethics across their education [13], suggesting that existing efforts at promoting ethical development are failing to capitalize on a major opportunity to shift ethical culture.

How would one go about empirically measuring the effects of tech ethics coursework—and hence take the first steps towards engineering courses that effectively promote ethical software? In this paper, we outline such a strategy through our REGAIN framework. Note that tech ethics courses are difficult to study because they vary substantially in both structure and content [14], [15], [16], [17], [18] and because tech ethics students are typically not yet in a position to make many measurable workplace decisions about technology development and design (see [19]). For this reason, here we focus more on the cognitive skills students may learn for anticipating possible consequences of their decisions.

Why study the ability to anticipate? To focus on one example, deceptive designs such as infinite scroll have exploited users' time and attention and amplified the addictive effects of social media [20]. Yet designers like Aza Raskin – credited with the invention of infinite scroll – often report that these downstream consequences were unanticipated, unforeseen, and

Funding for this research was provided by Google.

[1] This work involved human subjects in its research. Approval of all ethical and experimental procedures and protocols was granted by Princeton University.

unintended at the time of development [21]. Urgent work has begun on how researchers might better anticipate consequences, and how to understand claims about "unintended" effects [22], [23]. To address this at the college level, courses often use "case studies" of development gone wrong [24], but the cognitive skill of reflecting and passing judgment on known actions and their consequences is not the same as the skill of anticipating how the one leads to the other. There are important field reports of the effects of case studies [24], [25], but few specific measures have been developed to disaggregate students' moral judgments from their ability to anticipate consequences, and to measure the separate effects of tech ethics courses on both. Even fewer have been developed to measure which cognitive and decision-making processes might be important correlates of students' ability to anticipate the long-term ethical consequences of their decisions [26].

In this paper, we draw on evidence from the human psychology of belief and behavior change to develop a novel framework for measuring the effects of tech ethics coursework. We report a study using this framework to measure the effects of a course on tech ethics at a research institution in the northeastern United States. We conclude by highlighting both theoretical and practical suggestions for future empirical research. Though we cannot draw conclusions about causation, our data suggests that students who completed a tech ethics course showed higher moral awareness of potential tech harms compared to a control condition. Tech ethics students also differed in their reasoning strategies and metacognitive judgements, and they reported stronger intentions to seek diverse perspectives and prioritize society's goals more than their own goals as developers.

II. The REGAIN Framework

Our REGAIN framework focuses on measuring cognitive processes that we expect, based on psychology research, to be important aspects of moral decision-making that will generalize across contexts and individuals. Our framework measures students' (a) Reasoning about ethical decisions (e.g., by relying on deliberation vs. intuition), (b) Evaluation of their own ethical judgment and decision-making processes (e.g., confidence), (c) Goal priorities when making key decisions (e.g., how much they weigh societal consequences), (d) Awareness of moral issues (e.g., recognizing that a given situation has moral implications); (e) Information-seeking intentions (e.g., from diverse perspectives), and (f) Norm perceptions around ethical values in technology development (e.g., how tech developers normally prioritize different values).

A. Reasoning About Ethical Decisions

Despite common lay beliefs that moral judgments rely on deliberation, and philosophers' focus on moral deliberation as the key to changing moral judgment (e.g., [27]), growing empirical evidence suggests that intuition and emotion also play a key role in driving decision-making [28]. In some cases, deliberating about information and evidence can influence people's views (e.g., [29]), but many day-to-day moral judgments are driven by intuition instead [30], [31]. Some recent meta-analyses of ethics education have found that content focused on reasoning and "cognition" can systematically alter moral judgments (e.g., [32], [33]), but it is unclear from these studies which aspects of cognitive reasoning underlie the observed effects. How do tech ethics courses shift students' reliance on intuition and deliberation to make decisions, and how might these shifts relate to their moral awareness?

In a study looking at how ethics courses affect cognitive change, Oktar, Lerner, Malaviya, and Lombrozo [34] found that students taking a course in ethics changed their beliefs on key ethical controversies (e.g., abortion, eating meat), and that this change in ethical judgments was predicted by *lower* reported reliance on intuition over the course of the semester, though not by greater reported reliance on deliberation. These findings raise the fear of "ethics washing," especially when students are faced with difficult-to-anticipate social effects of tech: Could tech ethics courses lead students to deliberate more, but have no effect on their ethical judgments? Or might they shift judgments by lowering reliance on intuition instead? In our REGAIN framework, we measured students' self-reported reliance on intuition and deliberation when deciding about each scenario. For exploratory purposes, we also included scales validated in previous psychology research to measure people's overall willingness to change their minds (i.e., Actively Open-Minded Thinking; [35]), general preferences for intuition and deliberation [36], and attitudes about the importance of rationality[37].

B. Evaluating Ethical Reasoning Processes

In addition to measuring ethical reasoning, we also measured students' evaluations of their own reasoning processes. Reflections about one's own reasoning, or *metacognitive* evaluations [38], [39], can capture important aspects of reasoners' decision-making processes, such as confidence in their ability to make decisions and their appreciation for the limits of their understanding in the context of a complex and unpredictable system. Many existing strategies for teaching tech ethics focus on careful deliberation about the long-term and downstream consequences of technology development decisions (e.g., through role-play; [24], [40]). The goal is to help students think more about the macro-scale, structural consequences of individual actions within a complex system like human-computer interaction; this type of *systems thinking* is important for building more accurate representations of these complex systems [41], [42].

Foster-Hanson and Venkatagiri [26] suggested that interventions aimed at encouraging technologists to think more broadly about the systemic social effects of their actions (including interventions targeting moral imagination [43] or moral awareness [44]) could ultimately make students more hesitant to act. There is extensive literature on the importance of systems thinking for ethical technology development, such as reasoning about the system's affordances [45]. Thinking about interconnections across a sociotechnical system could help technologists confront its complexities and unpredictability, and heighten sensitivity to potential unforeseen consequences. For example, viewing technologies like vaccines or genetically modified foods as intervening in complex and unpredictable systems can make people more cautious about using those technologies [46], [47]. In a similar way, viewing technology development as complex and unpredictable could make designers more cautious and lead them to seek out additional

information [48] and consider diverse perspectives [44]. Such cognitive changes would achieve the most sought-after goals of development frameworks like Value Sensitive Design [49] . Can tech ethics courses convince students of system complexity, and thereby induce more cautious, perspective-seeking behavior? To this end, we asked students to rate the difficulty of making decisions about each scenario and the predictability of the long-term consequences.

Despite the importance of systems thinking, however, people are often motivated to justify their existing beliefs, and the process of deliberation can itself provide an opportunity for post-hoc rationalization [50], [51], [52] and even lead to overconfidence [53]—an open invitation for "ethics washing" [9]. To test whether tech ethics coursework shapes students' confidence, and how this might relate to their other metacognitive evaluations of their ethical reasoning processes, we measured students' confidence in their ability to make decisions about how to develop the products in each scenario.

C. Goal priorities

Human cognition is deeply social [54], yet the social aspects of tech ethics education may be the most difficult to measure and incorporate into coursework [24]. Focusing on the social nature of tech ethics includes recognizing that ethical decisions entail deciding *whose* ethics to prioritize, because different stakeholders may hold different—or even competing—goals and values.

In our REGAIN framework, we measured these value trade-offs directly by asking students to rate how much they valued four different types of goals: (i) the organization's goals, like revenue, efficiency, or meeting product deadlines, (ii) their own professional goals, like career advancement or supervisor approval, (iii) the end user's goals, like ease of use and flexibility, and (iv) society's goals, like maximizing fairness and minimizing social harm.

D. Awareness of Moral Issues

Many theories of moral decision making propose that moral behavior begins with moral awareness—i.e., recognizing that a given situation involves moral issues. From this perspective, people sometimes behave unethically simply because they have failed to identify the moral implications of a given situation and have therefore failed to initiate a process of moral reasoning. For example, [44] defined moral awareness (interchangeably called "ethical sensitivity") as understanding that one is in a morally relevant situation, and thereby perceiving the need to answer *what, who* and *how* questions about that situation: What actions are available, who (including oneself) would be affected by each, and how they will assess those actions/effects. Similarly, some researchers view ethical decision-making as the activation of "schemas" applicable to choice situations, where schemas might differ across cultures but still display the same general pattern of "development" across individuals [55].

Rest and colleagues theorized that decisionmakers have to first become aware of the moral implications of a situation before they can make a moral judgment about it. Nevertheless, most of the available tests of moral reasoning measure downstream moral reasoning and judgments, which can tell us little about the ability to anticipate problems or be aware of problematic situations. For example, one commonly used test in assessments of tech ethics education is the Defining Issues Tests (DIT and DIT-2), which classifies students' judgments into different stages of moral development [56], [57]. The Engineering and Science Issues Test (ESIT) similarly measures downstream moral judgments, rather than moral awareness itself [58]. In contrast, some tests have been designed to assess how students' awareness of moral issues, or ethical sensitivity, is impacted by coursework, including dental education (DEST; [59], short ethics modules for life sciences students (TESS; [60]), or ethics coursework for science and engineering students (TESSE; [61]).

The primary focus of our REGAIN framework was to measure aspects of moral decision-making that we expected would generalize across contexts and individuals, so we opted to use a set of questions developed to measure moral awareness in general [62]. An additional concern with existing tests is that the tests themselves may lead students to consider ethical issues when they would not otherwise have done so, which could provide limited insight into how students anticipate moral harms when making decisions about real-world scenarios in which relevant moral facts are not identified or labeled for them. For this reason, we first presented students with scenarios in which the ethical issues were not made explicit, and they answered a series of questions about the processes they would rely on to make decisions about these nonexplicit scenarios (see above). To capture the benefits of open-ended responses, respondents were also asked, "Which factors do you think need to be considered when making decisions about how to develop this product?" which were then coded for moral terms. Finally, after answering questions about two such nonexplicit scenarios, we asked students to rate their agreement (on a Likert scale) with three questions of moral awareness including "There are very important ethical aspects to this situation" and "This matter clearly does not involve ethics or moral issues," (reverse-scored; [62]). Students then repeated this process for two scenarios in which the ethical issues were made explicit. This allowed us to not only measure *whether* students were reasoning about the ethical implications of their decisions, but also to assess *how* their ethical sensitivity related to other key cognitive processes, because these processes had been assessed before the moral issues of the scenario were made explicit.

E. Information-seeking

When people are less confident in their ability to make decisions, they are more likely to seek out additional information before deciding [48], so we asked students to rate how likely they would be to seek additional information before deciding overall. Information-seeking in moral reasoning often means seeking out diverse perspectives [44], [56] so we also asked students whose information they would seek, including how likely they would be to consult with experts and how likely they would be to consult with people who have different perspectives and experiences.

F. Norm perceptions around ethical values

People tend to conform to the social norms within their groups [54], and ethics education should ultimately seek to shift the social norms around ethical behavior [26]. To assess students' perceptions of social norms in the tech industry, after

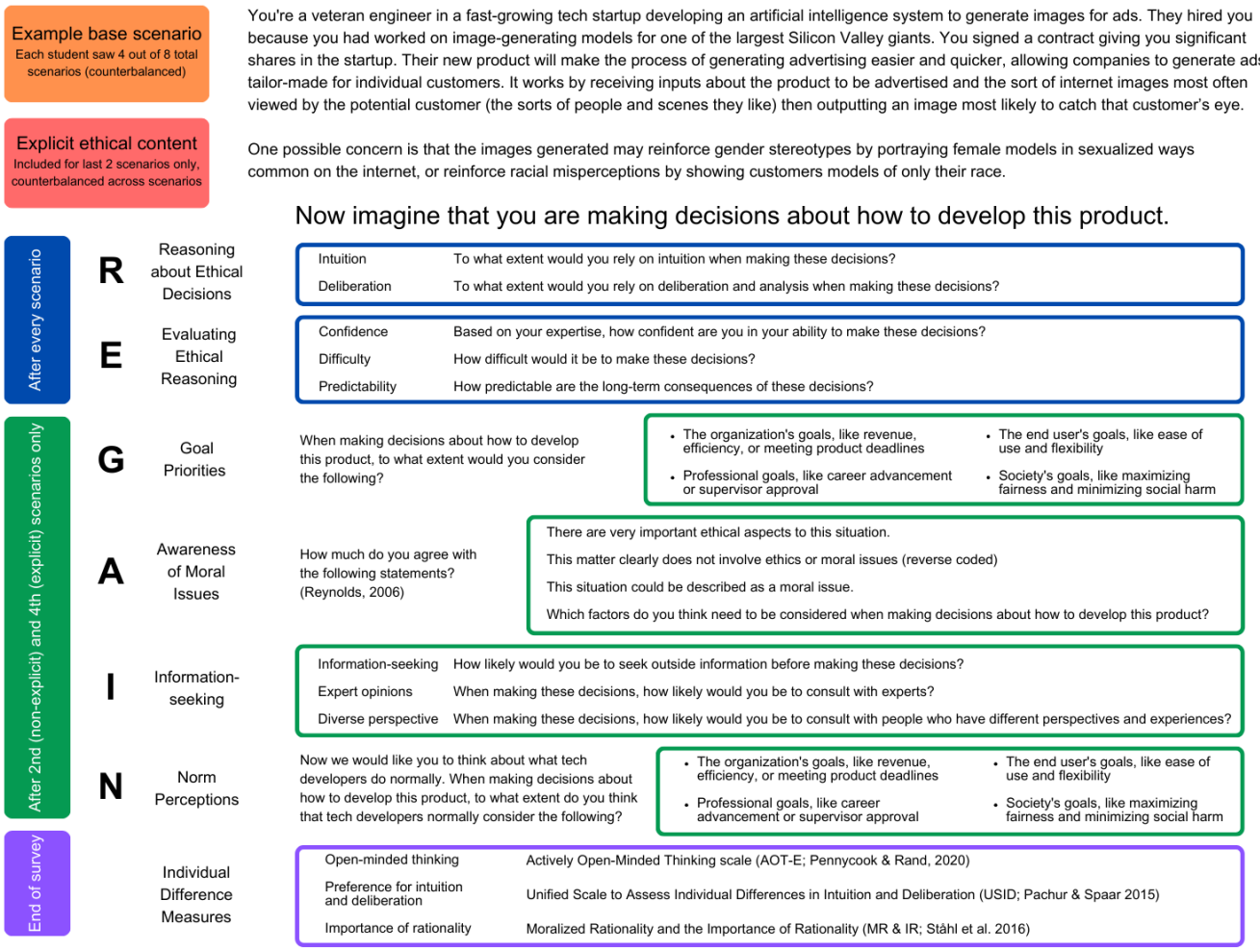

Fig. 1. General psychological mechanisms and the questions measuring them in the REGAIN framework.

students rated how they prioritized different types of goals, we also asked them to think about what tech developers normally do when making decisions about how to develop products like the ones in the scenarios. Students then rated how much they thought tech developers normally value the same four types of goals that they had previously rated for themselves (the organization's goals, their own professional goals, the end user's goals, and society's goals) when making decisions about technology design and development. See Fig. 1 for an example scenario and the set of questions students answered.

III. Empirical Study

To test our framework in practice, we conducted a study measuring the effects of an undergraduate course on technology ethics at a research university in the northeastern United States. The course enrolled 132 students in the spring semester of 2023 and had four main learning objectives: 1) improving moral deliberation through analysis of philosophical theories of ethics; 2) analysis of cases of social harm from tech, using theory to evaluate facts; 3) anticipation of these harms through role-playing real development cases; 4) awareness of organizational barriers to action through literature on behavioral economics and ethics.

The philosophical core of the class began with three theories commonly taught in ethics courses: Utilitarianism [63], Deontology [64], and Virtue Ethics [65], [66]. To give students a larger comparison set, the course also included Ethics of Care [67], [68], a philosophical analysis of Structural Justice theory (more common in courses on Science, Technology and Society; [69], [70]), and Computer and Information Ethics [71], [72]). Case studies of social harms included issues of loss of autonomy due to recommender algorithms [5], the difficulty of robot rule-following [73], [74], long-term threats of AI [75], racial and other serious biases in present AI and big data [2], threats to privacy from de-anonymization of data [3], [4], racial and other biases in visual recognition [76], [77], and others. The course also utilized role-plays called "Agile Ethics," intended to help students both to learn about the basic framework of Agile software development [78] and to experience the difficulty of making ethical decisions in low-information environments. To help assess the development framework they simulated, students also read literature about how complex systems cause ethical "blind spots" to emerge [79] and other relevant literature in organization psychology. These materials were also used to

assess other cases of institutional failures presented in lecture, in which whistleblowers struggled to sound the alarm (or never stepped forward; [80], [81]). To shed light on the cognitive processes of ethical decision making, and how they can be short-circuited in complex work environments, students also read psychological theory on the process of ethical awareness, judgment, intention and behavior [44].

A. Participants

All participants were undergraduate students at the same research-intensive university in the northeastern United States. We used two sets of participants: The test group was a subset of students enrolled in the course on technology ethics described above who chose to complete the study (N = 67). We also recruited a control group of students (N = 315) that was similar to the test group in terms of majors and demographic factors (interests, level in program, etc.). Participants received a $20 Amazon gift card for participating in the study. All participants were informed that their participation was voluntary, instructional staff would not know whether they completed the study, and participation would not affect their course grade[2].

We initially planned to compare pre- and post-course results for participants in the test and control groups using surveys distributed at the start and end of the semester. However, due to a smaller than expected number of participants in the test condition who completed both surveys (N = 33), we do not consider pre-test data further and here report only on the separately preregistered comparison between test and control condition participants within the larger sample of post-test responses. Nevertheless, exploratory analyses of individual difference measures among the portion of our sample for whom we had pre-test data revealed that students in the test condition were not significantly different from those in the control condition (with the only exception being that tech ethics students were in fact *less* open-minded than those control condition; see https://osf.io/nj29x/).

B. Procedure

Students in the study read different techno-moral scenarios [82]. To increase generalizability, each student read four out of eight total possible scenarios, counterbalanced across students. Each scenario asked the student to imagine developing a particular technology product (e.g., facial recognition software). For each scenario, one version explicitly described potential ethical concerns about the product, and a matched version did not explicitly describe any ethical concerns. Each student first read two scenarios without explicit ethical concerns, followed by two scenarios with explicit ethical concerns. The scenarios presented with and without explicitly ethical content were also counterbalanced across participants.

After reading each scenario, students were asked to imagine that they were making decisions about how to develop each product and were asked a series of questions, as described in Figure 1. For brevity, we report here the primary results included in our preregistration and key exploratory measures described above; analyses of the full set of measures are available in the Supplemental Online Material on the Open Science Framework, along with data and code, https://osf.io/nj29x/. Note that we asked explicitly ethical questions (questions about prioritizing ethical goals, social norms, and moral awareness) after the 2nd and 4th scenarios only, when students had already answered questions about their reasoning and evaluations for both nonexplicit scenarios. Presenting these explicitly moral questions last ensured that they would not influence students to reflect on the ethical content of the two nonexplicit scenarios if they had not already done so. We analyzed data in R version 4.4.2, using linear mixed models with the lme4 package including random intercepts per participant (for repeated measures) as well as counterbalance set and trial.

IV. Results

Overall, our data suggests that students who completed the tech ethics course showed higher moral awareness of potential tech harms compared to a control condition. Tech ethics students also differed in their reasoning strategies and metacognitive judgements, and they reported stronger intentions to seek diverse perspectives and prioritize their own goals as developers less than society's goals on the whole.

A. Reasoning About Ethical Decisions

Tech ethics students reported that they would rely more on deliberation when deciding about the scenarios (across types; M = 5.82, 95% CI [5.60, 6.04]) than students in the control condition (M = 5.43, 95% CI [5.27, 5.59]; $F(1, 378) = 11.36$, $p < .001$). Participants in the test condition also said they would rely *less* on intuition when deciding about the scenarios (across types; M = 3.54, 95% CI [3.15, 3.92]) than those in the control condition (M = 4.05, 95% CI [3.76, 4.34]; $F(1, 378) = 7.27$, p = .007). There were no main or interactive effects of scenario type on either measure.

To examine how these different cognitive processes relate to students' awareness of the moral implications of their decisions, we ran an exploratory linear mixed model analysis of students' moral awareness responses, with their reported reliance on intuition and deliberation as well as condition and scenario type included as predictors, and with random intercepts for each moral awareness question, counterbalancing condition, and participant. In addition to the previously observed main effects of condition ($F(1, 374) = 14.71$, $p < .001$) and type ($F(1, 1922) = 18.66$, $p < .001$), moral awareness across scenario types was predicted by both *more* reliance on deliberation ($\beta = 0.23$, $F(1, 1009) = 32.81$, $p < .001$) and *less* reliance on intuition ($\beta = -0.11$, $F(1, 748) = 15.62$, $p < .001$). This result supports the notion that tech ethics coursework can improve students' ability to anticipate possible ethical issues by leading them to both deliberate more, and to rely on their intuitive judgments less.

B. Evaluating Ethical Reasoning Processes

Participants' judgments of how difficult it would be to decide about the scenarios varied by a two-way interaction between condition and scenario type ($F(1, 1142) = 4.47$, p =

[2] In our preregistration, we planned to exclude participants who failed an attention check question at the start of the study; however, a larger than expected number of students failed this question (N = 39) so we opted to retain all participants for analysis.

.035), with students in the test condition rating explicit scenarios as more difficult than students in the control condition (pairwise contrast, p = .048) but no difference between conditions for nonexplicit scenarios (p = .894). Students in the test condition also viewed the scenarios (regardless of type) as less predictable (M = 3.49, 95% CI [3.18, 3.80]) than students in the control condition (M = 4.10, 95% CI [3.85, 4.35]; $F(1, 378) = 17.65$, $p < .001$). There were no significant main or interactive effects of condition or scenario type on students' confidence in their ability to make decisions. Thus, tech ethics students viewed the consequences of their decisions as more unpredictable overall, and they found it more difficult to make decisions about ethically explicit scenarios, but they were not more confident.

C. Goal Priorities

After the 2nd (i.e., nonexplicit) and 4th (i.e., explicit) scenarios, students indicated the extent to which (on 1-7 scales) they would consider four different types of goals: (a) the organization's goals, like revenue, efficiency, or meeting product deadlines, (b) their own professional goals, like career advancement or supervisor approval, (c) the end user's goals, like ease of use and flexibility, and (d) society's goals, like maximizing fairness and minimizing social harm. Tech ethics students said they would consider the goals of the organization less when making decisions (across types; M = 4.74, 95% CI [4.48, 5.00]) compared to students in the control condition (M = 5.13, 95% CI [4.95, 5.30]; main effect of condition, $F(1, 378) = 7.37$, p = .007). Participants in the test condition also said they would consider society's goals *more* when making decisions (across types; M = 5.72, 95% CI [5.35, 6.08]) than participants in the control condition (M = 5.36, 95% CI [5.00, 5.72]; main effect of condition, F (1, 377) = 5.35, p = .021); there were no main or interactive effects of scenario type on either measure. This result suggest that tech ethics education may help students expand the sphere of social value beyond the organization to society more broadly.

However, participants across conditions said they would consider the goals of the end user more in nonexplicit (M = 5.52, 95% CI [5.34, 5.71]) compared to explicit scenarios (M = 5.15, 95% CI [4.97, 5.34]; main effect of type, F (1, 381) = 17.43, $p <$ .001; there were no main or interactive effects of condition. That is, when the moral implications of a scenario were not explicit, all students were more likely to focus on the end user's goals, like ease of use and flexibility.

D. Awareness of Moral Issues

We assessed students' ethical sensitivity using three questions of general moral awareness (from Reynolds, 2006), which asked participants to rate agreement (from 1[Strongly Disagree] to 7 [Strongly Agree]) with three statements: (a) There are very important ethical aspects to this situation; (b) This matter clearly does not involve ethics or moral issues (reverse-coded); and (c) This situation could be described as a moral issue (Fig. 1). In our preregistration, we planned to collapse the three items into a composite moral awareness score for each scenario if the alpha was sufficiently high (> .7), however, the alpha did not reach this threshold (alpha = .55), so we did not create a composite. Rather, we analyzed responses on the 1-7 scales to all three questions using linear mixed models testing the main and interactive effects of group (test, control) and version

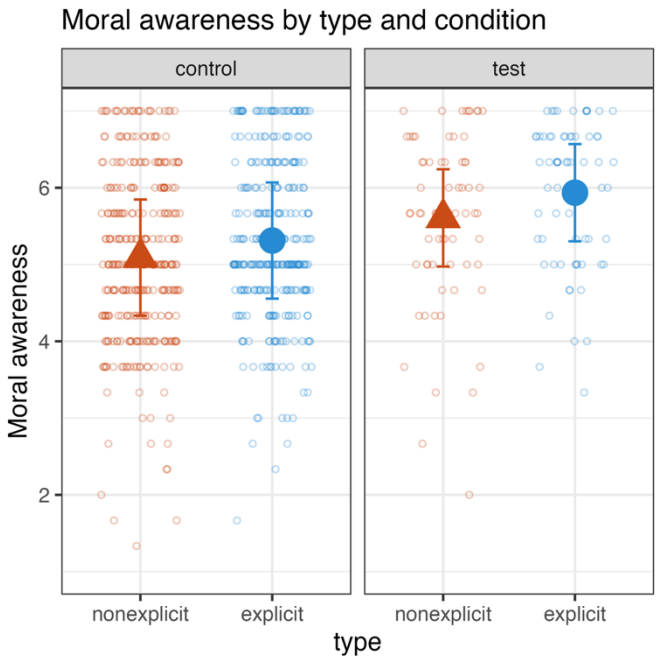

Fig. 2. Moral awareness responses by scenario type and condition. Large shapes are condition means with 95% Confidence Intervals; small circles are individual participant averages.

(explicit, nonexplicit), with random intercepts per question. As preregistered, we started out with maximal models, including random intercepts and slopes for each counterbalancing order and participant, but we dropped the random slopes because the models did not converge.

In line with our preregistered hypotheses, responses to moral awareness questions varied by both condition ($F(1, 378) = 22.47$, $p < .001$) and scenario type ($F(1, 1907) = 18.89$, $p < .001$). Participants in the test condition viewed the scenarios as more morally fraught overall (M = 5.77, 95% CI [5.12, 6.43]) than participants in the control condition did (M = 5.20, 95% CI [4.42, 5.98]; Fig. 2), and participants across both conditions viewed explicit scenarios as more morally fraught (M = 5.62, 95% CI [4.9, 6.35]) than nonexplicit scenarios (M = 5.35, 95% CI [4.62, 6.08]). However, contrary to our preregistered hypothesis, the two-way condition x scenario type interaction was not significant ($F(1, 1906) = 0.53$, p = .467). Students in the test condition were also significantly more likely to use ethical terms (i.e., referring to ethics, morals, justice, fairness, equality, bias, or racism; analyzed with a binomial linear mixed model) in their open-ended descriptions of the factors they thought should be considered when deciding about how to develop the products, regardless of scenario type ($X^2(1) = 5.08$, p = .024).

E. Information-seeking

Tech ethics students said they would seek out additional information before deciding (across both explicit and nonexplicit scenarios; M = 6.13, 95% CI [5.90, 6.37]) more than participants in the control condition (M = 5.72, 95% CI [5.60, 5.84]; $F(1, 380) = 10.38$, p = .001; there were no main or interactive effects of scenario type on this measure). Students' reports that they would consult experts before deciding, however, varied by a two-way interaction between condition and

scenario type ($F(1, 1142) = 4.53$, $p = .034$; subsumed main effect of condition, $F(1, 380) = 10.81$, $p = .001$): Students in the control condition said they would consult experts before deciding more for nonexplicit ($M = 5.91$, 95% CI [5.76, 6.07]) than explicit scenarios ($M = 5.71$, 95% CI [5.55, 5.86]; pairwise contrast, $p = .08$), but participants in the test condition were more likely overall to say they would consult experts about both types of scenarios (explicit: $M = 6.22$, 95% CI [5.97, 6.48]; nonexplicit: $M = 6.21$, 95% CI [5.95, 6.46]; pairwise contrast, $p = .89$).

Tech ethics students were also more likely overall than those in the control condition to say they would seek out diverse perspectives before deciding (across scenario types; $M = 5.92$, 95% CI [5.69, 6.17]) than students in the control condition ($M = 5.51$, 95% CI [5.40, 5.62]; $F(1, 379) = 9.36$, $p = .002$), but participants across both conditions said they would seek out diverse perspectives more before deciding about explicit scenarios ($M = 5.87$, 95% CI [5.72, 6.02]) than nonexplicit scenarios ($M = 5.57$, 95% CI [5.41, 5.72]; main effect of scenario type, $F(1, 1144) = 36.06$, $p < .001$; the two-way condition by scenario type interaction was not significant).

To examine how students' ethical decision making processes relate to their information-seeking, we again ran an exploratory linear mixed model analyzing students' reported likelihood of seeking additional information before deciding how to develop the product, with their reported reliance on intuition and deliberation, perceived difficulty and predictability, as well as condition and scenario type along included as predictors, and with random intercepts for counterbalancing condition and participant. In addition to the previously observed main effects of condition ($F(1, 369) = 4.58$, $p = .033$), information-seeking was also predicted by more reliance on deliberation ($\beta = 0.21$, $F(1, 2284) = 108.37$, $p < .001$), less reliance on intuition ($\beta = -0.11$, $F(1, 2160) = 52.44$, $p < .001$), and greater perceived difficulty of the decision ($\beta = 0.20$, $F(1, 2278) = 135.74$, $p < .001$). Neither scenario type nor perceived predictability were significant factors. One interpretation of this result is that tech ethics education might encourage students to seek out more information before deciding by leading them to deliberate more about the ethical implications of their decisions, to rely less on their own intuitive judgments, and to therefore evaluate the decision itself as more difficult to make on their own.

TABLE I. SUMMARY OF EMPIRICAL STUDY RESULTS

Process	Measures with significant effects		
	Effect of Condition	***Effect of Scenario***	***Interaction***
R	Intuition, Deliberation	None	None
E	Predictability	None	Difficulty
G	Organizational goals, Societal goals	End user's goals	None
A	Moral awareness, Open-ended codes	Moral awareness	None
I	Info-seeking, Experts, Diversity	Diversity	Experts
N	Professional goals	All goals	None

F. Norm perceptions around ethical values

To measure students' perceptions of social norms around ethical behavior in the tech industry, we then asked them to think about what tech developers do normally when making decisions about how to develop products like the one described in the scenario (also asked only after the 2nd and 4th scenarios). Students rated how much they thought tech developers normally consider each of the four goals they had previously rated for themselves.

Across conditions, students expected that developers who were making technology development decisions in scenarios where the ethical issues were not explicit (compared to explicit) would normally be more likely to consider the goals of the organization ($F(1, 381) = 5.75$, $p = .017$), their own professional goals ($F(1, 381) = 6.61$, $p = .011$), and the goals of the end user ($F(1, 381) = 11.75$, $p < .001$). Students in both conditions thought developers would normally be *less* likely in these nonexplicit scenarios to consider society's goals ($F(1, 381) = 11.75$, $p < .001$).

However, compared with participants in the control condition, tech ethics students thought that developers normally prioritize their own career goals more overall when making decisions (across scenario types; $F(1, 378) = 11.93$, $p < .001$). That is, one effect of tech ethics education (at least in this case) was to make students more pessimistic about the current social norms around ethical behavior in the tech industry.

V. DISCUSSION

In the current paper, we described a novel framework for measuring the cognitive processes shaped by tech ethics coursework, to shed light on which strategies might be most effective at shaping students' reasoning—and why. Key to this framework is a focus on the basic mechanisms of cognition and perception that we expect to be (a) susceptible to instruction, and (b) generalizable across contexts. Accordingly, the instruction in the tech ethics class we evaluated focused on teaching students to anticipate the downstream moral implications of technology development decisions. Our empirical study was intended as a proof of concept for the larger REGAIN framework, and the results provide preliminary support for this focus. We found that students in the course on tech ethics showed higher moral awareness about ethical considerations in technology decision-making: they both perceived more ethical danger when presented with techno-moral scenarios, and used more ethical terms in their free responses about those scenarios. However, we did not find significant differences between tech ethics students and those in the control condition on measures of moral awareness in nonexplicit scenarios (which would indicate that the students were better at anticipating harms that were not explicitly labeled for them). One possibility is that our sample size of tech ethics students was too small to detect an effect. It is also possible that a single course is not sufficient to make students with little real-world experience sufficiently aware of the ethical implications of their decisions. Future empirical research should test these different possibilities.

The primary novel contribution of the REGAIN framework, however, is testing how students' cognitive and metacognitive processes relate to their moral awareness. The finding that tech

ethics students reported that they would rely more on deliberation and less on intuition than did other students, and that their moral awareness was predicted by those very differences, suggests that our framework has promise. Furthermore, increased moral awareness was associated with significant change across several psychological determinants of ethical judgment and decision-making, including increased perceptions of difficulty deciding, decreased perception of the predictability of tech decisions, and higher expected information seeking (in general, and from diverse perspectives in particular). These results are promising, as past research has shown that changes in these cognitive processes can lead to downstream changes in ethical judgments [34] and decisions [83].

Our framework also helped to illuminate some of the social aspects of tech ethics education [24], [54]. For example, we measured students' perception of social norms around ethical decisions in the tech industry, and found that tech ethics students thought that developers normally prioritize their own career goals when making decisions more than students in the control condition did. One interpretation of this result is that the process of learning about the many social harms resulting from technology development, as well as ethical blind spots, institutional failures, and the difficulty of ethical decision making in complex work environments, may have left students feeling generally pessimistic about the current social norms surrounding ethical behavior among tech developers. However, when students rated their *own* goal priorities, tech ethics students reported valuing the goals of the organization less, and the goals of society more, than students in the control. These results paint a somewhat rosier picture—suggesting that tech ethics students hope to rewrite the social norms around ethical technology development in their own future careers.

Future research should measure how the processes our measurements target (motivated from prior psychology research) might relate not only to ethical reasoning, but to ethical behavior in practice. For example, do students who say they would prioritize society's goals, or seek information from people who hold diverse perspectives and experiences, actually do so when they become employees with bills to pay? Addressing these limitations may require not only devising new methods for measuring the effects of tech ethics coursework, but possibly also new methods for teaching tech ethics to begin with. Future research should also target ethics education among people who are actively making decisions about technology design and development, to examine what might happen when these students become employees in organizations whose incentive structures are at odds with those very ethical values they say they will uphold.

Our framework and study set the stage for further research that examines the consequences of ethics coursework. For instance, it is unclear which aspect(s) of the course were responsible for the changes in judgments. Examining the consequences of ethics education across multiple courses could enable factorial experiments that examine these aspects individually (for instance, by placing substantial weight on thought experiments in one group of classes, and very little in others). Similarly, though we only examined changes over the course of a semester, it is plausible that incorporating ethical content into curricula more deeply could result in more systematic changes. For instance, courses that have applied, internship- or project-based components could be used to examine the extent to which changes in responses to hypothetical scenarios predict changes in actual software development decisions. Conducting these studies would contribute not only to our understanding of the psychological processes underlying ethical change, but would also have direct implications for the design of effective (and efficient) tech ethics courses.

Our study had several important limitations. Most importantly, we only examined view change in one course in a northeastern United States university, and we do not know whether particular demographic (e.g., age) or contextual (e.g., instructor) characteristics played key roles in facilitating the effects we observed [84]. Second, we only examined view change in the context of eight scenarios; though they cover a broad range of actual ethical dilemmas that routinely arise in technology design and development, and students saw a counterbalanced four scenarios from a larger set of eight to improve generalizability, our scenarios could not possibly capture the span of possible scenarios and outcomes that students should ultimately consider. We do not know, therefore, the extent to which changes in students' moral awareness or reasoning would generalize to situations beyond those included in the current study, including real-world decision making. It is also important to note that students forget many aspects of their education, and we only examined view change over the course of a semester; future research should establish the extent to which the observed effects are long-lasting or temporally limited. Our results may also have been impacted by self-selection bias, both in terms of who elected to take the tech ethics course and who elected to participate in the study.

Finally, it is important to note that our research—and most similar research across moral psychology—focuses on the decisions and judgments of specific individuals. In line with our emphasis on the importance of evaluations of ethical reasoning, prioritizing social values and goals, and perceiving social norms among individuals, it is important for empirical research to start tackling ethical questions at the level of institutional structures [26], [85], [86], [87]. What enables groups to form institutional structures that place ethical considerations at the heart of their organization? Which kinds of educational practices can facilitate an understanding of the importance of these structures at the institutional level? Which individuals and timepoints might be the most advantageous for interventions to foster ethical norms change? Addressing these questions in future work—both theoretical and empirical—will help to center to social nature of human decision making.

ACKNOWLEDGMENT

We are grateful to the students who participated in this research and to Maya Malayiva for help with data collection.

REFERENCES

[1] A. Battista *et al.*, "Future of jobs report 2023," in *World Economic Forum*, Geneva, Switzerland, 2023. [Online]. Available: https://www.

[2] R. Benjamin, *Race After Technology: Abolitionist Tools for the New Jim Code*, 1st edition. Cambridge, UK ; Medford, MA: Polity, 2019.

[3] A. Narayanan and V. Shmatikov, "How To Break Anonymity of the Netflix Prize Dataset," Nov. 22, 2007, *arXiv*: arXiv:cs/0610105. doi: 10.48550/arXiv.cs/0610105.

[4] L. Sweeney, "Only You, Your Doctor, and Many Others May Know," *Technol. Sci.*, Sep. 2015, Accessed: Jan. 15, 2025. [Online]. Available: https://techscience.org/a/2015092903/

[5] J. Orlowski, *The Social Dilemma - More About the Film*, (2020). Accessed: Jan. 15, 2025. [Online Video]. Available: https://thesocialdilemma.com/the-film/

[6] H. Dai, K. L. Milkman, and J. Riis, "The fresh start effect: Temporal landmarks motivate aspirational behavior," *Manag. Sci.*, vol. 60, no. 10, pp. 2563–2582, 2014.

[7] M. J. Casañ, M. Alier, and A. Llorens, "Teaching Ethics and Sustainability to Informatics Engineering Students, An Almost 30 Years' Experience," *Sustainability*, vol. 12, no. 14, Art. no. 14, Jan. 2020, doi: 10.3390/su12145499.

[8] R. Ferreira and M. Y. Vardi, "Deep tech ethics: An approach to teaching social justice in computer science," in *Proceedings of the 52nd ACM Technical Symposium on Computer Science Education*, Mar. 2021, pp. 1041–1047.

[9] B. Green, "The contestation of tech ethics: A sociotechnical approach to technology ethics in practice," *J. Soc. Comput.*, vol. 2, no. 3, pp. 209–225, 2021.

[10] B. J. Grosz *et al.*, "Embedded EthiCS: integrating ethics across CS education," *Commun. ACM*, vol. 62, no. 8, pp. 54–61, 2019.

[11] R. Reich, M. Sahami, J. M. Weinstein, and H. Cohen, "Teaching computer ethics: A deeply multidisciplinary approach," in *Proceedings of the 51st ACM Technical Symposium on Computer Science Education*, Feb. 2020, pp. 296–302.

[12] J. Weinstein, R. Reich, and M. Sahami, *System error: Where big tech went wrong and how we can reboot*. Hachette UK, 2021.

[13] "Culture of Disengagement in Engineering Education? - Erin A. Cech, 2014." Accessed: Feb. 10, 2025. [Online]. Available: https://journals.sagepub.com/doi/full/10.1177/0162243913504305

[14] M. Bouville, "On Using Ethical Theories to Teach Engineering Ethics," *Sci. Eng. Ethics*, vol. 14, no. 1, pp. 111–120, Mar. 2008, doi: 10.1007/s11948-007-9034-5.

[15] E. Burton, J. Goldsmith, S. Koenig, B. Kuipers, N. Mattei, and T. Walsh, "Ethical Considerations in Artificial Intelligence Courses," *AI Mag.*, vol. 38, no. 2, Art. no. 2, Jul. 2017, doi: 10.1609/aimag.v38i2.2731.

[16] A. Colby and W. M. Sullivan, "Ethics teaching in undergraduate engineering education," *J. Eng. Educ.*, vol. 97, no. 3, pp. 327–338, 2008.

[17] C. Fiesler, N. Garrett, and N. Beard, "February). What do we teach when we teach tech ethics? A syllabi analysis," in *Proceedings of the 51st ACM technical symposium on computer science education*, 2020, pp. 289–295.

[18] J. L. Hess and G. Fore, "A systematic literature review of US engineering ethics interventions," *Sci. Eng. Ethics*, vol. 24, pp. 551–583, 2018.

[19] D. Horton, D. Liu, S. A. McIlraith, S. Coyne, and N. Wang, "Do Embedded Ethics Modules Have Impact Beyond the Classroom?," in *Proceedings of the 55th ACM Technical Symposium on Computer Science Education V. 1*, in SIGCSE 2024. New York, NY, USA: Association for Computing Machinery, Mar. 2024, pp. 533–539. doi: 10.1145/3626252.3630834.

[20] A. Monge Roffarello, K. Lukoff, and L. Russis, "Defining and identifying attention capture deceptive designs in digital interfaces," in *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, Apr. 2023, pp. 1–19.

[21] R. Botsman, "Tech Leaders Can Do More to Avoid Unintended Consequences," *Wired*, May 24, 2022. Accessed: Jan. 21, 2025. [Online]. Available: https://www.wired.com/story/technology-unintended-consequences/

[22] K. Do, R. Y. Pang, J. Jiang, and K. Reinecke, "That's important, but...": How Computer Science Researchers Anticipate Unintended Consequences of Their Research Innovations," in *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, Apr. 2023, pp. 1–16.

[23] N. Parvin and A. Pollock, "Unintended by design: on the political Uses of 'Unintended consequences,'" *Engag. Sci. Technol. Soc.*, vol. 6, pp. 320–327, 2020.

[24] A. Johri and A. Hingle, "Learning to link micro, meso, and macro ethical concerns through role-play discussions," in *2022 IEEE Frontiers in Education Conference (FIE*, IEEE, Oct. 2022, pp. 1–8.

[25] S. Kumar and M. Levis, "Reengineering ethics education for deeper student engagement through the creation of roleplaying and decision-making games [WIP Paper, Student Experiences," in *2023 ASEE Annual Conference & Exposition*, Jun. 2023.

[26] E. Foster-Hanson and S. Venkatagiri, "Promoting Ethical Technology Design Practices by Leveraging Human Psychology," in *Companion Publication of the 2024 ACM Designing Interactive Systems Conference*, Jul. 2024, pp. 157–161.

[27] D. Parfit, *On What Matters*. OUP Oxford, 2011.

[28] J. Greene, *Moral Tribes: Emotion, Reason, and the Gap Between Us and Them*. Penguin, 2014.

[29] Z. Horne, D. Powell, and J. Hummel, "A Single Counterexample Leads to Moral Belief Revision," *Cogn. Sci.*, vol. 39, no. 8, pp. 1950–1964, 2015, doi: 10.1111/cogs.12223.

[30] J. Haidt, "The emotional dog and its rational tail: A social intuitionist approach to moral judgment," *Psychol. Rev.*, vol. 108, no. 4, pp. 814–834, 2001, doi: 10.1037/0033-295X.108.4.814.

[31] J. Herec *et al.*, "Reflection and Reasoning in Moral Judgment: Two Preregistered Replications of Paxton, Ungar, and Greene (2012)," *Cogn. Sci.*, vol. 46, no. 7, p. e13168, 2022, doi: 10.1111/cogs.13168.

[32] E. P. Waples, A. L. Antes, S. T. Murphy, S. Connelly, and M. D. Mumford, "A Meta-Analytic Investigation of Business Ethics Instruction," *J. Bus. Ethics*, vol. 87, no. 1, pp. 133–151, Jun. 2009, doi: 10.1007/s10551-008-9875-0.

[33] L. L. Watts, K. E. Medeiros, T. J. Mulhearn, L. M. Steele, S. Connelly, and M. D. Mumford, "Are Ethics Training Programs Improving? A Meta-Analytic Review of Past and Present Ethics Instruction in the Sciences," *Ethics Behav.*, vol. 27, no. 5, pp. 351–384, Jul. 2017, doi: 10.1080/10508422.2016.1182025.

[34] K. Oktar, A. Lerner, M. Malaviya, and T. Lombrozo, "Philosophy instruction changes views on moral controversies by decreasing reliance on intuition," *Cognition*, vol. 236, p. 105434, 2023.

[35] G. Pennycook, J. A. Cheyne, D. J. Koehler, and J. A. Fugelsang, "On the belief that beliefs should change according to evidence: Implications for conspiratorial, moral, paranormal, political, religious, and science beliefs," *Judgm. Decis. Mak.*, vol. 15, no. 4, pp. 476–498, 2020.

[36] T. Pachur and M. Spaar, "Domain-specific preferences for intuition and deliberation in decision making," *J. Appl. Res. Mem. Cogn.*, vol. 4, no. 3, pp. 303–311, 2015.

[37] T. Ståhl, M. P. Zaal, and L. J. Skitka, "Moralized rationality: Relying on logic and evidence in the formation and evaluation of belief can be seen as a moral issue," *PloS One*, vol. 11, no. 11, p. 0166332, 2016.

[38] R. Cheruvalath, "Does studying 'ethics' improve engineering students' meta-moral cognitive skills?," *Sci. Eng. Ethics*, vol. 25, no. 2, pp. 583–596, 2019.

[39] P. R. Pintrich, "The role of metacognitive knowledge in learning, teaching, and assessing," *Theory Pract.*, vol. 41, no. 4, pp. 219–225, 2002.

[40] A. Hingle, H. Rangwala, A. Johri, and A. Monea, "Using role-plays to improve ethical understanding of algorithms among computing students," in *2021 IEEE Frontiers in education conference (FIE*, IEEE, Oct. 2021, pp. 1–7.

[41] D. Cabrera and L. Cabrera, "What Is Systems Thinking?," in *Learning, Design, and Technology*, J. M. Spector, B. B. Lockee, and M. D. Childress, Eds., Cham: Springer International Publishing, 2023, pp. 1495–1522. doi: 10.1007/978-3-319-17461-7_100.

[42] B. Richmond, "Systems thinking: Critical thinking skills for the 1990s and beyond," *Syst. Dyn. Rev.*, vol. 9, no. 2, pp. 113–133, 1993, doi: 10.1002/sdr.4260090203.

[43] P. H. Werhane, *Moral Imagination and Management Decision-making*. Oxford University Press, 1999.

[44] J. Rest, M. Bebeau, and J. Volker, "An Overview of the Psychology of Morality," in *Moral Development: Advances in Research and Theory*, New York, NY: Praeger, 1986, pp. 1–27.

[45] J. L. Davis, *How Artifacts Afford: The Power and Politics of Everyday Things*. MIT Press, 2020.

[46] G. D. Salali and M. S. Uysal, "COVID-19 vaccine hesitancy is associated with beliefs on the origin of the novel coronavirus in the UK and Turkey," *Psychol. Med.*, vol. 52, no. 15, pp. 3750–3752, 2022.

[47] P. Rozin *et al.*, "Preference for natural: instrumental and ideational/moral motivations, and the contrast between foods and medicines," *Appetite*, vol. 43, no. 2, pp. 147-154 , 2004.

[48] K. Desender, A. Boldt, and N. Yeung, "Subjective confidence predicts information seeking in decision making," *Psychol. Sci.*, vol. 29, no. 5, pp. 761–778, 2018.

[49] B. Friedman and D. G. Hendry, *Value Sensitive Design: Shaping Technology with Moral Imagination*. MIT Press, 2019.

[50] F. Cushman, "Rationalization is rational," *Behav. Brain Sci.*, vol. 43, p. e28, Jan. 2020, doi: 10.1017/S0140525X19001730.

[51] E. Schwitzgebel and F. Cushman, "Philosophers' biased judgments persist despite training, expertise and reflection," *Cognition*, vol. 141, pp. 127–137, 2015.

[52] E. L. Uhlmann, D. A. Pizarro, D. Tannenbaum, and P. H. Ditto, "The motivated use of moral principles," *Judgm. Decis. Mak.*, vol. 4, no. 6, pp. 479–491, 2009.

[53] M. L. Stanley, A. M. Dougherty, B. W. Yang, P. Henne, and F. Brigard, "Reasons probably won't change your mind: The role of reasons in revising moral decisions," *J. Exp. Psychol. Gen.*, vol. 147, no. 7, pp. 962–987, 2018.

[54] R. B. Cialdini and J. N. Goldstein, "Social influence: Compliance and conformity," *Annu. Rev. Psychol.*, vol. 55, pp. 591–621, 2004.

[55] S. J. Thoma, "Research on the Defining Issues Test," in *Handbook of Moral Development*, Marwah, NJ: Lawrence Erlbaum Associates, 2006, pp. 67–91.

[56] L. Kohlberg, "Moral stages and moralization: The cognitive-developmental approach," in *Moral development and behavior: Theory, research and social issues*, T. Lickona, Ed., Rinehart and Winston, 1976, pp. 31–53.

[57] J. R. Rest, D. Narvaez, S. J. Thoma, and M. J. Bebeau, "DIT2: Devising and testing a revised instrument of moral judgment," *J. Educ. Psychol.*, vol. 91, no. 4, pp. 644–659, 1999, doi: 10.1037/0022-0663.91.4.644.

[58] J. Borenstein, M. J. Drake, R. Kirkman, and J. L. Swann, "The Engineering and Science Issues Test (ESIT): A Discipline-Specific Approach to Assessing Moral Judgment," *Sci. Eng. Ethics*, vol. 16, no. 2, pp. 387–407, Jun. 2010, doi: 10.1007/s11948-009-9148-z.

[59] M. Bebeau, Rest, and C. Yamoor, "Measuring dental students' ethical sensitivity," *J. Dent. Educ.*, vol. 49, no. 4, pp. 225–235, Apr. 1985, doi: 10.1002/j.0022-0337.1985.49.4.tb01874.x.

[60] H. Clarkeburn, "A Test for Ethical Sensitivity in Science," *J. Moral Educ.*, vol. 31, no. 4, pp. 439–453, Dec. 2002, doi: 10.1080/0305724022000029662.

[61] J. Borenstein, M. Drake, R. Kirkman, and J. Swann, "The Test Of Ethical Sensitivity In Science And Engineering (Tesse): A Discipline Specific Assessment Tool For Awareness Of Ethical Issues," in *2008 Annual Conference & Exposition Proceedings*, Pittsburgh, Pennsylvania: ASEE Conferences, Jun. 2008, p. 13.1270.1-13.1270.10. doi: 10.18260/1-2--3253.

[62] S. J. Reynolds, "Moral awareness and ethical predispositions: investigating the role of individual differences in the recognition of moral issues," *J. Appl. Psychol.*, vol. 91, no. 1, p. 233, 2006.

[63] K. de Lazari-Radek and P. Singer, *Utilitarianism: A Very Short Introduction*. in Very Short Introductions. Oxford, New York: Oxford University Press, 2017.

[64] C. M. Korsgaard, *Creating the Kingdom of Ends*. Cambridge: Cambridge University Press, 1996. doi: 10.1017/CBO9781139174503.

[65] J. Annas, *Intelligent Virtue*, 1st edition. Oxford: Oxford University Press, 2011.

[66] S. Vallor, *Technology and the Virtues: A Philosophical Guide to a Future Worth Wanting*. Oxford, New York: Oxford University Press, 2016.

[67] C. Gilligan, *In a Different Voice*. Cambridge, Mass.: Harvard University Press, 1982. Accessed: Jan. 15, 2025. [Online]. Available: https://www.hup.harvard.edu/books/9780674970960

[68] J. C. Tronto, "An Ethic of Care," *Gener. J. Am. Soc. Aging*, vol. 22, no. 3, pp. 15–20, 1998.

[69] S. Haslanger, "Systemic and Structural Injustice: Is There a Difference?," *Philosophy*, vol. 98, no. 1, pp. 1–27, Jan. 2023, doi: 10.1017/S0031819122000353.

[70] K. Crenshaw, "Demarginalizing the Intersection of Race and Sex: A Black Feminist Critique of Antidiscrimination Doctrine, Feminist Theory and Antiracist Politics," *Univ. Chic. Leg. Forum*, vol. 1989, no. 1, pp. 139–167, 1989.

[71] J. Moor, "The Nature, Importance, and Difficulty of Machine Ethics," *IEEE Intell. Syst.*, vol. 21, pp. 18–21, 2006.

[72] L. Floridi, *The Ethics of Information*. Oxford, New York: Oxford University Press, 2013.

[73] I. Asimov, *I, Robot*. New York, NY: Gnome Press, Inc., 1950.

[74] D. Howard and I. Muntean, "A Minimalist Model of the Artificial Autonomous Moral Agent (AAMA)," *AAAI Spring Symp. Ser.*, 2016, [Online]. Available: https://cdn.aaai.org/ocs/12760/12760-56146-1-PB.pdf

[75] W. MacAskill, *What We Owe the Future*. Basic Books, 2022. Accessed: Jan. 15, 2025. [Online]. Available: https://www.hachettebookgroup.com/titles/william-macaskill/what-we-owe-the-future/9781541618633/?lens=basic-books

[76] J. Buolamwini and T. Gebru, "Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification," in *Proceedings of the 1st Conference on Fairness, Accountability and Transparency*, Proceedings of Machine Learning Research, Jan. 2018, pp. 77–91. Accessed: Jan. 15, 2025. [Online]. Available: https://proceedings.mlr.press/v81/buolamwini18a.html

[77] N. Meister, D. Zhao, A. Wang, V. V. Ramaswamy, R. Fong, and O. Russakovsky, "Gender Artifacts in Visual Datasets," Sep. 18, 2023, *arXiv*: arXiv:2206.09191. doi: 10.48550/arXiv.2206.09191.

[78] D. Rigby, S. Elk, and S. Berez, *Doing Agile Right: Transformation Without Chaos*, Illustrated edition. Boston: Harvard Business Review Press, 2020.

[79] M. H. Bazerman and A. E. Tenbrunsel, *Blind Spots: Why We Fail to Do What's Right and What to Do about It*, 1st edition. Princeton: Princeton University Press, 2011.

[80] J. Carreyrou, "Hot Startup Theranos Has Struggled With Its Blood-Test Technology - WSJ," *Wall Street Journal*, New York, NY, Oct. 16, 2015. Accessed: Jan. 15, 2025. [Online]. Available: https://www.wsj.com/articles/theranos-has-struggled-with-blood-tests-1444881901

[81] M. Goldstein, "A former Goldman Sachs banker says the plot to loot a Malaysian sovereign wealth fund was laid out in 2012.," *The New York Times*, New York, NY, Feb. 16, 2022. Accessed: Jan. 15, 2025. [Online]. Available: https://www.nytimes.com/2022/02/16/business/1mdb-goldman-sachs-roger-ng.html

[82] M. Boenink, T. Swierstra, and D. Stemerding, "Anticipating the Interaction between Technology and Morality: A Scenario Study of Experimenting with Humans in Bionanotechnology," *Stud. Ethics Law Technol.*, vol. 4, no. 2, Aug. 2010, doi: 10.2202/1941-6008.1098.

[83] E. Schwitzgebel, B. Cokelet, and P. Singer, "Do ethics classes influence student behavior? Case study: Teaching the ethics of eating meat," *Cognition*, vol. 203, p. 104397, 2020.

[84] J. Henrich, S. J. Heine, and A. Norenzayan, "The weirdest people in the world?," *Behav. Brain Sci.*, vol. 33, no. 2–3, pp. 61–83, 2010.

[85] S. J. Ali, A. Christin, A. Smart, and R. Katila, "Walking the Walk of AI Ethics: Organizational Challenges and the Individualization of Risk among Ethics Entrepreneurs," in *2023 ACM Conference on Fairness, Accountability, and Transparency*, Chicago IL USA: ACM, Jun. 2023, pp. 217–226. doi: 10.1145/3593013.3593990.

[86] N. Chater and G. Loewenstein, "Where next for behavioral public policy?," *Behav. Brain Sci.*, vol. 46, 2023.

[87] A. Madva, M. Brownstein, and D. Kelly, "It's always both: Changing individuals requires changing systems and changing systems requires changing individuals." 2023.