Augmenting Morality through Ethics Education: the ACTWith model

Abstract:

Recently in this journal, Jessica Morley and colleagues (AI & SOC 2023 38:411–423) review AI ethics and education, suggesting that a cultural shift is necessary in order to prepare students for their responsibilities in developing technology infrastructure that should shape ways of life for many generations. Current AI ethics guidelines are abstract and difficult to implement as practical moral concerns proliferate. They call for improvements in ethics course design, focusing on real-world cases and perspective-taking tools to immerse students in challenging situations with relatable personal, social and environmental implications. The present paper considers the ACTWith model of moral cognition for use in such training contexts. Morley and colleagues' paper and contemporary literature are reviewed. Whistleblowing as a source for relatable case studies is discussed, and the purpose of ethics education considered. The ACTWith model is introduced, and correlations with contemporaries noted. The ACTWith model has been successfully employed in a broad range of technology ethics courses in the Netherlands and South Korea. Illustrations are offered. Correlations with contemporary efforts are drawn in the context of software engineering. Summary discussion considers strengths and limitations of the ACTWith program. Overall, the paper emphasizes training high-level moral reasoning through guided inquiry education informing lifetime professional practice, and indicates an outstanding need for teachers able to do the job.

Keywords: engineering ethics education; empathy training; case study; metacognition; moral imagination

1. Introduction

Jessica Morley et al. (2023) review ethics education in AI engineering and design. They suggest that changes must be made to prepare students for their roles in establishing technological infrastructure for the coming generations. Contemporary education is inadequate. How can ethics be taken seriously by engineers and designers of technologies, when ethics education remains derivative of focal interests? They argue that practitioners require support beyond abstract guidelines, that ethics education must be "practically applicable to real-world ethical decisions" preparing students to "think through potential future scenarios" affecting diverse (and divergent) stakeholders (p 412). Though such an approach remains underdetermined, they consider recent progress in making most-common ethics principles accessible with "ethics as a service" (Morley et al. 2021) and offer recommendations.

Morley and colleagues (2023) survey AI practitioners about awareness of ethical issues and motivations to act on them. They ask about knowledge of and motivation to embed ethical principles into daily practice, about what hinders action from such principles, and about perceived needs for support in overcoming such obstacles. Respondents recognized that incorporating ethics into design "improves social impact" but comes with costs. Ethics is important for business. The focus is on avoiding risks of harm and to avoid contravention of conventional law, as reflected in focuses on privacy and security, attention to which reinforces consumer trust and company reputation (results summarized in Morley et al., 2023, fig. 4). Relatively few respondents were sensitive to deeper ethical concerns e.g. justice, autonomy, dignity (see figs. 1, 2). Especially worrisome given these results, most reported confidence in personal ability to design pro-ethical products (fig. 5), while who or what bears responsibility for ensuring that this happens remains unclear (fig. 6, discussion section 3.2).

One question is where to lay blame for less than ethical outcomes i.e. on poorly performing algorithms, their morally deficient designers, or the designers of those designers, e.g. their ethics teachers. Better tools are needed. Section 3.3 (Morley et al. 2023, Fig. 7) addresses gaps between available and necessary resources. Respondents need ethics resources which are "enforceable ... relatable and actionable" (p 417). Respondents "were particularly keen to stress the limitations of high-level principle-based frameworks" in informing pro-ethical design (p. 416). They complained that such frameworks appear "tick-boxy" in contrast to ethical practice which

seems grounded in context and practitioner. This tick-boxy nature encourages a sort of ethics-washing, with practitioners shopping-around for theories which, if not strictly endorsing divergent aims, do not forbid them (cf. Floridi 2019; Morley et al. 2020).

The focus is not on how ethics might inform design to do social good (Sect. 4). The aim is minimizing costly liabilities. As to where limited liability and social good intersect, Morley and colleagues (2023) emphasize that conventional law is an inadequate guide. "AI products that are merely legally compliant will not necessarily be ethically justifiable." (p 418) Even if conventional laws were adequate moral guides, legal constraints lag behind technological innovation. Citing Floridi (2018), they note that lawmakers often act only after prior "poor-decisions" deliver morally objectionable consequences. They argue for resources necessary to avoid such consequences.

Moral autonomy might not be the best business model, but if we outsource ethics to politicians and wait for regulation, it may be too late. Some things, regulation cannot fix. Rather, it is incumbent on practitioners to see the way through to right ends from the start. To train such an ability, however, Morley and colleagues (2023) conclude that a "cultural shift" is necessary. They stress that ethics courses must be mandated for all students of "data science, computer science, engineering, etc." with a focus on practical ethics "such as the use of empathy exercises" while training critical thinking skills (Sect. 4.1, p 418). Ethics must be made "relatable", challenging teachers to work past checklists of abstract principles to focus on preparing practitioners to respond to contemporary challenges.

Surveying the landscape in engineering ethics education reveals a broad migration in the direction of Morley and colleagues' recommended shift. Trbušić (2020) reports that students are more interested in simple skills than "understanding how social systems function and how engineers can or should function within the social system" (p 237). Yet, lacking such a "big picture" view and concomitant "social imagination"(unable to think through potential future scenarios; note correlations with Google team reports in section 4 of this paper), "engineers fail to promote their own ideas and goals to society because they do not possess sufficient knowledge of the social aspect of their engineering activities" (p 228). Trbušić stresses that "mere rules are not enough" and recommends integration of social and ethical training such as case-based critical thinking and communication into engineering curricula, advising that content is presented in ways that overcome student preferences for immediately applicable information. In the end, Trbušić recognizes enculturated bias against "soft" skills, notes organizational resistance slowing necessary changes to adequately train them, and encourages educators to find ways to speed the process.

Wang, Zhang and Zhu (2015) share concerns that rules are not enough. They find changeless ethical principles inadequate for dealing with new problems, and consider education involving interpretation, operation and dialogue as a model for professional practice. Marten Conlon and Bowe (2021a) advise that education must change through first understanding "what we are" in their focal sense of professional engineers embedded in society with distributed responsibilities, before "reorienting" engineering education "*for*" ethics concerning such a professional identity. Martin, Conlon and Bowe (2021b) echo needs to improve ethics education through "greater institutional or departmental support" (Sect. 4.4.4). They suggest that accreditation bodies may facilitate local networking around practical ethical issues, eventuating in guidelines for Engineering ethics programs (cf. Floridi 2018). Martin and colleagues (2023) propose that contemporary institutions teaching technology ethics consider obligations in terms of Scientific, Professional, Civic, Legal, and Intra-/Inter-generational, with the idea being that responsibility is the common factor (section "Articulating the Responsibility of Technological Universities").

Teaching responsibility is fundamental to applied ethics education. Wolfand et al. (2022) report that first year students were more likely to agree that engineers have the skills, knowledge and responsibility to solve social problems after an introductory course considering ethical issues, in this case with plastic pollution, pointing to motivational potential of ethics education. Shukla and colleagues (2023) review ongoing efforts at a "holistic" multidisciplinary engineering education in India involving critical thinking, problem-solving, communication, collaboration and research skills. They emphasize "universal human values" and personal integrity, "to create an environment where persons of integrity practice the morals that have been imparted to create a progressive world" i.e. not merely reinforce established patterns. They stress the need for "adequate faculty training"

to promote multi-disciplinary competencies coupled with an ability to "create interest" in students. Overall, they recommend a "T-shaped" curriculum reflecting the scope of social implications orthogonal to in-depth technical education, and recommend increased faculty capable of supporting such a curriculum (discussion p 142).

Chan and Lee (2021) also recognize a need for adequate faculty. They consider teaching generic "transversal" or "21st Century" competencies such as communication and problem-solving integrated into technical studies. They recognize global movement in that direction, noting mixed results, regional variation and institutional resistance. Focused on Hong Kong, they report that resources are predominantly allocated for research, demotivating educational development. They surveyed engineering faculty including program directors and teaching award winners about generic skill training in current course curricula, who reported some attention from a motivated "handful" of individual teachers, but who operated independently. All reported using problem-based learning with only indirect assessment of generic skill development. Most appeared unfamiliar with university goals for professional development, with "affective competencies" including advocacy for the disadvantaged, justice, and improving human conditions completely neglected. Chan and Lee note obstacles to implementing relevant skill training into course learning objectives. They recognize that few teachers can train such development, fewer want to, and no one seems able to assess competencies, regardless. In the end, allocated resources do not guarantee ability to inspire interest in, and do not motivate support for, a holistic engineering ethics education. Chan and Lee recommend that faculty develop competencies to train target skills, that curricula explicitly incorporate associated learning goals, and that academic leadership "seek improvement" in support of necessary change.

Morley and colleagues (2023) specify two ingredients for their cultural shift in ethics education. One requirement is empathy training including tools for perspective taking in order to apply ethical theory commensurate with personal values in professional practice. The other is ethical theory immediately accessible to non-specialists in philosophical ethics. The idea is to bridge gaps from ethical principles to social reform bottom-up, practitioner first. They propose the Digital Catapult Ethics Framework (DCEF) "to define and translate, transparently and contextually, high-level ethical principles into practice" (Box 1, p 419). The DCEF describes four levels, beginning with Floridi's (2018) unifying moral principles including beneficence, non-maleficence, autonomy, justice, and explicability at the first level of individual agent, principles as they bear on social concerns at the second level, Habermas' (2018) legitimization of governing principles through open access to information and free political discourse (ideally open to influence from marginal interests) at the third level, and finally purposeful development of technology to support such a practical pro-social ethical framework at the fourth level.

The next section considers empathy training through case-studies encouraging perspective-taking. The focus is on Floridi's first level, i.e. individual access to principles like justice and ability to explain why they are important. Section 3 introduces the ACTWith model as a potential contribution to a multi-layered, multidisciplinary, holistic ethics education. Section 4 concludes recognizing teachers able to support such a curriculum.

2. Case-based empathy training

The aim is to train designers and engineers of socially impactful technologies to discriminate what can be done from what should be done, and to do what should be done without defaulting to less. Empathy training and increased ethical awareness are fundamental to such efforts.

For better AI systems, Ball and Koliousis (2023) argue for "philosopher engineers" able to integrate ethical concerns throughout the design process in part by overcoming the "communicative impasse" between disciplines through multidisciplinary education. Trbušić (2020) reports that courses designed to train communication skills and "social imagination" (the ability to think-through potential scenarios, per Morley et al. 2023) help to assuage feelings of anxiety in students concerned for how their lives and work may contribute to society. Walther and colleagues (2019) emphasize that empathy training reveals personal values, affording students opportunity for reflective self-appraisal as individuals and as engineers. Howcraft and Mercer (2022) surveyed engineering and design faculty about empathy in teaching and in the profession. Ascribed importance ranged from moderate to extreme, with most self-identifying female

respondents ascribing very important, and the greatest portion of males ascribing extreme importance, to empathy teaching. However, how empathy should be taught remains unclear.

Walther and colleagues (2017) developed a model for empathy involving an individual level including perspective-taking, social level involving orientation to value pluralism, and holistic level representing service to self, others and the environment in the context of ethical principles including dignity (compare to the first three steps of the DCEF model). Sanz and colleagues (2023) propose the SPC model based on Walther et al.'s (2017), involving empathy skills, professional performance, and global citizenship. They consider that empathy involves affective and cognitive components which can be taught, including ethical awareness, perspective-taking and mode-switching (roughly, switching between bottom-up affect-driven and top-down analytic/intentional dynamics). Professional performance involves civic-mindedness, being open to diverse world views and practices, respect and responsibility. Global citizenship involves cooperation for the common good prioritizing human dignity and a healthy natural environment (again, compare with the DCEF). They stress that education must integrate all levels, with students learning "to value the commitment to social transformation in the different contexts in which they participate: classroom, educational institution, and local-global communities" (discussion, Sanz et al. 2023, Sect. 2). In effect, they propose a bottom-up practitioner-first framework to put high-level principles into practice and effect social change such as Morley and colleagues' cultural shift.

Empathy on Sanz and colleagues' (2023) account is a "transversal competence" unifying the framework. Along with empathy skills, practitioners require a "willingness to want to act" for pro-ethical social transformation (discussion, Sanz et al. 2023, Sect. 4). They propose a six-step course structure to "guide" focal training of ethical awareness and perspective-taking skills. They adapt Kolb's (1984; cf. Ball and Koliousis 2023) four-step learning cycle to computer science and engineering courses, with activities progressing through self- and peer-evaluation and educator feedback phases at the end (summarized in Sanz et al. 2023, Fig. 2). Their first step sensitizes students to a target situation with emphasis on SPC aspects, and clarifies empathy as an explicit goal of course exercises (cf. Walther et al. 2017 correlating prior conceptual understanding with learning outcomes). The second step directs students to take on the problem from their present understanding. The third step involves an empathy evaluation, with students aware of results. The fourth step tasks students with taking up the problem situation from the perspective of others affected in collaborative contexts. Sanz and colleagues suggest that educators introduce different communication strategies and discussion exercises at this stage, depending on expertise. The fifth step involves self, peer and course assessment questionnaires, and the sixth educator feedback. Results are discussed as a group with focus on difficulties in perspective-taking and communication during the exercises (discussion, Sanz et al. 2023, Sect. 4.B).

Sanz et al. (2023) describe an implementation of the proposed cycle in Sect. 5 and discuss results in 6. Students (first year, at two different universities, in Argentina and in Spain) embraced working in collaboration with peers, recognizing needs to improve communication skills (e.g. affective sharing). Overall, tested empathy scores were low, with 40% evidencing inflexibility in perspective-taking. Females scored higher on empathy, and overall students seemed to overestimate empathic awareness and perspective-taking abilities. These results are interesting, given practitioner over-confidence in pro-ethical design abilities reported in Morley et al. (2023), and supports their recommendation for improved empathy training.

Case-based studies are core to empathy training. Thiel et al. (2013) show that case-based studies designed to elicit emotions such as compassion facilitate learning. Martin, Conlon and Bowe (2021b) report on the use of case studies in engineering ethics education in Ireland. They find different instructors using case studies differently, hypothetical, real-world, stressing tragic outcomes exposing emotional consequences of "unethical decision-making". Martin and colleagues note the roles for cases resembling what a practitioner may expect in the field. They suggest that scientific content represented alongside societal concerns may be "provocative" in a way that motivates a sense that things can be done better. Hess et al. (2019) reported on five iterations of semester-length graduate engineering ethics courses employing the scaffolded, integrated, reflexive analysis of case-studies (SIRA; see Kisselburgh et al. 2014 for review). SIRA trains perspective-taking and leverages compassion for others in resolution of emotionally rich case-based problems. Kotluk and Tormey (2023) focus on compassion, finding that compassion and empathy encourage recall of technical and ethical content. However, they warn that overly

tragic cases can force a distancing response in students. They suggest instead that "even small-scale ethics cases that do not involve significant death and destruction are resonant with a range of moral emotions." (p 18) They also warn against ignoring moral emotions implicit in case studies, good or bad. They propose that emotional processing of ethics cases should be a focal aspect of mainstream teaching methodology, and offer Hess et al. (2019) as a model approach.

Integrating emotion into technical studies through case-studies is a key aspect of holistic ethics education. One traditional approach involves *phronesis* understood as practical wisdom to do the right thing at the right time. Han and Athanassoulis (2023) argue directly for the necessity of exemplar cases in training *phronesis*. Exemplars represent morally salient aspects of shared situations that other representations cannot. Exemplars demonstrate moral virtue which may be emulated. Considering human neurology including the mirror-system and self-related processes central to moral cognition (Han 2017, 2023; Koenigs et al. 2007; Young and Koenigs 2007), exemplars are imminently relatable. They are also motivational; more than remembering to apply some principle, students want to become that person. Han and Dawson (2023) consider the relatability and attainability of moral exemplars. They find that both contribute to reported "moral elevation and pleasantness" and suggest that in selecting cases, educators focus on relatability, then attainability, to promote moral motivation (cf. Han et al. 2022).

Whistleblowers are moral exemplars. Whistleblowing often attracts popular attention, making these cases recognizable entry points into a holistic ethics education. Whistleblowers are often interesting people whose stories are compelling, with the consequences of their moral agency evident in their demeanors (e.g. Allan McDonald of the famous Challenger disaster, as recorded in Ethics Case Study 1 for the American Society of Civil Engineers[1]). Whistleblower cases can be controversial even without tragic death tolls from identifiable engineering failures (e.g. Edward Snowden). Being relatable, with colorful media supporting their presentation, their study can encourage students to adopt perspectives of decision-makers at focal moments of moral agency, affording educators opportunities to introduce theoretical or metatheoretical tools to increase understanding of salient dynamics. The basic idea is to inspire interest, so that students operationalize their experience in more mundane contexts later on.

Whistleblowing is variously defined, with common features involving the disclosure of morally repugnant policies or actions undertaken by officers within an organization, usually in leadership positions, to some authority or interest receptive to such communication and more or less able to effect necessary reforms. Whistleblowing is characteristically not a preferred option for the whistleblower, presenting as necessity only absent other avenues for communication and reform. Near and Micelli (1985) define whistleblowing in the professional context as:

"... the disclosure by organization members (former or current) of illegal, immoral or illegitimate practices under the control of their employers, to persons or organizations that may be able to effect action." (p 4)

Near and Micelli (1985) considers whistleblowing a form of "organizational dissidence". Dissident action is motivated by felt duty or obligation to society or humanity at large that trumps local responsibilities. Such acts proceed from concern for others whether or not these have clear moral standing, communicating ethical concerns to motivate organizational reform rather than taking other actions, e.g. resignation, armed rebellion. As such, it is pro-ethical, and should not be considered "deviant" (p 3). Requisite changes to an organization may threaten privileged standings therein, inviting reprisal. Near and Micelli find whistleblowers motivated more by potential efficacy than fears of reprisal, however (p 6-7).

Near and Miceli (1985) analyze whistleblowing as a process composed of four decisions (reflected in four "propositions" on p 8). First, perception of wrong is the "discriminative stimulus for whistleblowing" (p 7). However, moral perception is variable; "individuals with higher levels of moral reasoning ... see different activities as wrong than would other observers" and are thus "more likely to blow the whistle" (p 8). Following steps depend on an evaluation of the significance of perceived moral infractions, and an understanding of how and to whom such a

---

[1] The American Society of Civil Engineers maintains this video on YouTube: https://www.youtube.com/watch?v=QbtY_Wl-hYI Accessed 21 July 2023.

report may be communicated with maximal potential for adequate reform (p 6). They discuss the whistleblower's personal situation, noting that stress and financial burdens may be crippling. So informed, the whistleblower may attempt to expose perceived wrongdoing. Near and Miceli (1985) then consider the organization's response, focusing on the interests of the "dominant coalition" including "silencing" the whistleblower with "the least costly strategy ... to discredit her charge" through destruction of reputation, for example (p 5).

The "dominant coalition" within an organization holds far greater power than the whistleblower. Thomas (2020) warns against seeking resolution of moral concerns from within such a power structure. Thomas considers power the ability to effect change through influence over others, getting them to get things done. With power relations in the context of whistleblowing typically hierarchical, he echoes Near and Miceli's (1985) concerns for the victimization of the whistleblower by the "dominant coalition" (compare Thomas's "enclaves" in 2020, Sect. 4). He summarizes their four step process in terms of first, recognition of third-party detriment due to actions or policies pursued by the organization's dominant coalition; second, the "local subversion" of this coalition as a challenge to their "authority"; third, an appeal to a person with "higher power" either internal to or external from the organization; fourth a "reasonable expectation" that this higher power will share moral concerns and take corrective action.

Thomas (2020) contrasts whistleblowing with three other forms of organizational dissidence; complaining to an authority with the self as the victim, empowered officers reporting subordinate misconduct, and snitching on peers which "may be justified on consequentialist or deontological grounds" but which differs from whistleblowing power relations (Sect. 2, fig. 2; see also Ceva and Bocchiola 2020 who consider the application of these two theories in tandem; Ceva and Bocciola 2019). Thomas provides realistic examples differentiating these phenomena which may be useful for ethics course design involving exemplar case-studies. Thomas then assesses different network structures and power relations with a focus on to whom and how a whistleblower should appeal given different embedding contexts. This analysis may be useful during exercises considering different cases.

In the classroom, whistleblowers are used to expose situational complexities and illustrate barriers to pro-ethical action, including institutional power structures serving dominant coalitions as their values run contrary to the common good. In its clearest form, whistleblowing involves pro-ethical action within a hierarchical power structure with high potential for personal suffering in effort to reform an organization by way of which a dominant coalition pursues policies and actions detrimental to third parties. Where there is no potential to seek reform within an organization, Thomas considers appealing to external authority, and Near and Miceli (1985) consider "going public". Here, the intention may be to redistribute discretion away from a dominant enclave. However, Thomas warns that appealing to external interests may cause harm should such information only strengthen central powers, e.g. federal agencies, that are also "malign" (Sect. 6; in the context of corrupted government Delmas 2015; Bellaby 2017; Berg 2020). Extreme measures may be necessary, i.e. "Snowden" becomes a verb.

Martin, Conlon and Bowe (2021a) are critical of whistleblower case-studies. They suggest that care must be taken to select cases matching the aims of the education, to prepare professional engineers. They contend that the "dominant" approach individualizes students, placing inadequate emphasis on their social embeddedness. Whistleblowing especially may under-represent extensive social factors. They warn of "unrealistic expectations" and "moralism" which may interfere with engineers acting as professionals responsible to their profession. Rather than on ethical inadequacies, Conlon (2015) argues that focus should be on aspects of the "overall sociotechnical system" which invite "accidents" and on their correction. It is from this perspective that Conlon and colleagues move focus away from personal responsibilities and values, toward macro-ethical issues including political economy, public policy, and the goals of engineering as a profession (cf. Conlon 2011; Trbušić 2020 recognizes similar areas).

Martin, Conlon and Bowe (2021a) recommend "centering ethics within the institutional culture" (p 24) with ethics courses re-conceived accordingly, i.e. leveraging examples from best practice over where these fail, e.g. whistleblowers. Their "agenda for a socio-technical orientation of engineering education *for* ethics" involves redefining "what it means to be an engineer" around their concept of "socio-technical identity" in order to "generate commitment to larger systematic

change to established practices over time, rather than suggest heroic responses to management wrongdoing" (p 26-7; compare Sanz et al. 2023). They recommend further research into "effectiveness and coherence" between "implementation, teaching, assessment methods" and "goals and theoretical frameworks envisioned for engineering ethics education" (p 25), stressing that evaluation metrics assessing effectiveness of different teaching approaches (including "development and application" of selected case studies) remain "underdeveloped" (p 26).

Conlon, Martin and Bowe (2018) challenge individualizing "holistic" approaches, and consider sociological metatheory for insight into how to represent system-level issues. Ritzer (1990) classifies sociological metatheory by product of application; understanding of theory, new theory, or unifying theory. Ultimately, the aim is a unifying theory. Some research concerns how theory shapes communication for constructive social change (cf. Dutta 2020). How we think about things affects how we talk about them, what we do with them together, and how the world turns out as a result. Recalled Morley and colleagues' DCEF, theory is a form of technology that may be purposefully developed to support ethical action. Ideally, we clearly communicate a unifying theory to speed their proposed cultural shift.

LaCroix (2022) considers model use in ethics teaching from a metatheoretical perspective. He considers trolley problems in the context of semi-autonomous vehicles. Trolley-style problems have become a norm in ethics classrooms. They represent moral dilemmas. They can be modified to challenge different intuitions. Different features may be stressed or incorporated to expose intuitions and afford reflective analysis. However, educators seem to miss the point in moral dilemmas (Sect. 3). LaCroix argues that Trolley problems are misused in ethics classrooms, failing to represent "how ethical the machine actually is relative to some meta-ethical standard", instead framing ethics as an exercise in aiming for lesser evils (Sect. 4). LaCroix considers the purpose of moral thought experiments in general, to "elicit normative intuitions" and to direct the felt confirmation of correlate conclusions, as a form of persuasion that exposes latent values to critical examination, as "validation proxies" rather than mere decision models. From a metatheoretical perspective, the point of the trolley problem is that neither option is a good one. Likewise, from a design and engineering standpoint, the entire situation is to be avoided. LaCroix warns against AI ethics standardization around such models, recommending further inquiry into adequate "proxies for moral decision making" (Sect. 5).

Case studies including whistleblowers serve as proxies for moral decision-making in the ethics classroom. Whistleblowers are often motivated by concerns for disaffected, powerless others external to or detrimental to central interests. Such cases lay bare endemic corruption resulting in identifiable harms due to morally repugnant policies pursued by dominant enclaves or coalitions who exercise power (often financial, over job security or status) to influence subordinate others to follow along. Scaled-up from company to country to globe, this is the dominant "sociotechnical system" into which a prospective engineer may find herself professionally employed, today. Ideally, this "system" and its officers would not require correction, but the world is not there, yet, cue the costs of ethics.

There is positive work in the direction of minimizing these costs. For example, Schuett (2023) integrates whistleblowing communication into an internal auditing team for (large) AI labs. Schuett's aim is to integrate such an active auditing team into current corporate best practices, reporting to executive management as a "layer of defense". To borrow from Near and Micelli (1985), this is an organization "conducive of dissidence" (p 8) with whistleblowing "role-prescribed" (and so, strictly speaking, not whistleblowing, p 2). The cost of ethics is 'baked-in' to such a model (more about this in the next section); and, "whistleblowing" remains part of the equation.

There is no guarantee that institutions which follow Schuett's protocol do not simultaneously pursue confounding policies on the side, such as selecting employees unlikely to blow a whistle, with lower-levels of moral reasoning, compartmentalizing operations so that each sees less of what may be found offensive, and so on. But, shifting focus away from whistleblowing because of such factors misses the point of ethics education. Yes, challenging a dominant coalition potentiates personal suffering, a neglected cost of ethics. The question becomes for the pro-ethical professional, are you ready to pay those costs to make things better?

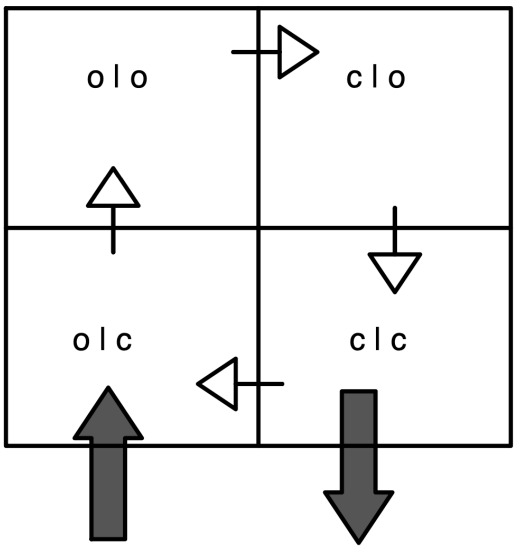
If the aim is to systematically train professionals to weather the emotional turbulence of compassionate consideration for the potentially poorly affected through perspective-taking, and inspire motivation to do the right things for humanity at large through the experience, then it is difficult to overstate the importance of exemplar cases including whistleblowers in ethics education. An expanded ability to relate to others as uniquely historically situated human beings resonates with the idea of the philosopher engineer as a way of life, as well.

Moreover, whistleblower cases teach something about the evolution of engineering as a profession. As bottom-up ethical input integrates with best practices, we see more than technological progress. We see what Kant considered moral progress. The job of ethics education is to train necessary high-level moral reasoning, leveraging theoretical proxies for moral decision-making, so that this progress continues. A suitable model should communicate unifying theory facilitating constructive social change, including Morley and colleagues' cultural shift. The ACTWith model was designed for such purpose, and is introduced in the next section.

3. The ACTWith model

The ACTWith model is a four-step information process. ACTWith stands for “As-if Coming-to-Terms With”. The model consists of four modes which can be considered in isolation and dynamically as a cycle. Each mode represents a combination of closed and/or open affective/rational operations represented dynamically in Fig. 1 as "the beating heart of conscience":

o/c: As-if (open) coming-to-terms with (closed)
o/o: As-if (open) coming-to-terms with (open)
c/o: As-if (closed) coming-to-terms with (open)
c/c: As-if (closed) coming-to-terms with (closed)



open || closed = affectively open
- feeling as-if

open || open = affectively and explicitly open
- feeling as-if and coming-to-terms

closed || open = explicitly open
- coming-to-terms

closed || closed = closed
- action

Figure 1: ACTWith model dynamics as “the beating heart of conscience”

The ACTWith account was inspired by Ron Sun's hybrid CLARION computational model (Sun 2001; cf. Bretz and Sun 2018; White 2013) and a dynamic systems interpretation of fundamental imaging research applied to studies in traditional ethics and phenomenology (White 2006). The ACTWith model represents iterative information processing upward and downward through a spatial-temporal hierarchy as in perception-action (cf. Fuster 1990) in an accessible way. Grounding research focused on the felt comparison of embodied situations via affective/effective mirroring and insular gating dynamics (e.g. Umilta et al. 2001; Wicker et al. 2003; Gallese, Keysers

and Rizzolatti 2004). One aim was to account for Kant's admonition that we must overcome disgust in order to open to situations of suffering others (White 2010, 2014).

The ACTWIth account is a general theory of cognition. The basic idea is that cognition involves the comparison of situations, intelligence involves finding ways to move from one situation to another, and agency involves getting there. These situations may be one's own or another's as informed through the mirroring of more or less similarly embodied conditions. The idea with ACTWith education is to train routine conscientiousness, afford introspective control over relevant dynamics, and to structure information flow at different levels of social organization in the way of a general model.

Processing begins with the subject agent opening to input bottom-up (o/c). Input is felt as difference from prior embodied determinations. Prior determinations are relaxed as explicit processes open to the felt situation (o/o). This information flows into the third stage (o/o -> c/o) as the agent closes to further low-level input, "coming-to-terms with" (c/o) input information as prior determinations are adjusted, ultimately to inform action (c/o -> c/c).

The aim at c/o is to minimize differences between the felt situation (one's own or another's) and a better one. As a hybrid model, adjustments which minimize this difference correspond with an explicit action plan, or policy, e.g. an ethical principle. With a policy formed, action may proceed toward the ideal that this policy expresses. The cycle is repeated, either in closed-loop reflection, or through action with feedback informing subsequent iterations.

ACTWith dynamics can be associated with traditional ethics accounts. Consider this passage from Adam Smith (appended with ACTWith shorthand):

By the imagination we place ourselves in his situation [O/C], we conceive ourselves enduring all the same torments [O/O], we enter as it were into his body [C/O], and become in some measure the same person with him [C/C], and thence form some idea of his sensations [O/C], and even feel something which, though weaker in degree, is not altogether unlike them [O/O]. His agonies, when they are thus brought home to ourselves [C/O], when we have thus adopted and made them our own [C/C], begin at last to affect us, and we then tremble and shudder at the thought of what he feels [O/C, understanding that the cycle repeats]. (1.1.1.2, Smith 1759)

This passage describes affective and rational processes iteratively employed affect-first in empathic perspective-taking. Smith describes an immersive condition with another's situation "brought home to ourselves". One feels as-if in that situation, affording a sense of value from that perspective. This processing cycle represents the basic movement of empathy training. One must open to a potentially embodied situation bottom-up, affect first, before 'bringing that situation home' through high-level reasoning processes.

Kouprie and Visser (2009) represent similar perspective-taking dynamics in their framework for empathy in design. They describe "stepping into and out of" a user's life through affective and cognitive processes (p 442-443). They consider affective empathy as felt identification with or "becoming" the user (i.e. as-if) and cognitive empathy as understanding that situation through imaginative perspective-taking (i.e. coming-to-terms with). Drawing from psychological theory, they propose four phases of empathy. The first involves willingness to enter the user's world without prejudice (o/c); the second involves immersion into that situation as-if one's own point of reference (o/o), absorbing without judging; the third involves explicit processing (c/o), finding common meaning recognized in cognitive and emotional "resonance"; the fourth step completes detachment, with increased understanding to design from the user perspective (c/c). Kouprie and Visser (2009) suggest that the immersive step is the most important in empathic design, as without taking time "to wander around in and be surprised by various aspects of the user's world ... knowledge of the user's world will not increase" (p 446).

Morley and colleagues' (2023) four-level DCEF model also leans heavily on ACTWith o/o dynamics. The DCEF begins with personal moral value at the first level (bottom-up, o/c), opens to broad social concerns at the second (o/o), informs legitimate principle through open bottom-up information in the third (c/o), and develops support for continuation of this processing cycle in the fourth (c/c). In essence, the DCEF describes ACTWith beating heart dynamics spanning levels of

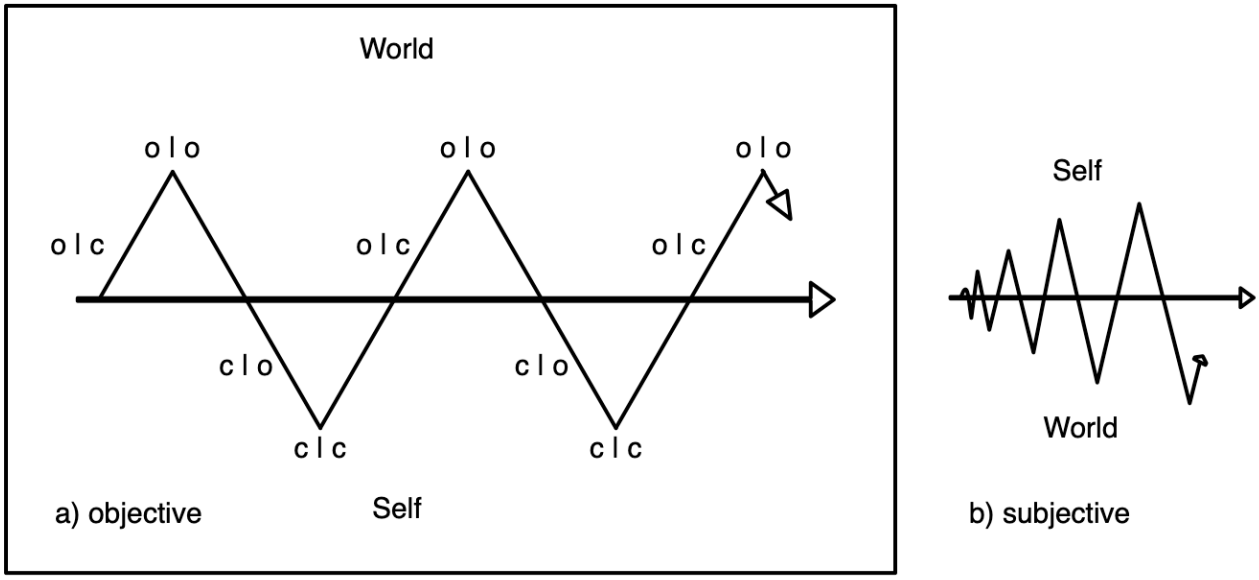


Figure 2: Stitching one's self into the world: a) objective plot; b) subjective plot

organization from self to global social system (recalling Sanz' SPC). The focus of their cultural shift in education is ethical awareness through empathy-training, i.e. o/c -> o/o, to ground principle (o/o -> c/o) to inform (broad social) action (c/o -> o/o).

In the classroom, the focus of ACTWith exercises is empathy training and metacognition. The basic idea with empathy training is to change reflexive information processing. The basic idea with metacognition is to autonomously modulate that processing.

Sanz and colleagues' empathy skills are reflected in phases of their course design, correlative with ACTWith information flow. Phases 1 and 2, sensibilization and appropriation correspond with open (input) modes (o/c -> o/o), 3 is an empathy assessment informing forward processing (o/o -> c/o) and phases 4 (deployment including collaborative planning and problem analysis (more o/o -> c/o)) and 5 (reflective self-evaluation) involving closed (output c/c) modes (c/c -> o/c). Phase 6 is feedback represented by input/output arrows in Fig. 1.

During empathy training, the ACTWith cycle bridges differences between situations by prompting students to adopt others' situations as their own, gradually revealing hidden significances of objects and opportunities as they research and discover prior personal limitations, raising ethical awareness. In this way, conscience is educated through perspective-taking, potentiating practical wisdom or *phronesis* (per Han and colleagues) as understanding of the world and one's place within it grows. Lifelong practice may shape philosopher engineers. Example exercises are considered in the next section.

Moral growth and development over the life-course is illustrated with ACTWith dynamics in Fig. 2, "stitching one's self into the world". Fig. 2a illustrates a textbook processing cycle, regularly opening to information and closing in action. The self is a process that exposes the world, in co-discovery, including the "built" environment (cf. care structure in Heidegger 1996). Fig. 2b represents the subjective point of view. As the human being develops, understanding of self and world grows and becomes more complex, responsibilities are taken on, principles adopted in routine practice, potentiating the formation of purpose in life and something like Martin and colleagues' professional identity through late adolescence and into adulthood. The basic idea is that we stitch ourselves into the world as we open to it, care for it, and act to make it better.

Different cognitive styles can be associated with different static modes, as these may predominate over time, i.e., o/c for characteristically spontaneous and gregarious, o/o caring and compassionate, c/o analytic and in command, c/c active; or, with characteristic emphases on

modes routinely enacted during context-dependent iterative processing. Conscientiousness for example emphasizes empathic (o/c -> o/o) opening to understanding a situation as one's own (o/o -> c/o) to advise forward action (c/o -> c/c), while remaining invested in the result (habitually open to associated situations and considering these in reflection c/c -> o/c). This investedness is the stitch that holds self and world together.

Relative de-emphasis on empathic modes corresponds with instrumentality and selfishness, characteristic marks of psychopathy. Such a cognitive style emphasizes c/o -> c/c with relative de-emphasis of input o/c -> o/o. With each iteration, as o/c -> o/o is de-emphasized, affective investment in the outcome of a given situation does not develop as deeply over time. Policies c/o informing action are considered absent such information. Such imbalance in processing is consistent with Shalev, Eran and Uzefovsky (2023) who describe psychopathy as "empathic disequilibrium" over-emphasizing cognitive over affective processes in a similar way. Basically, diminished mirrored affect trains moral deficiencies potentiating psychopathy recognized as instrumentally aggressive behavioral dispositions, i.e. acting as-if others' situations don't matter, or matter less (White 2012). Rather than stitching into the world, such policy stitches others out (cf. Magnani 2011 on embubblement). In these ways, routine processing can be associated with differential development of self and world. Empathy training with the ACTWith model teaches routine conscientious.

Metacognition involves cognition of cognition, presenting cognitive processes in a way that affords self-appraisal, with tools like the ACTWith model making these processes transparent and accessible to reflective moderation. Fundamentally, the model is an information processing cycle, with iterations experienced as the felt difference between embodied situations. One practical idea is that autonomous modulation of such processing can be taught with the ACTWith model. Another is that understanding common cognitive dynamics grounds consistent and constructive interpretation of various program accounts, even where such common grounds are not originally intended or already recognized. In grounding various initiatives universally, the ACTWith account offers a general organizational principle.

ACTWith consistent metacognitive dynamics are evident in Sanz and colleagues' (2023) "mode switching" for example. Their "affective sharing" (o/c) and the ability to "feel with others and experience their internal world" (o/o) correlate with empathic perspective-taking. Mode-switching between "empathic and analytic cognitive mechanisms" (Sanz et al. 2023, Sect. 4) is represented by more or less selective opening or closing to embodied situations in empathy (o/c and o/o) or considering what can and must be done from that information (c/o and c/c). Deep self-understanding, emotion regulation correlate with reflective cycle dynamics modulating c/c -> o/c and o/o -> c/o. It is expected that embodied capacities may develop through associated training, especially during critical developmental periods including adolescence and university education.

Situatedness is a key aspect of the ACTWith model, important in the classroom for making otherwise distant or abstract ideas relatable (compare Trbušić's social imagination, also Lange and colleagues' "moral imagination" in the next section of this paper). Consider the difference between asking students to imagine a situation as-if their own, to live and to die within and so determined, or to consider a possible action in terms of three normative ethical theories, with written results due Monday. The former is accessible, relatable, and maybe interesting. The latter is distant, abstract, and pushes against student preferences for practical ideas. One practical idea is that moral cognition is a bottom-up process. Model proxies and class exercises should represent salient processes consistently. Whistleblower cases help to illustrate such dynamics.

Here, consider whistleblowing in ACTWith model terms. Near and Micelli (1985) describe whistleblowing as a four-step decision-process involving the corporation. First, the observer determines if a perceived action is "illegal, immoral or illegitimate" and if this concern should be made explicit so that it may be communicated to inform action (o/c -> o/o revealing a morally repugnant situation, and o/o -> c/o to communicate this repugnance). The aim of the whistleblower is an ideal, i.e. a situation in which a whistle need not be blown (so that personal processing cycles may proceed uninterrupted). At the organizational level, the cycle represents the individual as the bottom-up input into executive decision-making. Then, these officers exercise the same processing cycle individually (either opening to upstream information flow, or not), choosing to modify corporate policy and/or seek reprisal through organizational dynamics.

Power dynamics are clearer in Thomas's (2020) treatment. His four-step process (appended with ACTWith notation): first, recognition of third-party detriment due to actions or policies pursued by the organization's dominant coalition (o/c -> o/o); second, the "local subversion" of this coalition as a challenge to their "authority" (o/o -> c/o, in coming-to-terms with the fact that the situation must change); third, an appeal to a person with "higher power" either internal to or external from the organization (c/o -> c/c, effectively handing-off information to another whose beating heart of conscience may effect necessary change); fourth, a "reasonable expectation" that this higher power will share moral concerns and take corrective action (c/c -> o/c, trusting that future input will evidence requisite correction).

Recall Floridi's (2018) explainability alongside holistic communication skills; it is up to the whistleblower to make concerns clear, relatable. It is up to the executive to inform forward action. In the classroom, collaborative exercises may emphasize o/o -> c/o dynamics representing whistleblower-executive dynamics, so that students may train in either role of the process. Sanz and colleagues' (2023) deployment phase for example involves collaborative perspective-taking exercises which may frame such discourse. Schuett's (2023) "layer of defense" is effectively the voice of conscience at the corporate level of organization, advising against egregious action (the arrow from c/c to o/c in Fig. 1) and impelling empathic reflection (o/c -> o/o -> c/o) until pro-ethical action-plans emerge at the organizational level, recalling Near and Micelli's (1985) sense that whistleblowing represents constructive information flow within the organization.

The ACTWith model may augment introductory courses representing traditional ethical theories, typically the main three: deontology, utilitarianism and virtue theory. One problem with introductory courses is that original theories are misrepresented when over-simplified in effort to make them relatable. For example, Kant, Mill and Aristotle have more in common than they differ, but associated moral theories are often represented as if they were unique, i.e., as three different theories. These theories can instead be mapped to ACTWith model dynamics as cognitive styles characteristic of their exponents. Mill emphasizes applying policy (c/o -> c/c) then assessing results (o/c -> o/o). Normally, we do as always already done, pausing for reconsideration only when established routines deliver poor results (compare Floridi on legislation). Kantian moral theory emphasizes upstream empathic modes, goodwill, opening to situations of all other agents able to do the same (Kant calls them "rational"), and understanding this universal-fellow-feeling as moral law (essentially, you have a duty to o/o -> c/o). The practical idea is: Do not put others in situations that you would not seek for your own. Ideally, everyone proceeds accordingly, self-assembling Kant's Kingdom of Ends on Earth.

Both Kant and Mill emphasize the role of conscience in determining right from wrong. The categorical imperative can be considered the voice of conscience, and Mill is explicit that conscience is the ultimate judge of right action. They differ in emphases on processes informing such judgments. As for Aristotle, his emphasis is on the cycle in its routine operation, as a whole, "understanding of understanding" as represented by the "beating heart of conscience" (Fig. 1). In this way, three theories commonly represented as distinct are accessibly represented in common terms of a general information processing dynamic. More broadly, relationships between different inherited traditions can be clarified.

When introducing a course using the ACTWith model, ethical theory is not an add-on; it is useful. Kant and others are describing something normal. We are situated, coming to terms with our situations. Empathy means coming to terms with bad situations. Opening to bad situations means overcoming disgust. We must overcome disgust to practice empathy; Kant did it, too. Conscience objects when bad situations are avoidable, this is normal. In this way, ethics becomes accessible, relatable, cue Millian reconsideration of prior convention, whistleblowing and all the rest. How can we do better?

This is the situation into which the practicing designer and engineer of powerful technologies wakes, daily. Teaching to take up this place should be the aim of applied ethics education. By integrating traditional ethical frameworks into holistic empathy training with the ACTWith model, abstract ethical principles become practical, and with practice, familiar. The ACTWith approach is on target. Supporting such a model in the classroom depends on other factors, however. Illustrations from past classroom exercises and comparisons with contemporaries are offered in

the next section, to help to overcome educator limitations, lower barriers of uptake, and catalyze necessary institutional changes.

4. Illustrations

The ACTWith account offers a general theory of cognition. Cognition on the ACTWith account is co-gnition, with -gnition from the Greek "*gnosis*" here signifying the felt situated and familiar, grounding, knowledge differing from higher intellect, "*eídein*". The basic idea is that awareness of determinable things arises in the felt comparison (the co- for the -gnition) of embodied situations in a bottom-up way. Routine associations are learned as habits of mind which can be represented as guiding principles. Autonomy requires that prior guiding principles and routine associations be revised, with further ideation or action proceeding from iteratively enriched understanding.

Moral autonomy is this process involving others, recognizing that the situation is shared in a strongly embodied way, through differential development of the mirror-neural system. The basic idea with the ACTWith education is that routine dynamics involving these and other systems can be trained, and when made explicit, subject to self-modulation. Opening to other-embodied situations involves taking up those situations as-if one's own, and moral agency involves formulating and executing associated actions from this perspective (writ large, from the perspective of Kant's kingdom of ends). ACTWith model dynamics represent this process for metacognitive self-appraisal. As a strongly embodied account, information is understood in the strong way. One is in-formed when opening or closing to situational information. Neural processes change, develop, are destroyed, are weakened or reinforced. As one opens to other-situational information in this strong way, and long-term value associations are altered in light of the shared situation with others similarly embodied and affected through the ACTWith o/o mode, moral cognition is exercised, and through self-modulation of these processes over time, augmented.

As a general theory, the ACTWith model can be applied in different contexts, including to whistleblower case studies, stakeholder analyses and management as regularly taught as aspects of engineering and business ethics, and in technical training contexts for practitioners in and out of the classroom. The following subsections illustrate the modification of typical undergraduate ethics training centered on a whistleblowing example (4.1), the coordination of different materials around the ACTWith model for design of applied ethics courses dedicated to different disciplines and degree programs including Information Technology, Engineering, and Business Administration (4.2), application of the ACTWith model account to interpret emerging technical issues in the context of autonomous aerial vehicles or "drones" (4.3), and active parallel efforts in private enterprise in the context of software development at Google (4.4).

4.1 Shifting the standard approach

A common way to approach applied ethics in the classroom is through standard ethical theories applied to whistleblower cases. One example comes from ongoing courses in the Netherlands. Students are tasked with assessing a fictional situation in terms of virtue, deontological, and utilitarian theories. Simplified versions of these accounts are presented and students are expected to independently research when compiling written responses to a case-based challenge such as the following:

You are a member of an AI start-up experimenting with brains-on-chips using stem cells programmed using viruses. The company mission statement acknowledges risks as essential to any new technology, and claims that benefits are expected to outweigh them. The legal team has assured you that nothing that the company is doing is illegal, but one member of the research team had come forward in public interviews about the use of stem cells and viruses, contrary to signed non-disclosure agreements. Public attention has been attracted, and the company is deciding how to proceed. As you have some expertise in ethics, we need you to explain how such a whistleblower's disclosure might be morally - if not legally - justified.

Students should from here apply the three main theories in composition of an essay. Though case-based, and maybe interesting, such an activity suffers some infelicities that make it inadequate for contemporary ethics education. One, given accessible generative AI, there is increasing potential for machine-generated reports. In places with competitive scoring, current AI generated responses are unlikely to satisfy most students. Under such conditions, instructors may be more cognizant of potentials for AI to re-compose best performing essays from prior course iterations in apparently novel ways, instead. In other student cultures, direct machine generation of such an essay may yield desired results. Many students may aim for minimal passing scores, especially in a side-line subject like ethics at a large technical university where long-form responses are mandatory for course grades and a single instructor may be responsible for scoring five- to seven-hundred students of many different degree programs every quarter. This is troubling, because such dedicated technology students are the target demographic for applied ethics education, and the most likely to best utilize technical solutions when approaching course assignments, i.e. large language models such as ChatGPT. Setting aside faculty limitations, a course designed to minimize potential for shortcuts to gaining an operational understanding of ethics is preferable.

Consider the following ACTWith model adjustment to the preceding illustration. Students are tasked with producing a stakeholder analysis diagram, a risk assessment based on that stakeholder analysis, and a brief presentation. The presentation should be a slide-show applying the ACTWith model to identify non-obvious risks for selected stakeholders. Students are encouraged to use content from lectures on Kant, Mill, and Aristotle at appropriate stages, perhaps by describing what is learned in adopting an affected stakeholder's perspective, e.g. user, neighbor, competitor, advertiser, investor, client, supplier and so on. Consider what may be ideal for each, for example ask them, consider similar cases, or consult established convention. Recommend possible forward actions that do not contradict Kant's imperatives, and that proceed from goodwill, i.e. inclusive of all affected stakeholders. Which possible paths of action maximize utility according to selected materials from Mill introduced during course lectures? Which actions represent virtue according to course materials from Aristotle. Final deliverables are conditioned on feedback from this presentation, including peer feedback.

An ethics course module entitled "Four step applied ethics cycle: incorporating ethics into the daily work and information flow" confronted University of Twente students with a similar problem during Spring 2020. That course module assigned different small groups to focus on specific steps in the ACTWith process. The aim was to embed ethics into a larger organization in real time around final project preparation. With duties spread out, members of different groups had to communicate findings with members assigned different steps of the cycle. Some students could focus on dangerous situations to be avoided or sought, others on expanding the ethical circle to include diverse perspectives, and others on operationalizing action plans while setting up the organization for feedback. Students presented results, and final assignments were as follows (adapted for a general case):

Company stock is down. Your group has been assigned to understand how executive management failed to assess risks and proceed with all stakeholder interests in mind. Your recommendation will advise new leadership how to re-structure corporate information and work flow so that morally sensitive stakeholders are not compelled to risk everything to protest perceived injustices in the future. Public attention in this way is bad for company reputation, corrections can be costly and are to be avoided. Further background information is offered during lectures and is available on class slides uploaded to the course page.

As ACTWith model dynamics have mostly been the subject of coursework exercises and assignments, AI homework assistance is little concern. Stakeholder analyses and risk assessments are assigned as diagrams supported by short textual explanations, and incorporated into presentation slides. To date, an AI is unavailable that can provide such solutions. Most importantly, through such exercises, students operationalize ethics, making ethics part of their routine work and information flow considered intuitively as habits of mind. Ideally, co-workers and colleagues are similarly trained. In such cases, as in the course reviewed above, small groups can be assigned to focus on different steps in the cycle and associated analyses including stakeholder and risk assessments, with the anticipation that members of different groups inform others about assessments with different foci, thereby training expectations

facilitating collaboration around ubiquitous ethics issues in large institutions with distributed technical responsibilities.

4.2 General applications

One upshot of the ACTWith education is that traditional courses can be supplemented with various ACTWith consistent programs to suit different contexts, while ACTWith processing dynamics remain constant across them, binding the course together around a common theme. When teaching business ethics in the Netherlands, the ACTWith model was paired with Freeman's value creation stakeholder management theory (Freeman 1984/2010). Associated assignments drew from Freeman's (2012) "Stakeholder management and Reputation" (openly accessible here: https://www.bbvaopenmind.com/wp-content/uploads/2013/02/Stakeholder-Management-and-Reputation_R.Edward-Freeman.pdf). A central tenet of Freeman's theory is that economic and social conditions are inseparable and complex. Courses emphasized Freeman's treatment of the "separation fallacy" involving extraction of profit through business without regard to distributed, potentially negative, social and environmental impacts. Freeman emphasizes a shift in emphasis away from profit as the primary measure of success in business to reputation gained through contribution to community including local stakeholders such as employees and their families, friends, and supporting organizations. Freeman and colleagues encourage business leaders to anticipate local stakeholder needs through symbiotic relationships that are stable and mutually rewarding. One neglected upshot of this strategy is that employees feel pride in company reputation, inviting dutiful work practices and harmony in and outside of the office and worksite. The ACTWith model was applied in this context, at individual and company levels, as perspective-taking exercises considering stakeholder interests and company values statements.

In ethics of engineering and design, the ACTWith model offers a simplified framework for teaching and learning Vallor and colleagues' (2018) seven-step applied ethics cycle. Their seven steps (appended with ACTWith correlates) proceed from "risk-sweeping" for worst-case scenarios (o/c), to "expanding the ethical circle" and case-based analyses to contextualize current work (o/o), to reflecting on the intended good as action plans are formulated, considering possible tragic outcomes (c/o), and if sound then act while setting up for feedback in order to re-enter the cycle through iterative practice (c/c). These steps can be mapped to the ACTWith model explicitly and taught alongside Vallor and colleagues' ethics toolkit. During undergraduate Information Technology and Business Information Technology courses at the University of Twente from 2019-2020, one sample case involved a software company whose most recent game release has been poorly received by the gaming community, perceived as a "cash-grab" with micro-transactions creating a pay-to-win ethos that proved divisive and socially damaging. The ACTWith model was used to structure discussion of different stakeholder perspectives. The aim is to train moral sensitivities to stakeholders marginal to profit-driven projects, correlative with Freeman's (2012) re-imagination of business as symbiotic with communities that support them.

Bezuidenhout and Ratti (2021) work from Vallor (2016) in focusing on "self-cultivation" through selective "moral attention" in discernment of salient aspects of situations for informed moral judgement (section 4.2 "Foregrounding virtuous behavior"). Vallor's (2016) selective moral attention driving self-cultivation parallels the ACTWith education of conscience though iterations of more or less selective opening and closing to other-embodied situations. Their focus is on expanding capacities for moral attention to the right things in the right ways at the right times in a data science context. They focus on bringing ethical concerns into the spheres of influence of working data scientists as "citizens" and as employees. They consider training operational ethics in terms of "microethics" by illustrating macroethical consequences of systematically applied routine decisions, with the basic idea being that each of these "microdecisions" is an opportunity to do the right thing. They describe the approach as inculcating moral virtue through repetition during everyday work activities including engaging with stakeholders and communicating with others about safety and rights protections. Their table 3 considers four daily actions with corresponding micro-decisions and ethical significance as opportune moments for ethical practice which can be roughly correlated with the modes of the ACTWith model, including responsible use of another's code, just credit for code used, decisions to share created code, and considering misuse of code reminiscent of Vallor and colleagues' fifth step.

"Bridging" macro- and micro-ethics is a fundamental aspect of case-studies in applied ethics education (cf. Kline 2010, "Engineering Case Studies: Bridging macro and micro ethics"). The focus of such exercises is to make ethics part of the everyday work and information flow. Cases immerse students in different roles in realistic, sometimes controversial, decision-making situations that they might encounter on a job. The aim in making ACTWith model dynamics transparent for practice during class exercises is to offer an easily accessible and memorable model for metacognitive self-appraisal in self-cultivation of target reactive attitudes through the professional life, with habitual employment potentiated through independent assignments including presentations, case analyses and progress interviews through appropriately structured course progress.

4.3 Technical application: UAVs

One upshot of the ACTWith education is that it represents a grounding framework for educators to interpret and apply complimentary programs in different degree programs. Regardless of focal area, moral cognitive dynamics remain constant. Understanding essentials affords insight into emerging as well as established efforts in ethics education, such as reviewed in this paper.

ACTWith education trains common conceptual language and interactional dynamics which facilitate communication across disciplines regardless of area-specific representations and differences in background after students leave university and enter the workforce. This aspect of the ACTWith program is especially important in inter- and trans-disciplinary research and development contexts of technologies involving multinational teams and global institutions, for example, with teams from different areas working together in product development and innovation for AI advisory systems deployed worldwide, rapidly and in competition with regulatory oversight lagging behind.

Consider the following coursework example concerning unmanned areal vehicles (UAVs). During a graduate course in Ethics of Technology at the University of Twente in 2019, students were asked to consider Boyd's famous OODA loop in the context of machine ethics (cf. Moor 2006) and machine autonomy of UAVs or "drones" as discussed in Marra and McNeil's "Understanding 'The Loop': Humans and the next drone generations" (2012). Boyd's OODA loop in its original form is a simple four-step input-output process; Observe (open to situational information), Orient (interpret situational information in terms of values and goal conditions), Decide (determine optimal action), Act (and re-enter the loop sampling situational information so affected). Originally considered from the perspective of human pilots during wartime combat, the OODA loop Orientation step extends to all personal knowledge including genetic heritage and cultural traditions.

Marra and McNeil (2012) approach the idea of machine autonomy from the perspective of designers and engineers. They distinguish between autonomy and automation. Current and anticipated robots are automated, as goal conditions are given and operations regardless may be interrupted by operators at any time."True autonomy" involves control over ends as goal conditions. Increased autonomy involves decreasing need for operator interaction to "navigate" "environmental uncertainty" and the "level of assertiveness" to enact selected actions to achieve goal conditions when unexpected conditions arise. They associate increasing autonomy with automation of the OODA loop, from observation to planning, limited by assertiveness, and illustrate with a robot that is "stuck" in a decision loop with no further action toward goal conditions possible. Though able to navigate, without operator interaction, the machine may abort. An assertive robot might take a risk, executing an action with a low success estimation in order to get itself "unstuck" facilitating further action toward goal conditions (discussion p 21-2). "True autonomy" on the other hand is not programmed into machines. Marra and McNeil argue that humans will remain "in the loop" of decision-making for this reason, assigning goal conditions regardless of other measures of autonomy.

One concern with humans in the loop of UAVs is lasting effects of trauma pursuant to lethal engagements. 2019 Ethics and Technology students were confronted with headline reports of post-traumatic stress in drone operators as prevalent as in combat soldiers (including https://www.npr.org/2017/04/24/525413427/for-drone-pilots-warfare-may-be-remote-but-the-trauma-is-real and https://cimsec.org/drone-pilots-statistically-front-lines/). The ACTWith model structured discussion, considering operator perspectives. Students may consider sources of operator stress

leading to rumination over lethal scenarios with characteristic depressive symptoms. For example, counterintuitively, the distancing of the mediating technology corresponds with an unusual absence of affective information, inviting rumination in efforts to reconstruct these missing aspects of the experience. Such an illustration underscores the value of empathy.

The OODA loop compared with the ACTWith cycle helps to clarify differences between human and machine autonomy. OODA loop dynamics were compared with ACTWith model dynamics in the context of Moor's (2006) levels of moral agency. Correlations with the OODA loop were drawn with Andersen and colleagues' (2008) "central circuit of the mind" which describes situational awareness (drawing also from memory), representational composition of actionable alternatives, end selection, and motor processing as an information processing loop involving four regions or "modules" of the perception-action processing cycle of a biological brain. The ACTWith model emphasizes moral affect through affective mirroring, feeling as-if the source of input informing action planning, absent in the OODA loop and de-emphasized in the perception-action central circuit.

Marra and McNeil's (2012) discussion offers opportunities for class exercises applying the ACTWith model in contexts of autonomy, emphasizing that as human beings more or less selectively open and close to information, they embody these patterns as routine prejudice. Where Marra and McNeil emphasize the speed with which machines can cycle through the OODA loop, updating in light of situational information and adjusting action plans accordingly, in order to give increasingly automated armed forces advantages over adversaries, the ACTWith model is a model of moral cognition. Where Boyd's Orientation step interprets input in terms of enculturated values, to identify a threat, ACTWith open-open emphasizes informing and revising prior established associations on the basis of other-situational information bottom-up. Ethical end-selection and action-planning proceed from understanding so informed, from the perspective of the shared situation.

Advanced courses for ethics educators might consider mirroring processes and their development, consider cases of maldevelopment of such processes and dysfunctional dynamics, e.g. psychopathy. Should we worry if a drone operator doesn't feel stress with remote murder? Technical courses might focus on machine autonomy by considering the Orientation step of the OODA loop in terms of cultural heritage including religious values or guiding ethical principles, e.g. Kant's categorical imperative. Students of any level of sophistication may be challenged to redraw the OODA loop as a moral information process, and to consider under what conditions"true" machine autonomy may be desirable, and how it may be programmed. Throughout, ACTWith model dynamics are a common touchstone holding machine and human information processing together for constructive comparison.

Like Kouprie and Visser, the aim is not to cycle through situational updates as quickly as possible in order to respond to cues with competitive advantage, but rather to "wander around" the shared object environment as-if one's own in order to ground that shared perspective. However, there is competitive advantage in empathy training. Brief considerations follow.

4.4 Parallel efforts: Moral Imagination training at Google

The ACTWith education offers a window into correlative accounts across disciplines and is suited for advanced coursework in ethics and technology. It is also useful in interpreting reported recent efforts internal to private industry. Lange and colleagues (2023) describe "moral imagination" training at Google for two years involving a four-step cycle that corresponds directly with ACTWith dynamics. They advocate for a "responsibility shift" from management removed from product development to individual project team members and upwards (compare Schuett et al. 2023; this paper section 4.1). Their work is motivated by the recognition of a regulatory "vacuum" in which powerful AI and other emerging technologies are developed. The basic idea is that engineers and designers should appreciate these contexts from user and other stakeholder perspectives during product ideation and design, bottom-up, and that current practices do not train relevant capacities.

The paper applies four steps in different analyses (appended, with ACTWith notation). First, technology companies have a practical responsibility to consider social contexts in terms of

which their products are used (open to other situations, stakeholder engagement, o/c). Second, reminiscent of OODA loop Orientation, enculturated values determine how well that practical responsibility is met (bottom-up informing prior understanding through active empathy, capacity for moral imagination, o/c -> o/o). Currently, Lange and colleagues note a "gap" in project team training to coordinate engineers around this responsibility (o/o -> c/o) and developed their moral imagination training program to address this emerging need (c/o -> c-c, in publishing their report and setting themselves up for feedback c/c -> o/c, ...). Moral imagination involves "creative" imagination of other perspectives to reveal otherwise hidden aspects of a given situation (e.g. negative outcomes for neglected stakeholders). The aim of training is to make moral imagination in the work context routine, effecting their target "role obligation shift" (section 3.1; compare the aim of the ACTWith education as described in this paper).

Moral imagination involves three "key ethical abilities" (Lang et al. 2023, Sect. 3.2) that map directly to ACTWith modes (Fig. 1, this paper). "Ethical Awareness" involves opening to salient moral information in order to inform colleagues about perceived risks or value violations (o/c -> o/o -> c/o -> c/c). Moreover, commitment to expanding this ability involves a meta-level understanding of values and their implications for product development, in this way focusing ethical awareness around the practical responsibility to social factors motivating the role obligation/responsibility shift (ACTWith model metacognition). Moral imagination training makes this obligation explicit. "Ethical Deliberation and Decision-making" (o/o -> c/o) focuses on forward action-planning with considerations including possible principled conflicts complimentary to pragmatic or profit oriented utility calculations. "Ethical Commitment" involves plans for forward action guiding work tasks including product development and research (c/o -> c/c).

Lange and colleagues describe their Moral Imagination workshops in four steps, as well (Sect. 3.3). "Reflection" considers prior value associations and ideal outcome situations - "a world where the technology has been successfully deployed" - and challenges teams to test current plans against shared values, making these explicit in the process (c/c -> o/c). "Expansion" (o/c -> o/o) involves considering neglected stakeholder perspectives. They describe an exercise involving unexpected consequences of successful deployment years in the future. Trainees are challenged to adopt different perspectives, and take these perspectives in role-play. The aim of the exercise is to illustrate that personal perspectives are limited, to emphasize the value of diverse input, and to train sensitivity to potential negative consequences. "Evaluation" (o/o -> c/o) involves explicit reasoning such as in the form of ethical theories and principles applied to moral decision contexts. Lange and colleagues report that moral theories are presented"schematically" (compare ACTWith diagrammatically) and teams are presented tools for weighing different values to "resolve" tensions between them, making relative commitments explicit. "Action" (c/o -> c/c) in their workshop context involves further reflection over training eventuating in actionable value statements, in this way orienting team members around the practical responsibility to develop technologies while sensitive to their broad implications over the long term.

5. Summary discussion, limitations and future work

Institutions change as information processing changes, as officers therein are retrained or replaced and communications between them opened or closed. In the context of engineering ethics education, to align institutional with social and professional objectives, Chan and Lee (2021) like Shukla and colleagues (2023) stress the need for adequate teaching faculty; there are not enough of them. The first problem is resources (institutional and social support) and the second motivation (teachers doing the work to develop necessary backgrounds, then spending their lives becoming better teachers). A third is talent. A fourth is lack of integrative frameworks for systematic applied ethics training of that talent. The ACTWith model integrates similar efforts under a common umbrella, appropriate for applied ethics training in diverse degree programs, directly for practitioners, and for teaching requisite ethics teachers how to teach applied ethics.

The ACTWith model trains moral reasoning through perspective-taking exercises in practical contexts from whistleblowing to institutional reform, and may be useful in speeding Morley and colleagues' (2023) cultural shift in ethics education. Their approach relies on Habermas' (2018) theory of legitimacy involving influence from the marginalized and disaffected, a dynamic that may be expedited by routinely empathic perspective-takers (ACTWith "REPs"), implying potential for ACTWith application at scale.

Immersive exemplar studies including whistleblower cases afford relatable contexts for demonstration of metacognitive tools such as the ACTWith model. Different theories can be illustrated using ACTWith dynamics. The model fits course design principles from Sanz et al. (2023) and holistic ethics education on the model of Han's (2023) *phronesis*. As a tool for empathy training through perspective-taking, the ACTWith model represents affective and effective mirroring (as in Thioux, Gazzola and Keysers 2008; Bastiaansen, Thioux and Keysers 2009; cf. Wu et al. 2022) reflected in the moral psychology of Kouprie and Visser's (2009) empathic design. Initial ACTWith research focused on insular gating dynamics and moral repugnance, consistent with Menon and Uddin's (2010) saliency switching. Connections with different theories are made in White (2014). Further connections are beyond the scope of this paper (such as with Fermin et all 2022).

The ACTWith model is a perception action model of empathy as described in Gunatilake et al. (2023). Gunatilake and colleagues assessed different empathy models in the context of software engineering. They consider empathy a "competitive advantage" improving end-user service, and emphasize the value of empathy training in designing systems that find better solutions by focusing-in on what is most salient to human beings. However, they leave best methods and models to represent empathy undetermined. This paper has considered the ACTWith model as the best method for empathy training.

As a model of moral cognition, the ACTWith model is limited. The model is most appropriate for perspective-taking exercises, and can be easily applied to information processing dynamics at larger scales of organization (cf. White 2012, section 7). The model is appropriate for metacognitive self-appraisal and modulation of cognitive dynamics through practice, and as a generic framework for interpreting correlative programs in different disciplines. As a general model, it has not been developed in detail within a specific expertise. The model relies on complimentary accounts in different areas to support ACTWith education in different contexts, such as VCSM in business and Vallor and colleagues' toolkit in engineering and design contexts, or correlations with the OODA loop in contexts involving UAVs. Once a practitioner is fluent in ACTWith dynamics, the model can be spontaneously applied to current events to charge student interest in the classroom, and to develop training programs such as Lang's.

ACTWith publications are limited. Most ACTWith model-related work has been for the classroom since 2010. Though applied in design of more than thirty Philosophy and Ethics courses on two continents for thousands of students, since, the present report is the first publication on ACTWith model related teaching. A monograph has been prepared for publication, and course workbooks including case-based exercises are being considered. Current practice relies on course feedback, which has been positive, but there has been no study on ACTWith model effectiveness as a teaching tool. Ideally, opportunities to design structured courses with surveys in assessment of empathy training efficacy, perhaps following Sanz and colleagues, can be pursued with results communicated in future reports. Opportunities for popular engagement through public lectures considering cases in ACTWith terms might also present; the model is accessible, the subject current, the moniker memorable. To date, there has been no effort to advertise the ACTWith model outside of existing publications and implicitly through students and colleagues. A tenured professorship should be the most suitable platform for this work.

Future work should meet emerging needs in common efforts at providing adequate applied ethics education for current and future generations, including teaching teachers to teach ethics. As such efforts as these and from other researchers reviewed in this paper gain momentum, as ethics is operationalized and embedded in the daily work and information flows of universities and private enterprises, with business reconceived pursuant to Freeman's final challenge, whistleblowing may be made obsolete, and with historical cases remaining worthy of consideration for their cautionary lessons, nonetheless.

6. Conclusion

If a man say, I love God, and hateth his brother, he is a liar: for he that loveth not his brother whom he hath seen, how can he love God whom he hath not seen?

- 1 John 4:20 (King James version)

Ethics education is a critical moment for intervention in technology design, especially in contexts of rapidly emerging technologies when innovation outpaces regulatory capacities. By training to feel out and forecast possible future situations from perspectives of extant and especially marginalized stakeholders, issues such as justice and dignity become accessible in a non-"check-listy" way. They mean something. The ACTWith model was designed to do this work.

Given increasing momentum in operationalizing ethics, and with significant formal overlap with popular programs in different domains, the usefulness of the ACTWith model account is evident. As with the ACTWIth education, the goal of Lange and colleagues' (2023) moral imagination training varies depending on institutional objectives; but, the grounding method remains the same. These grounds are exposed in this paper. The ACTWith model accounts for various initiatives in common cognitive dynamics, potentiating a unification of similar efforts in a single simple approach.

This paper considered how empathy training contributes to holistic education of lifelong philosopher engineers. The ACTWith focus on situations emphasizes what is missing in abstract principles. Recalling Kouprie and Visser (2009), understanding without judging and without agenda is key to empathy, and to any constructive social reform dependent on its exercise (o/o, love). An AI might excel in demonstrating such a capacity. Until these arrive, we must rely on ourselves and moral exemplars including whistleblowers and religious heroes to exemplify the civic courage necessary to effect needed changes.

Resources are limited. If the aim remains to entrain highest-level moral reasoning through engineering ethics education pursuant to Morley and colleagues' (2023) cultural shift, the situation is difficult. For one thing, the majority of practicing teachers are poorly prepared. Most were not hired to teach, few trained to teach well, and fewer still in a style suiting an appropriately shifted culture. Why should teaching be taken seriously by philosophy professors, when education remains derivative of focal interests e.g. securing grants, becoming dean? The result is faculty unable to support necessary holism or generate authentic interest, and a culture in need of shifting. The immediate way forward is to look to the rather limited pool of practitioners for whom holistic ethics education is their dedicated vocation and who remain up to requisite cultural-shifting, independent of dominant enclaves under-performing in the classroom to this point. The rarity of such talented teachers underscores the need for an integrative approach to training future teachers. The ACTWith model was designed to do this work, as well.

The ACTWith model represents a general theory appropriate for training teachers to design class activities and course curricula appropriate to students of different levels in different disciplines. Such training should prove important in rapidly emerging contexts involving AI and related technologies, where the idea is to get ahead of rapid changes to avoid otherwise unforeseeable negative impacts while optimizing development for the common good.

Works consulted:

Anderson J, Fincham J, Qin Y, Stocco A (2008) A central circuit of the mind. Trends in Cognitive Sciences 12(4):136-143. https://doi.org/10.1016/j.tics.2008.01.006.

Ball B, Koliousis A (2023) Training philosopher engineers for better AI. AI & Soc 38:861–868. https://doi.org/10.1007/s00146-022-01535-7

Bastiaansen JA, Thioux M, Keysers C (2009) Evidence for mirror systems in emotions. Philosophical Transactions of the Royal Society B 364(1528):2391–2404. https://doi.org/10.1098/rstb.2009.0058

Bellaby RW (2018) The ethics of whistleblowing: Creating a new limit on intelligence activity. Journal of International Political Theory, 14(1):60–84. https://doi.org/10.1177/1755088217712069

Berg KT (2020) The Ethics of Whistleblowing. Journal of Media Ethics 35(1):60-64. https://doi.org/10.1080/23736992.2020.1702671

Bezuidenhout L, Ratti E (2021) What does it mean to embed ethics in data science? An integrative approach based on microethics and virtues. AI & Soc 36, 939–953. https://doi.org/10.1007/s00146-020-01112-w

Bretz S, Sun R (2018) Two Models of Moral Judgment. Cogn Sci 42:4-37. https://doi.org/10.1111/cogs.12517

Ceva E, Bocchiola M (2019) Is Whistleblowing a Duty. Polity Press, Cambridge

Ceva E, Bocchiola M (2020) Theories of whistleblowing. Philosophy Compass 15:e12642. https://doi.org/10.1111/phc3.12642

Chan CKY, Lee, KKW (2021) Constructive alignment between holistic competency development and assessment in Hong Kong engineering education. Journal of Engineering Education, 110(2):437–457. https://doi.org/10.1002/jee.20392

Conlon E, Martin DA, Bowe B (2018) Holistic Engineering Ethics? EESD 2018 Proceedings Creating the Holistic Engineer, pp 52-60. https://pure.tue.nl/ws/portalfiles/portal/168478777/EESD_Holistic_Engineering_Ethics.pdf Accessed 21 July 2023

Delmas C (2015) The Ethics of Government Whistleblowing. Social Theory and Practice 41(1):77–105. https://doi.org/10.5840/soctheorpract20154114

Dutta, M. (2020). Introduction: A Framework for Communicating Social Change. In: Communication, Culture and Social Change. Palgrave Studies in Communication for Social Change. Palgrave Macmillan. https://doi.org/10.1007/978-3-030-26470-3_1

Fermin ASR, Friston K, Yamawaki S 2022 An insula hierarchical network architecture for active interoceptive inference. R. Soc. Open Sci. 9: 220226. https://doi.org/10.1098/rsos.220226

Floridi L (2018) Soft ethics, the governance of the digital and the General Data Protection Regulation. Phil. Trans. R. Soc. A 376:20180081. https://doi.org/10.1098/rsta.2018.0081

Floridi L (2019) Translating Principles into Practices of Digital Ethics: Five Risks of Being Unethical. Philos. Technol. 32:185–193. https://doi.org/10.1007/s13347-019-00354-x

Freeman RE (1984/2010) Strategic Management: A Stakeholder Approach. Cambridge University Press, Cambridge

Freeman RE (2012) Stakeholder Management and Reputation. In: Values and Ethics for the 21st Century (p 363–381). BBVA Open Mind https://www.bbvaopenmind.com/en/books/values-and-ethics-for-the-21st-century/ Accessed December 15, 2023

Fuster JM (1990) Prefrontal cortex and the bridging of temporal gaps in the perception-action cycle. Ann N Y Acad Sci. 608:318-29

Gallese V, Keysers C, Rizzolatti G (2004) A unifying view of the basis of social cognition. Trends Cogn Sci. 8(9):396-403. doi:10.1016/j.tics.2004.07.002

Gunatilake H, Grundy J, Mueller I, Hoda R (2023) Empathy models and software engineering — A preliminary analysis and taxonomy. Journal of Systems and Software 203:111747. https://doi.org/10.1016/j.jss.2023.111747

Habermas J (2018) Between facts and norms: Contributions to a Discourse Theory of Law and Democracy. John Wiley & Sons.

Han H (2017) Neural correlates of moral sensitivity and moral judgment associated with brain circuitries of selfhood: A meta-analysis. Journal of Moral Education, 46(2):97–113. https://doi.org/10.1080/03057240.2016.1262834

Han H (2023) Considering the Purposes of Moral Education with Evidence in Neuroscience: Emphasis on Habituation of Virtues and Cultivation of Phronesis. Ethical Theory and Moral Practice. https://doi.org/10.1007/s10677-023-10369-1

Han H, Dawson KJ (2023) Relatable and attainable moral exemplars as sources for moral elevation and pleasantness. Journal of Moral Education, 1–17. https://doi.org/10.1080/03057240.2023.2173158 Accessed 21 July 2023

Han H, Workman CI, May J, Scholtens P, Dawson KJ, Glenn AL, Meindl P (2022) Which moral exemplars inspire prosociality? Philosophical Psychology 35(7):943–970. https://doi.org/10.1080/09515089.2022.2035343

Heidegger M (1996) Being and Time: A Translation of Sein und Zeit. SUNY Press.

Hess JL, Beever J, Zoltowski CB, Kisselburgh L, Brightman AO (2019) Enhancing engineering students' ethical reasoning: Situating reflexive principlism within the SIRA framework. J Eng Educ. 108:82– 102. https://doi.org/10.1002/jee.20249

Howecroft J, Mercer K (2022) where we are: understanding instructor perceptions of empathy in engineering education. Proceedings 2022 Canadian Engineering Education Association (CEEA-ACEG22) Conference CEEA-ACEG22, York University; June 19 – 22, 2022; Paper 72

King James Bible. (1769). King James Bible Online. https://www.kingjamesbibleonline.org/

Kisselburgh L, Zoltowski CB, Beever J, Hess JL, Iliadis AJ, Brightman AO (2014) Effectively Engaging Engineers in Ethical Reasoning about Emerging Technologies: A Cyber-Enabled Framework of Scaffolded, Integrated, and Reflexive Analysis of Cases Paper presented at 2014 ASEE Annual Conference & Exposition, Indianapolis, Indiana. 10.18260/1-2--20349

Kline RR (2010) Engineering Case Studies: Bridging Micro and macro ethics. IEEE Technology and Society Magazine, 29(4):16–19. https://doi.org/10.1109/mts.2010.939188

Koenigs M, Young L, Adolphs R, Tranel D, Cushman F, Hauser MD, Damasio AR (2007). Damage to the prefrontal cortex increases utilitarian moral judgements. Nature, 446(7138):908–911. https://doi.org/10.1038/nature05631

Kouprie M, Visser FS (2009) A framework for empathy in design: stepping into and out of the user's life. Journal of Engineering Design 20(5):437–448. https://doi.org/10.1080/09544820902875033

Kolb DA (1984) Experiential Learning: Experience as the Source of Learning and Development. Prentice-Hall, Englewood Cliffs, New Jersey

Lange B, Keeling G, McCroskery A, Zevenbergen B, Blascovich S, Pedersen K, Lentz A, Aguera y Arca B (2023) Engaging Engineering Teams Through Moral Imagination: A Bottom-Up Approach for Responsible Innovation and Ethical Culture Change in Technology Companies. AI Ethics. https://doi.org/10.1007/s43681-023-00381-7

Magnani L (2011) Understanding Violence. Springer, Heidelberg, Berlin. https://doi.org/10.1007/978-3-642-21972-6

Marra W, McNeil S (2012) Understanding “The Loop”: autonomy, system Decision-Making, and the next generation of war machines. Harvard Journal of Law and Public Policy 36(3). Available at SSRN: http://dx.doi.org/10.2139/ssrn.2043131

Martin DA, Conlon E, Bowe, B (2021a) A Multi-level Review of Engineering Ethics Education: Towards a Socio-technical Orientation of Engineering Education for Ethics. Sci Eng Ethics 27:60 https://doi.org/10.1007/s11948-021-00333-6

Martin DA, Conlon E, Bowe, B (2021b) Using case studies in engineering ethics education: the case for immersive scenarios through stakeholder engagement and real life data, Australasian Journal of Engineering Education 26(1):47-63. https://doi.org/10.1080/22054952.2021.1914297

Martin DA, Bombaerts G, Horst M, Papageorgiou K, Viscusi G (2023) Pedagogical Orientations and Evolving Responsibilities of Technological Universities: A Literature Review of the History of Engineering Education. Sci Eng Ethics 29(40). https://doi.org/10.1007/s11948-023-00460-2

Menon V, Uddin LQ (2010) Saliency, switching, attention and control: a network model of insula function. Brain Structure & Function, 214(5–6):655–667. https://doi.org/10.1007/s00429-010-0262-0

Moor JH (2006) The nature, importance, and difficulty of machine ethics. IEEE Intelligent Systems, 21(4):18–21. https://doi.org/10.1109/mis.2006.80

Morley J, Floridi L, Kinsey L, Elhalal A (2020) From what to how: an initial review of publicly available AI ethics tools, methods and research to translate principles into practices. Sci Eng Ethics 26(4):2141–2168. https://doi.org/10.1007/s11948-019-00165-5

Morley J, Elhalal A, Garcia F, Kinsey L, Mökander J, Floridi L (2021) Ethics as a service: a pragmatic operationalisation of AI ethics. Mind Mach 31(2):239–256. https://doi.org/10.1007/s11023-021-09563-w

Morley J, Kinsey L, Elhalal A, Garcia F, Ziosi M, Floridi L (2023) Operationalising AI ethics: barriers, enablers and next steps. AI & Society 38(1):411–423. https://doi.org/10.1007/s00146-021-01308-8

Near JP, Miceli M (1985) Organizational dissidence: The case of whistleblowing. Journal of Business Ethics 4(1):1–16. https://doi.org/10.1007/BF00382668

Ritzer G (1990) Metatheorizing in sociology. Sociol Forum 5:3–15. https://doi.org/10.1007/BF01115134

Rivas DA, Husein S (2022) Empathy, persuasiveness and knowledge promote innovative engineering and entrepreneurial skills. Education for Chemical Engineers, 40:45–55. https://doi.org/10.1016/j.ece.2022.05.002

Sanz C, Coma-Roselló T, Aguelo A, Álvarez P, Baldassarri S (2023) Model and Methodology for Developing Empathy: An Experience in Computer Science Engineering. IEEE Transactions on Education 66(3):287-298

Schuett J (2023) AGI labs need an internal audit function. https://doi.org/10.48550/arXiv.2305.17038 Accessed 21 July 2023

Shalev I, Eran A, Uzefovsky F (2023) Empathic disequilibrium as a new framework for understanding individual differences in psychopathology. Front. Psychol. 14:1153447. doi:10.3389/fpsyg.2023.1153447

Shukla B, Soni KM, Sujatha R, Hasteer N (2023) Roadmap to inclusive curriculum: a step towards Multidisciplinary Engineering Education for holistic development. Journal of Engineering Education Transformations 36(3):133-144

Smith A (1759) The theory of moral sentiments. Oxford Text Archive, http://hdl.handle.net/20.500.12024/3189 Accessed 21 July 2023

Sun R (2001). Duality of the Mind: A Bottom-up Approach Toward Cognition (1st ed). Psychology Press, New York. https://doi.org/10.4324/9781410604378

Thiel CE, Connelly S, Harkrider L, Devenport LD, Bagdasarov Z, Johnson JF, Mumford MD. Case-based knowledge and ethics education: improving learning and transfer through emotionally rich cases. Sci Eng Ethics. 2013 Mar;19(1):265-86. doi: 10.1007/s11948-011-9318-7

Thioux M, Gazzola V, Keysers C (2008) Action Understanding: How, What and Why. Current Biology 18(10):431-434

Trbušić H (2020). Holistic education: the social reality of engineering. The Journal of Education, Culture, and Society, 4(2), 227–238. https://doi.org/10.15503/jecs20132.227.238

Umilta MA, Kohler E, Gallese V, Forgassi L, Fadiga L, Keysers C, Rizzolatti G (2001) I Know What You Are Doing: A Neurophysiological Approach. Neuron 31:155-165

Vallor S (2016) Technology and the virtues - a philosophical guide to a future worth wanting. Oxford University Press, Oxford

Vallor S, Green B, Raicu I (2018). Ethics in Technology Practice. The Markkula Center for Applied Ethics at Santa Clara University. https://www.scu.edu/ethics/

Walther J, Miller SE, Sochacka N (2017) A Model of Empathy in Engineering as a Core Skill, Practice Orientation, and Professional Way of Being. Journal of Engineering Education 106(1):123–148. https://doi.org/10.1002/jee.20159

Walther J, Brewer MJ, Sochacka NW, & Miller S (2019) Empathy and engineering formation. Journal of Engineering Education 109(1):11–33. https://doi.org/10.1002/jee.20301

Wang Q, Zhang W, Zhu Q (2015) Directing engineering ethics training toward practical effectiveness. Technology in Society 43:65–68. https://doi.org/10.1016/j.techsoc.2015.02.004

White JB (2006) Conscience: toward the mechanism of morality. Dissertation, University of Missouri-Columbia. https://doi.org/10.32469/10355/4327

White J (2010) Understanding and Augmenting Human Morality: An Introduction to the ACTWith Model of Conscience. In: Magnani L, Carnielli W, Pizzi C (eds) Model-Based Reasoning in Science and Technology. Studies in Computational Intelligence, vol 314, Springer, Berlin, Heidelberg, pp 607-621. https://doi.org/10.1007/978-3-642-15223-8_34

White J (2012) An Information Processing Model of Psychopathy. In: Fruili & Veneto (eds), Moral Psychology. Nova Science, pp 1-34

White J (2013) A general theory of moral agency grounding computational implementations: The ACTWith model. In: A. Floares (ed), Computational Intelligence. Nova Science, pp 163-209

White J (2014) Models of Moral Cognition. In: Magnani, L (ed) Model-Based Reasoning in Science and Technology. Studies in Applied Philosophy, Epistemology and Rational Ethics, vol 8. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-37428-9_20

Wicker, Bruno, Christian Keysers, Jane Plailly, Jean-Pierre Royet, Vittorio Gallese and Giacomo Rizzolatti (2003) Both of Us Disgusted in My Insula: The Common Neural Basis of Seeing and Feeling Disgust. Neuron 40: 655-664

Wolfand, JM, Bieryla KA, Ivler CM, Symons JE (2022) Exploring an Engineer's Role in Society: Service Learning in a First-Year Computing Course. IEEE Transactions on Education, 65(4):568-574. doi: 10.1109/TE.2022.3148698

Wu WY, Cheng Y, Liang KC, Lee RX, Yen CT (2022) Affective mirror and anti-mirror neurons relate to prosocial help in rats. iScience. 26(1):105865. doi: 10.1016/j.isci.2022.105865

Young L, Koenigs M (2007) Investigating emotion in moral cognition: a review of evidence from functional neuroimaging and neuropsychology. Br Med Bull. 84:69-79. doi: 10.1093/bmb/ldm031