TEACHING ETHICS THROUGH THE BACK DOOR? – EMPLOYING IDEAS FROM ASSEMBLAGE THEORY TO FOSTER A RESPONSIBLE INNOVATION MINDSET

C. Herzog[1]
Ethical Innovation Hub, University of Lübeck
Lübeck, Germany
https://orcid.org/0000-0003-2513-2563

H. Diebel-Fischer
Dresden University of Technology, Faculty of Arts, Humanities and Social Science
Dresden, Germany
https://orcid.org/0000-0002-6505-3563

Conference Key Areas: *Ethics in Engineering Education; Fostering Engineering Education Research*
Keywords: *Engineering Ethics; Assemblage Theory; Responsible Innovation; Socio-Technical Systems; Artificial Intelligence*

ABSTRACT

Adding ethics courses to engineering curricula seeks to equip students with the critical mindset that enables careers committed to serving humanity. Yet, the knowledge of ethical theories is neither a necessary, let alone sufficient condition for being good [1]. There is no automatism that translates ethical knowledge into action, overriding attitudes that were developed during the enculturation of a student. However, we deem teaching assemblage theory a promising means to achieve a sustained commitment to responsible innovation practice. We base our argument on assemblage theory's (cf. [2, 3]) capacity to conceptualize the interplay of human actors and technological artefacts in terms of dynamic evolutionary systems. The notion of an assemblage as a collection of potentially heterogeneous elements that—despite displaying consistency—remains malleable through reorganization, interconnection and, (re-)attribution forms the ontological basis that guides a conceptual approach to thinking in-between the extremes of technological

[1] *Corresponding Author: C. Herzog, Christian.Herzog@uni-luebeck.de*

doi:10.5821/conference-9788412322262.1390

determinism and social constructivism. Information algorithms, e.g., can be regarded as having the power to facilitate ethical action as part of a larger assemblage [4] and artificial intelligence can arguably only be understood as “trustworthy” within socio-technological systems in which a shared responsibility realizes both epistemic and moral conditions for trust [5].

Ultimately, we intend engineering students to realize the extent of their influence on the world and, therefore, their responsibility for contributing to a prosperous community. Thus, ethics is not only taught by conveying its classical normative theories but rather explored by discovering the entangledness of technology and society.

1 INTRODUCTION—DRAWBACKS OF COURSES IN ETHICAL THEORY

Ethics has always been part of engineering and engineering education at universities and elsewhere—but mostly implicitly, to that extent that everyone who is involved, i.e., educators as well as students, is always endowed with an ethical configuration resulting from their respective socialization process. Our upbringing as well as the experiences we make have a further impact on how we assess moral questions and make judgments. So, if everybody already has an implicit understanding of what is right and wrong that reflects their personal views, why bother using a part of the already limited time for courses at university to add ethics education to an engineering curriculum? What difference can an ethics course make? And what difference would it make if it were based on assemblage theory, cf. [2, 3]?

First, we should look at how a course that teaches ethics by conveying ethical theory works, what assumptions it is based on, and what its (possible) drawbacks are. This will shed light on why we propose a different way which circumvents the pitfalls in ‹classical› ethics courses. Ethical theories, be they deontological (based on duties) or consequentialistic, provide normative frameworks. These start out at a moral problem and help its users to identify a path of action to take, which—according to the respective theory—is identified as either ‹good,› ‹right,› or ‹one's duty.› That is, an action is demonstrated as morally justified by outlining a sound argument.

On the other hand, ethical theory tends to be perceived as eluding the practical side of life. The criticisms voiced regarding the effectiveness of ethics education take place on (at least) two levels: (1) Skeptics of the teachability of ethics in general are claiming that one either is a moral person or not, as this is the result of a socialization process. (2) Skeptics of the knowledge of ethical theories being an asset to a person's ability to better deal with moral problems are claiming this knowledge does not contribute to the formation of moral judgments as it only serves as an *ex post* justification of decisions made without an ethical theory [6, p. 37–38].

Both, Johnson and Fischer, do not deny the teachability of ethics. Johnson thinks that skeptics in (1) have «oversimplistic notions of ethics and of human behavior» and declares «knowledge [of] codes and standards,» «skill,» «reasoning,» and «motivation» the core components of engineering ethics [1, p. 64]. Fischer pursues a different approach and points to evidence from courses in applied ethics for doctors and nurses, which were perceived as being disconnected from the needs of medical practice because the appeal to normative ethics fails by giving the justification of actions precedence over understanding the «morally significant situation». While according to Fischer we recognize such a situation and know what to do without going step by step through a (quasi-)algorithmic judgment process, Johnson regards it a teachable and thus learnable «skill» to «identify ethical issues» [1, p. 64].

Johnson is aware that ethics education will not automatically make people «good» in general but claims that this increased the probability that engineering students will be «better prepared to handle the ethical issues that arise in their professional lives» [1, p. 64]. Based on her assumptions that the knowledge of policies and professional ethics codes equips future engineers with a compass that suffices to handle ethical problems, hand back the verdicts of having «oversimplistic notions of ethics» back to her. Relying on such codes' results and their application is little different from relying on the knowledge of frameworks delivered in theories of ethics. Both are teachable, yet both represent a technical approach to ethics, meaning that a tool is provided to ‹solve› a problem. While theories of ethics are often employed this way, it is better to use them as concepts for analyses and approach real world problems differently. This however exposes a gap between ethical theory and moral practice.

2 AN ALTERNATIVE PROPOSAL—USING ASSEMBLAGE THEORY FOR ENGINEERING ETHICS EDUCATION

To fill this gap, we propose teach engineering in ethics not by ethical theories and their application to moral problems, but in a way that acknowledges the complexity of the world and society as socio-technical conglomerates, or *assemblages*.

2.1 A Brief Characterization of Assemblage Theory

We would like to precede a brief characterization of assemblage theory by disclaiming that our portrait may strike scholars of Deleuze and Guattari as overly simplistic. To the best of our knowledge, assemblage theory eludes a concise summary, perhaps with the exception of [7]. Scholars like Buchanan do, in fact, reject the very idea that something like this should even be pursued. At the core of Buchanan's argument lies a characterization of assemblage theory as an unfinished project whose principles can be extended [3, p. 6], but for which it is impossible to give an exhaustive and model-like description, because this would betray its very project of working against mechanistic ways of perceiving reality [3, p. 5].

If we accept these premises, then whatever we can sketch here will not do assemblage theory as originally conceived of by Deleuze and Guattari in *A Thousand Plateaus* proper justice. However, aware of this, we will try to sketch the general concept, hoping to evoke an idea of why it can be useful and how. In doing so, like Buchanan, we will try to remain true to the original idea that extends beyond—and rejects core tenets of-—the notion of assemblages-as-systems-of-things due to DeLanda [2]. DeLanda takes that properties of a given assemblage are generated by its components, whereas, according to Buchanan, «desire» is primary, i.e., it is giving selected things the properties they have in an assemblage [3, p. 56].

Crucial to understanding assemblages is to realize that assemblages are not defined by their components, but rather that they are defined by their products [3, p. 47]. The virtual (think, e.g., concept) is actualized by the elements' relations and agents. Assemblages are «alive» in the sense that they do not disappear, even if the material things they arrange are removed [3, p. 60]. What matters most are the ideas and notions that remain. In that sense, at the core of assemblage theory as per Buchanan (and hence as per Deleuze and Guattari) is a reversal of the «virtual» and the «actual», which can be exemplified in capitalism: Advertisements based on particular characters from, e.g., movies would only be *virtual*, because they are unaffecting without the stories that *actually* matter to prospective buyers [3, p. 60].

An *assemblage* (an unfortunate translation of *agencement*), then, is an «arrangement or layout of heterogeneous elements» [7, p. 22]. In contrast to portraying unity described by an essence, understanding assemblages as arrangements emphases multiplicity and events. While one cannot extract parts from unities without destroying them (e.g. a heart from a body), assemblages allow for recombinations or removals of parts and are defined by the relations between these. Hence, an assemblage «constructs or lays out a set of relations between self-subsisting fragments» [7, p. 23]. Instead of being about essence, presupposing a static, defining finality, assemblages are dynamic constructs of contingent features.

All assemblages are defined by three kinds of features: their *Abstract Machine* (or conditions, C), their *Concrete Assemblage* (or elements, E), and their *personae* (or agents, A) [7, p. 24–28]: *The Abstract Machine* are the external relations holding the elements together. While the conditions are not tangible objects and thus abstract, the external relations are real. *The Concrete Assemblage* are the actual elements from which the abstract machine is composed. The relation between the concrete assemblage and the abstract machine is reciprocal; they are mutually co-adapting to each other. *The Personae* are the agents that connect the concrete elements according to their abstract relations and are immanent to the assemblage [7, p. 27].

All assemblages are arranged as combinations of basic political types: *Territorial Assemblages (TAs)* are arranged to divide the world into coded segments that define the «natural» norms of life in terms of «this is how things are done, how they have always been done», [7, p. 28–29]. *State Assemblages (SAs)* are arranged to unify all concrete elements in the assemblage [7, p. 30]. *Capitalist Assemblages (CAs)* are arranged such that conditions, elements and agencies form abstract quantities [7, p. 31] that can be treated as globally exchangeable [7, p. 32]. *Nomadic Assemblages (NAs)* are arranged such that conditions, elements and agencies can recombine to allow qualitative transformation and expansion of the assemblage [7, p. 33]

In summary, territorial assemblages are based on essentialist meanings, state assemblages allow centralized command, capitalist assemblages allow exchange in terms of generic quantities and nomadic assemblages allow for participation and self-management. All assemblages constantly change according to different kinds of change (or «deterritorialization» as the way in which assemblages continuously transform and reproduce themselves [7, p. 34]): *Relative negative* processes maintain and reproduce an established assemblage. *Relative positive* processes do not maintain or create a new assemblage. *Absolute negative* processes undermine all assemblages. *Absolute positive* processes create new assemblages.

The above is a concise and ultimately incomplete terminology used in structuring concepts from assemblage theory. Its use will be exemplified next in the context of algorithmic accountability.

2.2 An Example — Algorithmic Accountability

Algorithms are generalizable descriptions of methods designed to accomplish certain well-defined tasks. Ananny frames algorithms as «unstable objects of study» [4, p.

109], which require a framework in which to address ethical implications that can keep track of the dynamic relations. Assemblage theory is a promising candidate.

Algorithms promote quantitative and oftentimes (deliberatively) reductive views that allow for efficacy or increasing the efficiency of various routines. Algorithms can hence be characterized as drivers of *absolute positive processes* that typically create *capitalist assemblages*. However, while not be wrong, this characterization may also be too limited [4, p. 97]. For instance, it fails to capture that algorithms also facilitate new forms of human interaction. Be it through, e.g., match-making, encryption or compression, algorithms have led to qualitative breakthroughs in the conditions under which humans can communicate, organize themselves and act together. Accordingly, they can be conceived of as drivers of *absolute positive processes* creating *nomadic assemblages* as well. Like institutions, algorithms have the power to structure and influence behaviour [4, p. 99] and an ethical analysis of questions of accountability needs to account for all elements, relations and agents involved.

The *concrete assemblages* of algorithms in social media, for instance, consist of actual code of both the social media-based algorithm as well as of systems interacting with it, human practices, i.e., those of the developing individuals, users of social media, policy, and regulatory bodies, etc., as well as norms. A particular *abstract machine* might be identified in the notion of «clout» as the power or influence an institution or individual might exert on social media. *Personae* connecting elements of the concrete assemblage are statistics, aggregations and other abstract means bringing about effects contingent on the abstract machine.

Algorithms working on data, e.g., influence associations by statistical means previously beyond the recognition of humans. More broadly, «[p]eople that fail to leave data that can be categorized are effectively invisible to the … algorithm» [4, p. 101], while those that share data may be viable targets for adverts but may also be offered interactions that induce opportunities. Accordingly, algorithms facilitate processes that maintain, undermine, or create associations and, hence, corresponding assemblages. Much of the effects of social media algorithms strongly depend on how users appropriate a particular platform, similar to, but perhaps far extending beyond the capacity of citizens appropriating urban infrastructure to their needs. In addition, the effects do seldom originate from the utilization of a single social media platform alone, but rather depend on the interaction with further platforms of, e.g., commerce, such as ad-targeting, aggregate news outlets and search engines. Hence, locating the origin of, e.g., the societal effects of a single post going viral, is difficult and can only be possibly understood or traced by examining interactions between the concrete elements of the assemblage. Furthermore, in case of machine learning algorithms, effects may also depend on past and present interactions encoded by various sources of data over time.

Assemblage theory may be a difficult concept to fully grasp. However, we believe that the above sketch outlines that it provokes a way of thinking about relations and events. In Buchanan’s words «[t]he assemblage is intended to answer several types of question, ‹how?›, ‹why?›, ‹when?›, and not just a ‹what?› question,» [3, p. 13].

Ananny makes a point that more traditional perspectives from ethical theories, such as deontology, teleology and virtue ethics are at a loss for answering or critiquing algorithms appropriately. For instance, the effect of emerging categories would have to be assessed by outside standards in a deontological sense, by efficaciousness in a teleological sense, or by its alignment with virtue-based expectations in a virtue ethics sense. All of this is useful but overlooks that, e.g., an algorithm could bring about new collectives of ethical concern and that assemblages might incur transitions under which ethical assessments might alter. Ananny's argument for inquiring about the ethics of algorithms via assemblage theory, hence, hinges on a demand to go beyond lists of ethical guidelines requiring «transparency» [4, p. 109] or «justice», but to rather highlight the dynamics of relations and how these could bring about unethical states and conditions.

3 A PROPOSED COURSE DESIGN FOR HIGHER EDUCATION

Above we have elaborated on assemblage theory as a potential candidate for ethical analysis, whose use might mitigate the drawbacks of traditional ethical theory-based engineering ethics education. In the following, we will propose a course design rational and structure, arguing that using assemblage theory, it will be possible to educate engineers to routinely consider the ethical implications of their work «through the back door», i.e., implicitly through sensitizing for ethically relevant relations, conditions and events—in other words, through assemblages.

3.1 Rationale

To avoid any misunderstanding: by ‹employing assemblage theory› to foster a responsible innovation mindset, we do not seek to teach assemblage theory to create a fixed view on ethics. We also do not seek to teach assemblage theory as a kind of blueprint to be mapped onto any given situation, which would go against the intentions of its conceivers Deleuze and Guattari [3, p. 5]. Even though such an approach might be most familiar to engineering students, given the conceptual difficulties in providing for a concise and complete characterization of assemblage theory, it appears hardly possible to design a course that begins with outlining assemblage theory as such and then proceeds to apply it. Hence, to bring insights from assemblage theory to bear in an engineering course, lecturers might be best advised to keep explicit references to assemblage theory terminology to a minimum. In fact, even when scanning the (mostly sociological) literature covering aspects of the ethics of technology by means of assemblage theory, an explicit mentioning of assemblage theory-related terminology in terms of features such as *abstract machines* or political types such as *nomadic assemblages* is rarely found.

Accordingly, an approach to «teaching ethics through the back door» using assemblage theoretical concepts must succeed by conveying the relevance of socio-technical dynamics when addressing innovative technological disruption, the network of influences, responsibilities, and possibilities in flux as well as the interconnectedness of agents and the necessity of a shared sense of collective ethical action to achieve desired outcomes.

At times, conclusions might boil down to rather simple insights. E.g., when making «trustworthiness» the theme of discussion on artificial intelligence-based algorithms, a first realization may be that algorithms cannot be proper recipients of a moral account of trust because vulnerable trustors cannot attribute (well meaning) motivations to an artificial artefact, the trustee [5]. Algorithms may be reliable, but actual trust is to be lend to the socio-technical assemblage, in which conceiving, developing, marketing, auditing, competing, and regulating agents interact with, experience, challenge and support each other. Institutions are formed and categories (of algorithms) emerge that support trust by evoking expressions that portray algorithms—even though potentially opaque—as essentially based on good (or malicious) intentions, as trustworthy or not. Perhaps rarely will multiple agents consciously flog towards a common idea of advancing a specific product actively through the means of responsible innovation. But if this happens, an assemblage can be identified. Different from an indifferent, almost Kafkaesque, pursuit of an engineering career, acquiring a world view in terms of assemblages will hopefully spark the students' interest to contribute to the responsible innovation assemblage.

3.2 Course Structure

An actual engineering ethics course based on assemblage theory may obviously take many forms. We outline one possible concept of a structure that, admittedly, still needs to prove effective. Based on the above rationale, however, we believe it best to split the course into two parts: A first part is driven by case studies and works towards establishing the main concepts of assemblage theory one by one. A summary can then outline the theory itself and retrospectively link concepts and terminology to the case studies that were discussed in the first part. Its purpose lies in highlighting conceptual similarities and patterns as a concession to student expectations of take-away messages, which needs to be balanced with the open-ended nature of assemblage theory. A second part could task students to penetrate a new case study along the conceptual lines taught in the first part. More ideally than ready-made case studies as hand-outs even, a current situation from the news or reports could provide an analytical challenge a group of students could work on.

4 CONCLUSIONS

In this contribution, we have advocated the adoption of assemblage theory as a conceptual approach for ethical analyses in an engineering ethics educational context. We have outlined the limits of teaching ethical theory to increase commitment to responsible innovation practice in engineering students as a core motivation for approaching engineering ethics education differently. Instead, we believe that assemblage theory's focus on conditions, relations and events is well suited to shed light on the ethical challenges that innovative technologies, the perpetuated use of algorithms and artificial intelligence first and foremost, bring about. However, even though assemblage theory is a viable engine for ethics education, its conceptual intricacies may be best introduced in a step-by-step fashion without overbearing terminology.

Hermann Diebel-Fischer's part in this work was supported by the German Federal Ministry of Education and Research (BMBF, 01IS18026A-D) by funding the competence center for Big Data and AI ScaDS.AI Dresden/Leipzig.

REFERENCES

[1] Johnson, D. G. (2017), Can Engineering Ethics Be Taught? *The Bridge – Linking Engineering and Society*, Vol. 47, No. 1, pp. 59–64.

[2] DeLanda, M. (2016), Assemblage Theory. Edinburgh University Press.

[3] Buchanan, I. (2021), Assemblage Theory and Method. Bloomsbury Academic.

[4] Ananny, M. (2016), Toward an Ethics of Algorithms: Convening, Observation, Probability, and Timeliness. *Science Technology and Human Values* Vol. 41, No. 1, pp. 93–117.

[5] Rieder, G., Simon, J., & Wong, P-H. (2020), Mapping the Stony Road toward Trustworthy AI: Expectations, Problems, Conundrums. *SSRN Electronic Journal*, 2020, pp. 1–14.

[6] Fischer, J. (2014), What Kind of Ethics? – How Understanding the Field Affects the Role of Empirical Research on Morality for Ethics, In *Christen, M., van Schaik, C., Fischer, J., Huppenbauer, M., Tanner, C. (eds) Empirically Informed Ethics: Morality between Facts and Norms. Library of Ethics and Applied Philosophy*, Springer, pp. 29-43.

[7] Nail, T. (2017), What is an Assemblage? *SubStance*, Vol. 46, No. 1, pp. 21–37.