De Ethica

A Journal of Philosophical, Theological and Applied Ethics

Vol. 6, No. 1 (2020)

D E E T H I C A

A JOURNAL OF PHILOSOPHICAL, THEOLOGICAL AND APPLIED ETHICS

De Ethica seeks to publish scholarly works in philosophical, theological and applied ethics. It is a fully peer-reviewed, open-access publication hosted by Linköping University Electronic Press. We are committed to making papers of high academic quality accessible to a wide audience.

De Ethica is published in cooperation with *Societas Ethica, the European Society for Research in Ethics*. *Societas Ethica* was founded in Basel, Switzerland in 1964; today it has more than 200 members from more than 20 countries, representing a variety of theological and philosophical traditions. The annual conferences of *Societas Ethica* draw speakers from across the globe and provide a lively forum for intellectual exchange. Like *Societas Ethica*, *De Ethica* aims to create dialogue across national, political, and religious boundaries.

We welcome contributions from all philosophical and theological traditions. While we do welcome historically and empirically oriented work, our focus is on normative ethical questions. We also have a special interest in papers that contribute to ongoing public debates.

It is our aim to facilitate intellectual exchange across disciplinary and geographical boundaries and across the gaps between different philosophical and theological traditions. Thus we seek to publish papers that advance a clear and concise argument, avoid jargon, and are accessible to a non-specialized academic audience.

De Ethica is an Open Access publication with no publication fee. This means that the contents of the journal are published free of charge for authors, and can be accessed free of charge by readers. *De Ethica* uses Creative Commons BY-NC licencing, meaning that anyone may make full use of an article for non-commercial purposes as long as a full reference (including DOI) is given to the original publication. Additionally authors are free to archive the copy-edited version of the article at his or her web page, or at their university's web page.

By publishing exclusively online, *De Ethica* is not only more accessible than most traditional journals, but its production is also considerably more environmentally friendly. *De Ethica* is a not-for-profit operation; hence we welcome donations that might help us to improve the journal's quality, but we do not rely on them.

De Ethica was founded in 2013. It published its first issue in 2014, under the guidance of its first Editor-in-Chief, distinguished professor Brenda Almond.

Cover photo courtesy of
Huzeyfe Turan

From the Editors

For many ethicists, the spring of 2020 has been marked by an adaptation of academic life to extensive restrictions aimed at reducing the effects of the Covid-19 pandemic. Some researchers within the field of ethics are more directly involved in addressing social challenges related to the crisis, while others continue their research and deliver academic courses in as normal a manner as possible. Societas Ethica – the European Society for Research in Ethics – has cancelled its annual meeting in autumn 2020, and is simultaneously planning for a digital conference related to the ongoing pandemic. De Ethica – one of the Society's most important initiatives – will soon announce a special issue on the pandemic.

Meanwhile, we are delighted to present this issue, which is not related to the pandemic, but instead is highly representative of the main strategy of the journal and Societas Ethica: to promote European research in ethics through dialogue between philosophical, theological, and applied ethics.

The first article in the issue offers a discussion of the concept of moral agency in the context of AI. It is titled 'Moral Agency without Responsibility? Analysis of Three Ethical Models of Human-computer Interaction in Times of Artificial Intelligence.' This analysis is a contribution to applied ethics and the growing field of AI ethics in particular. At the same time, it is a philosophical contribution to the fundamental theoretical issue of moral agency and responsibility. Alexis Fritz, Wiebke Brandt, Henner Gimpel, and Sarah Bayer scrutinize three philosophical models that all describe forms of computer systems in terms of moral agency. The authors are skeptical about these attempts, arguing instead in favor of a concept of moral agency that considers human responsibility to be crucial.

In his article 'Violence, Shame, and Moral Agency – An Exploration of Krista K. Thomason's Position,' Jan-Olav Henriksen elaborates on the issue of moral agency from a different perspective. He discusses Krista K. Thomason's effort to explain violence as a response to the loss of agency. The starting point of Thomason's approach is the observation that people can respond to shame with violence. Violence thus becomes a way of regaining agency. Henriksen scrutinizes Thomason's understanding and suggests an alternative. According to his account, violent reactions that appear during the experience of shame need not be described as rational if we view shame as a manifestation of the lack of 'ability to fulfill the intended project or achieve the desired aim in a coherent manner.'

In an article titled 'Distributive Energy Justice and the Common Good,' Anders Melin discusses the issue of distributive justice in relation to energy production and energy consumption. The author compares Martha Nussbaum's capabilities approach with a contemporary model of the tradition of common good within Catholic theology. While

both approaches offer arguments in favor of global redistribution of energy production and consumption, the theological approach is more radical and therefore a more reasonable response to global injustices as they appear in the context of energy production and distribution.

In her article ''What Do We Do with the Art of Monstrous Men?' Betrayal and the Feminist Ethics of Aesthetic Involvement,' Sarah Stewart-Kroeker reflects on the consumption of artwork. Her feminist approach is related to the context of the #MeToo movement. The author focuses on the fact that aesthetic evaluation of an artist's work might be highly personal, thus creating special dilemmas when the artist is accused of sexual abuse. Stewart-Kroeker argues that a proper response to these dilemmas requires reflexive and social-structural examination.

'On Some Moral Implications of Linguistic Narrativism Theory' is an attempt to explicate normative components within a metatheory called linguistic narrativism. Natan Elgabsi and Bennett Gilbert follow Iris Murdoch, who claimed that abstract theoretical descriptions might imply evaluative components to the degree of normativity of moral visions. The authors argue that linguistic narrativism contains what they view as "undesirable moral agnosticism," and they believe that such a metatheory should be normatively evaluated.

I hope that the reader finds the articles of this issue to be of interest for reflection as well as critique as the main instrument of philosophical enquiry. I would also like to encourage those working in different traditions of ethics to submit their articles to De Ethica.

Elena Namli, Editor in Chief

Moral agency without responsibility? Analysis of three ethical models of human-computer interaction in times of artificial intelligence (AI)

Alexis Fritz, Wiebke Brandt, Henner Gimpel and Sarah Bayer

Philosophical and sociological approaches in technology have increasingly shifted toward describing AI (artificial intelligence) systems as '(moral) agents,' while also attributing 'agency' to them. It is only in this way – so their principal argument goes – that the effects of technological components in a complex human-computer interaction can be understood sufficiently in phenomenological-descriptive and ethical-normative respects. By contrast, this article aims to demonstrate that an explanatory model only achieves a descriptively and normatively satisfactory result if the concepts of '(moral) agent' and '(moral) agency' are exclusively related to human agents. Initially, the division between symbolic and sub-symbolic AI, the black box character of (deep) machine learning, and the complex relationship network in the provision and application of machine learning are outlined. Next, the ontological and action-theoretical basic assumptions of an 'agency' attribution regarding both the current teleology-naturalism debate and the explanatory model of actor network theory are examined. On this basis, the technical-philosophical approaches of Luciano Floridi, Deborah G. Johnson, and Peter-Paul Verbeek will all be critically discussed. Despite their different approaches, they tend to fully integrate computational behavior into their concept of '(moral) agency.' By contrast, this essay recommends distinguishing conceptually between the different entities, causalities, and relationships in a human-computer interaction, arguing that this is the only way to do justice to both human responsibility and the moral significance and causality of computational behavior.

Introduction: Exemplary harmful outcomes

Artifacts have played a substantial role in human activity since the first Paleolithic hand axes came into use. However, the emergence of an (ethical) discussion about which roles can be attributed to the people and artifacts involved in an action is only a consequence of the increasing penetration of artifacts carrying 'artificial intelligence' (AI) into our everyday lives.

Let us consider three examples of the potentially harmful effect of sophisticated machine learning approaches:

1) Google's search engine shows ads for high-paying executive jobs to men, but not so much to women.[1] Google's photo tagging service incorrectly labeled photos showing African-American people as showing 'gorillas.'[2] Even years after being alerted to this racist behavior, Google did not fix the machine learning approach itself, instead simply removing the word 'gorilla' from the set of possible labels.[3]
2) Amazon developed a machine learning system designed to analyze the résumés of job applicants and rate them with respect to their technical skills. The system was shown to be sexist in how it distinguished between applicants: 'It penalized résumés that included the word 'women's,' as in 'women's chess club captain.' And it downgraded graduates of two all-women's colleges.'[4] Amazon eventually shut down the system after failing to fully prevent discrimination.
3) In pretrial, parole, and sentencing decisions in the U.S., machine learning algorithms frequently predict a criminal defendant's likelihood of committing a future crime. The calculation of these so-called 'recidivism scores' is made by commercial providers that do not disclose the workings of their models. It was demonstrated for a widely used criminal risk assessment tool that used 137 features concerning an individual that the model performs no better than a simple logistic regression using just two features: age and the defendant's total number of previous convictions.[5] Yet, the seemingly more sophisticated 137-feature black box is being used in practice and has been accused of having a racial bias.[6,7]

We do not suggest that Google, Amazon, or the providers of criminal risk assessment tools are sexist, racist, or discriminatory by purpose in any other way. These examples merely illustrate that even well-intentioned initiatives using subsymbolic AI black boxes can lead to harmful outcomes. These systems may do very well with respect to some performance measures but may have inductive biases which are hard to detect and hard to fix. Overall, applications of AI, and especially subsymbolic machine learning-based

[1] Cf. Julia Carpenter, 'Google's Algorithm Shows Prestigious Job Ads to Men, But Not to Women. Here's Why That Should Worry You', *The Washington Post* (July 6, 2015), online at https://www.washingtonpost.com/news/the-intersect/wp/2015/07/06/googles-algorithm-shows-prestigious-job-ads-to-men-but-not-to-women-heres-why-that-should-worry-you/ (accessed 2019-11-10).
[2] Cf. Alex Hern, 'Google's Solution to Accidental Algorithmic Racism: Ban Gorillas', *The Guardian* (January 12, 2018), online at https://www.theguardian.com/technology/2018/jan/12/google-racism-ban-gorilla-black-people (accessed 2019-11-10).
[3] Cf. ibid.
[4] Reuters, 'Amazon Ditched AI Recruiting Tool that Favored Men for Technical Jobs', *The Guardian* (October 11, 2018), online at https://www.theguardian.com/technology/2018/oct/10/amazon-hiring-ai-gender-bias-recruiting-engine (accessed 2019-11-10).
[5] Cf. Julia Dressel and Hany Farid, 'The Accuracy, Fairness, and Limits of Predicting Recidivism', *Science Advances* 4:1 (2018).
[6] Cf. Anthony W. Flores, Kristin Bechtel and Christopher T. Lowenkamp, 'False Positives, False Negatives, and False Analyses: A Rejoinder to 'Machine Bias: There's Software Used Across the Country to Predict Future Criminals. And It's Biased Against Blacks'', *Federal Probation Journal* 80:2 (2016), pp. 38-46.
[7] Cf. Sam Corbett-Davies et al., 'A Computer Program Used for Bail and Sentencing Decisions was Labeled Biased Against Blacks. It's Actually Not That Clear', *The Washington Post* (October 17, 2016), online at www.washingtonpost.com/news/monkey-cage/wp/2016/10/17/can-an-algorithm-be-racist-our-analysis-is-more-cautious-than-propublicas (accessed 2019-11-10).

systems, are part of complex socio-technical systems. There is no doubt that AI systems have moral impact, but do they act and reason morally?[8]

The question of whether it is possible to create ethically acting machines represents an ongoing discussion.[9,10] Additionally, the dominant approaches of technical philosophy and sociology currently emphasize the moral significance of AI systems, and have moved towards calling them '(moral) agents' and attributing them 'agency.' The principal argument of this approach is that it allows us to describe both the moral effect of an action's technological components and the complex network of human-computer interaction in a sufficiently descriptive and ethical manner. It is therefore crucial to elucidate the semantics of 'agency' and 'moral agency,' as well as their connection to the concept of responsibility, in order to provide more clarity in settings involving hybrid human-computer intelligence. The central issue is whether we can better grasp the descriptive and normative dimensions of AI and especially subsymbolic machine-learning-based systems with the help of the 'agency' attribution.

In the first part of this research, we provide basic information on symbolic and subsymbolic AI, the black box character of (deep) machine learning, and the complex relationship networks in the supply and application of machine learning.

The second part elaborates ontological and action-theoretical basic assumptions of agency attribution regarding the current teleology-naturalism debate, as well as an explanatory model of Actor-Network Theory (ANT).

Thirdly, three technical philosophical models describing computer systems as '(moral) agents' are critically analyzed with regard to whether an extended agency attribution really illuminates the descriptive and ethical-normative structure of human-computer interaction, or whether it obscures this.

Background on artificial intelligence

AI describes a computer 'system's ability to correctly interpret external data, to learn from such data, and to use those learnings to achieve specific goals and tasks through flexible adaptation.'[11] Different levels of AI include narrow AI (below human-level intelligence, outperforming humans in specific domains but not being potent in other domains), general AI (human-level intelligence across many domains), and artificial super intelligence (above human-level intelligence). Contemporary AI systems show narrow AI (also known as weak AI).

Early computer programs solved tasks that can logically be described with a set of rules and are therefore easy for computers but require prolonged effort for people. A branch of AI still follows this route: computers are equipped with a formal representation of knowledge about the world and the rules of logical reasoning. Thus, they deductively generate new insights. This type of AI is *symbolic AI* because it builds on explicit symbolic

[8] Cf. ibid.
[9] Cf. Michael Anderson and Susan Leigh Anderson, 'Machine Ethics. Creating an Ethical Intelligent Agent', *AI Magazine* 28:4 (2007), pp. 15-26.
[10] Cf. Gordana Dodig Crnkovic and Baran Çürüklü, 'Robots: Ethical by Design', *Ethics and Information Technology* 14:1 (2012), pp. 61-71.
[11] Andreas Kaplan and Michael Haenlein, 'Siri, Siri, in My Hand: Who's the Fairest in the Land? On the Interpretations, Illustrations, and Implications of Artificial Intelligence', *Business Horizons* 62:1 (2019), pp. 15-25.

programming and inference algorithms. IBM's chess computer Deep Blue defeating the chess world champion Gary Kasparov in 1997 is an example of a symbolic (narrow) AI system. The other type of AI is *subsymbolic AI* using machine learning. The challenge for today's computer programs is to solve tasks that for humans are hard to describe formally, as they are more intuitive; for example, speech recognition, face recognition, or emotions.[12] Machine learning aims to build computers that automatically improve through experience.[13] A computer program learns from experience with respect to a class of tasks and a specific performance measure, if its performance on tasks of that class improves with experience.[14] However, this focus on experience might lead to an inductive bias if training data is not representative of the data and situations a machine learning model will face after training. Within AI, 'machine learning has emerged as the method of choice for developing practical software for computer vision, speech recognition, natural language processing, robot control, and other applications.'[15] Contemporary voice assistants, such as Amazon's Alexa, Apple's Siri, and Microsoft's Cortana, leverage such subsymbolic (narrow) AI.

Symbolic AI is easier to debug, easier to explain, and easier to control than subsymbolic AI, as symbolic programming lends itself to human inspection. Subsymbolic AI requires less upfront knowledge, builds on learning from data more successfully and shows better performance than symbolic AI in many domains, especially on perceptual tasks.

Deep learning is a form of machine learning that has gained popularity in recent years due to advances in (big) data availability, (cloud-based) massive computing power, algorithms, and openly available libraries for using these algorithms. In this context, the 'depth' refers to the number of layers in the network's structure; for example, in an artificial neural network (ANN). In the training phase, the strength of the connections (an analogy to brain synapses) between different nodes (an analogy to brain neurons) in the network is identified and learned. The more nodes and connections a network has, the better the network can acquire structural descriptions of the domain (if sufficient training data is available). Some of the largest artificial neural networks have millions of nodes and billions of connections.

Black box character of (deep) machine learning

Machine learning models, especially deep ANN, are frequently perceived as a black box.[16] Once such a model is then trained, and calculating the output based on a given input is rather simple. In principle, all the weights and functions to apply can be inspected manually. However, the sheer number of nodes and connections in a deep ANN, as well as the non-linearity of the calculations, make it practically very difficult, if not impossible, to fully understand the model's behavior for all but the most trivial examples. It is even more difficult to ex-ante predict the outcome of the statistical learning process. Thus, many people effectively perceive deep learning as a black box.

[12] Cf. ibid., p. 15.
[13] Cf. Stuart J. Russell and Peter Norvig, *Artificial Intelligence. A Modern Approach* (Boston: Pearson, 2016).
[14] Cf. Tom M. Mitchell, *Machine Learning* (Boston, Mass.: WBC/McGraw-Hill, 1997).
[15] Russell and Norvig, *Artificial Intelligence*, p. 255.
[16] Cf. Davide Castelvecchi, 'Can we open the black box of AI?', *Nature* 538:7623 (2016), pp. 20-23.

Over recent years, applications of AI became more sophisticated in terms of high-impact and high-risk tasks, such as autonomous driving or medical diagnosis. This has led to an increasing need for explanations.[17] At the same time, this rising complexity has made it more difficult to get insights and to understand and trust the system's functions – not just for users, but also for the programmers of those algorithms.[18] A logical model, like a decision-tree with statements involving 'and,' 'if-then,' etc., is comprehensible for the user. The larger the decision tree, the longer it takes, but humans are able to work through this process. Understanding deep learning models with millions or even billions of connections can be compared to understanding human predictions: we might anticipate what the system predicts, based on prior experience with the system, but we will never be completely sure if our assumption about the system's operating principles is correct.

This lack of transparency stands at the core of the discussion about the accountability and responsibility of humans regarding AI systems: can the user trust a prediction or be responsible for a decision made by a system that she or he cannot understand? To solve this issue, the research stream of *explainable AI* discusses two main options: white box and black box approaches. White box approaches aim at transparency, for instance, by displaying verbally or graphically the 'information contained in the knowledge base,' or via explaining the evidence, such as displaying the symptoms and test results that indicate the existence of a disease.[19] As the operating principles of linear models or decision trees are easier to understand, those models still dominate in many application areas.[20] Nevertheless, complex machine-learning models are in the fast lane and should offer explanations of their predictions to users. Due to the rising complexity of such systems, we cannot expect users to understand how the models work.[21]

Taking the example of an ANN, black box approaches focus on, for example, visualizing the input-output relationship, thus showing which input is most responsible for reaching a certain output.[22,23] These approaches help users and programmers shed light on the black box, but they do not reveal the whole complex functions of the ANN. Therefore, such approaches make AI 'more of a grey than a black box.'[24] Still, these highly performant black and grey box machine learning systems pose challenges in terms of agency, especially as these artifacts are part of complex systems involving multiple actors.

[17] Cf. Jichen Zhu et al., 'Explainable AI for Designers: A Human-Centered Perspective on Mixed-Initiative Co-Creation', *IEEE Conference on Computational Intelligence and Games* (2018), pp. 1-8.
[18] Cf. Mitchell, *Machine Learning*.
[19] Cf. Carmen Lacave and Francisco J. Díez, 'A Review of Explanation Methods for Bayesian Networks', *The Knowledge Engineering Review* 17:2 (2002), pp. 107-127.
[20] Cf. Grégoire Montavon, Wojciech Samek and Klaus-Robert Müller, 'Methods for Interpreting and Understanding Deep Neural Networks', *Digital Signal Processing* 73 (2018), pp. 1-15.
[21] Cf. Or Biran and Kathleen McKeown, 'Human-Centric Justification of Machine Learning Predictions', *Proceedings of International Joint Conferences on Artificial Intelligence* (2017), pp. 1461-1467.
[22] Cf. Zhu et al., 'Explainable AI for Designers'.
[23] Cf. Ruth C. Fong and Andrea Vedaldi, 'Interpretable Explanations of Black Boxes by Meaningful Perturbation', *Proceedings of the IEEE International Conference on Computer Vision* (2017), pp. 3429-3437.
[24] Zhu et al., 'Explainable AI for Designers'.

Complex relationship networks in the supply and application of machine learning

Figure 1 is a stylized picture of the value chain from algorithm development, all the way through to the human being affected by a decision. It is an abstract depiction of the processes behind the examples given above. By showing the different types of human actors involved, it can thereby illustrate the complex interplay between different human actors and artifacts.

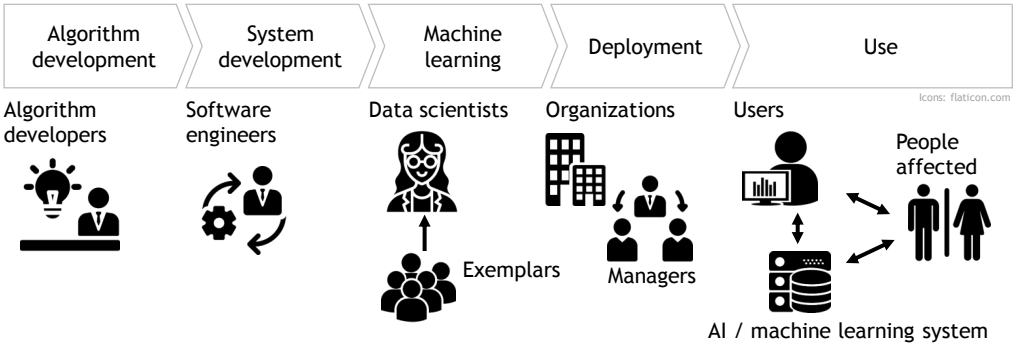

Figure 1: Stylized value chain from algorithm development to use of machine learning systems

Algorithm development conceives general-purpose machine learning algorithms. System development embeds these algorithms in a software system, typically for a specific purpose like criminal risk assessment or personnel decisions. The system is trained on the basis of data that originates from it (e.g., prior decisions by humans like evaluating résumés or sentencing criminals). Organizations like a court system or a company – or, more specifically, managers within an organization – then decide to use the system. Finally, individual users (like a clerk in the personnel department or a judge) interact with the machine learning-based system to obtain information and make decisions that affect others, like applicants or defendants.

If this overall socio-technical system harms people, who is responsible? There are eight candidates: (1) the technical AI system, despite it being an artifact; (2) the users obliged to use a system they do not understand; (3) the managers who neither understand the black box nor make individual decisions; (4) the organization; (5) the data scientists, despite the fact they do not make decisions concerning individual persons; (6) the people providing the training data, oftentimes unknowingly; (7) the software engineers, despite their inability to foresee the system's behavior after learning; and (8) the algorithm developers who created the multi-purpose black boxes in the first place. Is any single candidate responsible, several of them (each to a certain degree), is the overall socio-technical system responsible without individual responsibility, or are none of them responsible?

Pre-assumptions of agency attribution based on action theory

Asking what an actor or an action is and how it can be explained leads to a branched discussion of very different approaches to action theory. This makes it clear that agency attribution depends on several ontological and action-theoretical basic assumptions. Whoever uses concepts of action must not shy away from reflecting on these fundamental implications. Only against this background can different positions and their possible conclusions be adequately understood and discussed.

The teleology-naturalism debate concerns whether we can adequately describe and understand human actions and natural events by the same language and at the same level. Actor-Network Theory seeks to overcome the distinction between humans and non-humans by describing an actor as the symmetrical interplay between social, technical, and natural entities.

The teleology-naturalism debate in action theory

In order to determine the ways in which an action differs from a natural event, it is instructive to take a closer look at how we talk about it. We usually explain actions through the intentions of the person doing them ('She opened the window to air the room'), thus attributing the mental capacity to have goals, make decisions, etc. In contrast, we consider a natural event as the (provisional) end of a causal chain, and name the previous chain links as an explanation for its taking place ('The window opened because a gust of wind blew against it').[25] Obviously, we distinguish between a 'mental' language, which refers to actions, and a 'physical'[26] language, which refers to natural events.[27] As long as both are applied only in their respective fields, there is no problem. However, it is questionable whether the same event can be expressed in both languages: is the window opening perhaps also due to certain neuronal states that triggered the woman's arm movement? Is such a physical description perhaps even more accurate than referring to mental states and abilities?

How do these different descriptions of the same event relate to each other? Are both of them legitimate perspectives that are able to coexist, or do they exclude each other so that at least one of them must be wrong? As a third option, one language might be translatable into the other.[28]

This is exactly the basic assumption of the naturalistic approach: anything expressed in mental language can be translated into physical language without any loss of meaning. Ultimately, there is no ontological difference between actions and natural events.[29] Accordingly, actions are subject to the same causal laws as natural events. Therefore, they can, in theory, be retrospectively deduced from a certain set of necessary and sufficient conditions, as well as predicted for the future if those very conditions are fulfilled (deductive-nomological explanatory scheme) – even if an accurate prediction is practically difficult to realize due to the complex interplay of numerous internal and external conditional factors.[30] In order to avoid this problem, a simpler action pattern is declared the object of investigation: the so-called 'basic action,' which consists of only a

[25] Cf. Edmund Runggaldier, *Was sind Handlungen? Eine philosophische Auseinandersetzung mit dem Naturalismus*. Stuttgart: W. Kohlhammer, 1996, p. 17, 106; Christoph Horn and Guido Löhrer, 'Einleitung: Die Wiederentdeckung teleologischer Handlungserklärungen', in *Gründe und Zwecke. Texte zur aktuellen Handlungstheorie*, edited by Christoph Horn and Guido Löhrer (Berlin: Suhrkamp, 2010), pp. 7-45, at p. 8.

[26] Cf. Runggaldier, *Was sind Handlungen?*, p. 18.

[27] Cf. Runggaldier, *Was sind Handlungen?*, p. 106.

[28] Cf. Scott R. Sehon, 'Abweichende Kausalketten und die Irreduzibilität telologischer Erklärungen', in *Gründe und Zwecke. Texte zur aktuellen Handlungstheorie*, edited by Christoph Horn and Guido Löhrer (Berlin: Suhrkamp, 2010), pp. 85-111, at p. 87; Horn, 'Einleitung', pp. 15f.

[29] Cf. Runggaldier, *Was sind Handlungen?*, pp. 15, 24-26.

[30] Cf. ibid., pp. 26, 106f, 110; Josef Quitterer, 'Basishandlungen und die Naturalisierung von Handlungserklärungen', in *Soziologische Handlungstheorie. Einheit oder Vielfalt*, edited by Andreas Balog and Manfred Gabriel (Opladen: Westdeutscher Verlag, 1998), pp. 105-122, at pp. 106f.

simple body movement (e.g. bending a finger).[31] If one regards the different levels of an action as an 'action tree,' then this 'basic action' represents the lowest, most basal level, which cannot be further explained by other partial actions. You get to higher levels by asking 'why?': he bent his finger to pull the trigger of a weapon, to fire a bullet at a person, to kill that person, etc. By contrast, you reach a lower level by asking 'how?': he killed him by shooting at him, by using the trigger, by bending the finger, etc. At this point, where you cannot break down the question of 'how?' any further, you have reached the lowest level.[32] Regardless of whether you consider these levels to describe the same action or many different actions,[33] both positions agree that the 'basic action' is the main, essential action on which further analysis has to concentrate.

The teleological approach contrasts with the naturalistic approach, and its followers criticize the orientation towards 'basic actions': in order to do justice to the nature of an action, it cannot be reduced to a body movement. On the contrary, the higher levels of the action tree are to be examined, where the actor's intentions, systems of rules and signs, the situational context with possibly involved third parties, etc. are situated.[34] Certain actions (e.g. greeting, betting, lecturing) are not dependent on a certain movement of the body, and therefore cannot be reduced to it.[35] But even actions whose correlation to body movements is evident, such as firing a weapon, are principally comprehensible only against the background of their circumstances and references: not the bending of the finger, but the intention to kill, the connection with the victim, etc., which constitute the action.[36] The reference to lower levels of action can be misleading, and even be used to deliberately conceal the essence of the action: 'I have only...'[37]

Teleologists agree that intentions are the criterion that distinguishes an action from a natural event.[38] In contrast to the naturalistic translation thesis, they insist that mental language cannot be reduced to physical language, since intentions cannot be equated with the links of a causal chain.[39]

Not only is it practically impossible to completely determine all the causal conditions for an action taking place, but this is also theoretically opposed by the conviction that a human being is fundamentally free in his decision to act.[40]

[31] Cf. Quitterer, 'Basishandlungen', pp. 107f.
[32] Cf. Runggaldier, *Was sind Handlungen?*, pp. 46-48; Quitterer, 'Basishandlungen', pp. 115f; Georg Kamp, 'Basishandlungen', in *Handbuch Handlungstheorie. Grundlagen, Kontexte, Perspektiven*, edited by Michael Kühler and Markus Rüther (Stuttgart: J. B. Metzler Verlag, 2016), pp. 69-77, at pp. 69f.
[33] According to the 'unifiers'/'minimizers' bending the finger and killing the victim represent a single action; from the point of view of the 'multipliers'/'maximizers' these are numerically different actions (cf. Runggaldier, *Was sind Handlungen?*, pp. 50f; Quitterer, 'Basishandlungen', pp. 116f; Christian Budnik, 'Handlungsindividuation', in *Handbuch Handlungstheorie. Grundlagen, Kontexte, Perspektiven*, edited by Michael Kühler and Markus Rüther (Stuttgart: J. B. Metzler Verlag, 2016), pp. 60-68, at p. 60).
[34] Cf. Runggaldier, *Was sind Handlungen?*, pp. 55, 59, 62; Quitterer, 'Basishandlungen', p. 106.
[35] Cf. Runggaldier, *Was sind Handlungen?*, pp. 65f.
[36] Cf. ibid., p. 62; Quitterer, 'Basishandlungen', pp. 118f.
[37] Cf. Runggaldier, *Was sind Handlungen?*, p. 62f.
[38] Cf. Friedo Ricken, *Allgemeine Ethik* (Stuttgart: W. Kohlhammer, 2013 [1983])., pp. 103f; Horn, 'Einleitung', p. 9; Sehon, 'Abweichende Kausalketten', p. 85; Runggaldier, *Was sind Handlungen?*, pp. 12, 68; Donald Davidson, 'Handlungen, Gründe und Ursachen', in *Gründe und Zwecke. Texte zur aktuellen Handlungstheorie*, edited by Christoph Horn and Guido Löhrer (Berlin: Suhrkamp, 2010), pp. 46-69, at p. 48.
[39] Cf. Runggaldier, *Was sind Handlungen?*, p. 76; Horn, 'Einleitung', p. 8; Sehon, 'Abweichende Kausalketten', p. 110.
[40] Cf. Runggaldier, *Was sind Handlungen?,* pp. 110-113.

Donald Davidson, a representative of a moderate naturalism, takes this objection seriously and does not claim any principal predictability of human action. In the case of a broken windowpane, it can be stated afterwards, without any doubt, that a certain stone caused its breaking. However, to move from such a causal analysis to a prognosis about how hard one has to throw a stone against a window to break it in the future is something completely different.[41] For actions, it applies analogously that individual, concrete actions can be explained causally and, in these individual cases, be translated into physical language. However, there are no laws either in the mental realm or between the mental and the physical sphere according to which predictions about future actions can be made. The name of this position, 'anomalous monism,' derives from the negation of such overarching laws.

Teleologists reply that such a concept devalues the mental side, since it is causally effective only insofar as it can be translated into physical terms.[42] Again, the intentionality of the actor is reduced.

Instead of searching for mental or physical events within the actor that have produced his action, one should simply accept the actor himself as the origin of his action ('agent-causality').[43]

The concept of 'agency' in Actor-Network Theory (ANT)

Both naturalistic and teleological theories of action require a distinct separation between the subject and the object of an action. ANT criticizes this basic assumption. It opposes mechanistic, quasi-automatic explanations of actions, as well as models of understanding that presuppose the intention, autonomy, or consciousness of the human actor. But how are the terms 'action' and 'agency' to be understood if there is no subject-object difference, no primary principle, or no modern concept of the subject?

ANT is a challenging alternative to traditional theories of action, and has become one of the classic approaches of technical sociology.[44] Bruno Latour, Michel Callon, and John Law founded this theory in the 1980s and continue to develop it further to this day. Despite the diversity and complexity of the concepts within this family of ANTs, some key aspects shall be briefly highlighted.[45]

ANT does not ask why an actor acts in this way and not differently. Rather, it describes how an actor is transformed into an agent through the interplay of social, technical, and natural entities. The surprising thing is not so much that action always refers to others, but that non-humans are not simply passive objects of human action. Instead, they act themselves in a heterogeneous network.[46]

[41] Cf. Davidson, 'Handlungen, Gründe und Ursachen', pp. 63f.
[42] Cf. Runggaldier, *Was sind Handlungen?,* pp. 122-127, 132; Quitterer, 'Basishandlungen', pp. 109; 112-114.
[43] Cf. Runggaldier, *Was sind Handlungen?*, pp. 144-147.
[44] Cf. Roger Häußling, *Techniksoziologie. Eine Einführung* (Opladen, Toronto: Verlag Barbara Budrich, 2019), pp. 240-252.
[45] A differentiated introduction to ANT in German is offered by Andréa Belliger and David J. Krieger, 'Einführung in die Akteur-Netzwerk-Theorie', in *ANThology. Ein einführendes Handbuch zur Akteur-Netzwerk-Theorie*, edited by Andréa Belliger and David J. Krieger (Bielefeld: transcript, 2006), pp. 13-50; Ingo Schulz-Schaeffer, *Sozialtheorie der Technik* (Frankfurt am Main: Campus-Verlag, 2000).
[46] Cf. Bruno Latour, 'Social Theory and the Study of Computerized Work Sites', in *Information Technology and Changes in Organizational Work*, edited by W. J. Orlinokowsky and Geoff Walsham (London: Chapman and Hall, 1996), pp. 295-307, at pp. 303ff.

This basic assumption is formulated by ANT as the general principle of symmetry, which claims a radically equal treatment of humans and non-humans. Social, technical, and natural factors are equal and depend on each other.[47] In order to clarify the concept that not only humans are capable of acting, ANT replaces the 'actor' with an 'actant.' An actant is generally someone or something with the ability to act and to exercise activity.[48] Both human and non-human actants begin to create heterogeneous networks by themselves. They do not precede their networking but are produced by the networking process. The results of such networking are hybrids (i.e. hybrid forms of the social, the technical, and the natural).[49]

Actants transform into actors when a role and interests are assigned to them in the process of building networks (figuration).[50] The successive and different steps of the network-building process are summarized under the term 'translation.' This is 'the continuous attempt to integrate actors into a network by 'translating' them into roles and interests.'[51] Translations create the 'identities, characteristics, competences, qualifications, behaviors, institutions, organizations and structures necessary to build a network of relatively stable, irreversible processes and procedures.'[52] A 'network' is not an external social reality, but a theoretical term for a concept that 'is traced by those translations in the scholars' accounts.'[53] Statements about actants and actors are always moments in the process of network building or translation.

Latour exemplified his ANT by closing a door.[54] He understands this process as a network in which both human (= the user) and technical (= the door) actants are involved. If you regularly forget to close the door, this can quickly become a problem. This problem can then be solved, for instance, by introducing a sign, hiring a porter, or implementing a door-closing mechanism. If, for instance, a door-closing mechanism is installed, the new technical actant changes the characteristics and behavior of the existing network. For example, people have to adapt to the speed of the closing door.

While humans determine technical behavior, technical artifacts can also lead to human behavioral changes. In ANT, there is no clearly assignable making and being made; instead, there is only the network of actants (e.g. texts, people, animals, architectures, machines, or money).[55]

[47] Cf. Bruno Latour, *Wir sind nie modern gewesen. Versuch einer symmetrischen Anthropologie* (Berlin: Akad.-Verl., 1995), pp. 125ff.
[48] Cf. Madeleine Akrich and Bruno Latour, 'A Summary of a Convenient Vocabulary for the Semiotics of Human and Nonhuman Assemblies', in *Shaping Technology/ Building Society. Studies in Sociotechnical Change*, edited by Wiebe E. Bijker and John Law (Cambridge, Mass.: The MIT Press, 1992), pp. 259-264, at p. 259.
[49] Cf. Latour, *Wir sind nie modern gewesen*, pp. 7f.
[50] Cf. Michel Callon, 'Einige Elemente einer Soziologie der Übersetzung: Die Domestikation der Kammmuscheln und der Fischer der S. Brieuc-Bucht', in *ANThology. Ein einführendes Handbuch zur Akteur-Netzwerk-Theorie*, edited by Andréa Belliger and David J. Krieger (Bielefeld: transcript, 2006), pp. 135-174, at pp. 146f; Bruno Latour, *Reassembling the Social. An Introduction to Actor-Network-Theory* (Oxford: Oxford University Press, 2007) p. 53.
[51] Belliger and Krieger, 'Einführung in die Akteur-Netzwerk-Theorie', p. 39 (translated by authors).
[52] ibid., p. 39 (translated by authors).
[53] Latour, *Reassembling the Social*, p.108.
[54] Cf. Jim Johnson, 'Mixing Humans and Nonhumans Together: The Sociology of a Door-Closer', *Social Problems* 35:3 (1988), pp. 298-310.
[55] Cf. Michel Callon, 'Techno-ökonomische Netzwerke und Irreversibilität', in *ANThology. Ein einführendes Handbuch zur Akteur-Netzwerk-Theorie*, edited by Andréa Belliger and David J. Krieger (Bielefeld: transcript, 2006), pp. 309-342, at p. 313.

This sometimes results in controversial, even irritating formulations in Latour's writing. Thus, a clumsy hotel key chain acts more morally than its human user. Due to its size, it forces the guest to hand in the key at the reception desk before leaving the hotel.[56] When asked whether a person or a weapon was responsible for killing a person, Latour replied: 'It is neither people nor guns that kill. Responsibility for action must be shared among the various actants.'[57] It is a hybrid that cannot be reduced to a technical or human actant. Agency emerges from a connection of actants in the network: 'Action is a property of associated entities.'[58] Action and agency are always distributed among different entities. According to the sociologist M. Wieser, the notion of the agency in terms of non-human things must 'not be understood as animism or as the naive intentionality of things, but as the power of things, highlighting their resistance.'[59] 'Agency' is not a substance, but a process.[60] In this sense, non-humans also possess the ability to act, for which the English term 'Agency' or 'Material Agency' has prevailed in technical sociology.[61]

Three technical-philosophical approaches

It turned out that 'agent' or 'agency' are multifaceted concepts in the field of action theory. Their semantics and language practice depend on controversial and sometimes contradictory basic assumptions. The following technical-philosophical approaches are not identical with any of the action-theoretical directions discussed above. Nevertheless, the basic concerns, the course, or the focus of the following technical-philosophical approaches can each be traced back to one of the previously discussed theories of action.

The following approaches aim to describe and ethically evaluate the complex human-computer interaction appropriately and descriptively with the help of the terms '(moral) agent' or 'agency.'

The original problem and the basic concern of the three systemic models coincide. Nevertheless, Floridi's, Johnson's and Verbeek's answers compete with each other, and thus cannot be sensibly combined. To put it simply, we can describe Floridi's model as 'techno-centric,' Johnson's as 'anthropocentric,' and Verbeek's as 'constructivist.'

L. Floridi: Artificial Agency

According to Floridi, the so-called standard ethics (i.e. deontological – like discourse-theoretical and contractualistic – or teleological – like virtue-ethical or consequentialist

[56] Cf. Bruno Latour, 'Technology is Society Made Durable', in *A Sociology of Monsters? Essays on Power, Technology and Domination*, edited by John Law (London/ New York: Routledge, 1991), pp. 103-131.
[57] Bruno Latour, *Pandora's Hope. Essays on the Reality of Science Studies* (Cambridge, Mass.: Harvard Univ. Press, 1999), p. 180.
[58] Ibid., p. 182.
[59] Matthias Wieser, *Das Netzwerk von Bruno Latour. Die Akteur-Netzwerk-Theorie zwischen Science & Technology Studies und poststrukturalistischer Soziologie* (Bielefeld: transcript, 2012), p. 182 (translated by authors).
[60] Cf. ibid., pp. 184f.
[61] Cf. Latour, *Reassembling the Social*, p. 45; Werner Rammert, *Technik – Handeln – Wissen. Zu einer pragmatistischen Technik- und Sozialtheorie* (Wiesbaden: Springer VS, 2016 [2007]), p. 14; Wieser, *Das Netzwerk von Bruno Latour*, pp. 175-184.

ethics) are hopelessly overwhelmed by the challenges of human-computer interaction.[62] The first reason for this is that in conventional philosophy, only human beings (and thus no AI), are considered 'moral agents.' Thus, the human actor is burdened by a disproportionally great responsibility.[63] Secondly, actions are judged on the basis of the actor's intentions:[64] it is morally relevant whether a person is injured intentionally or unintentionally. However, this focus on intentions does not help us where AI is used. In fact, the impact of a self-learning computer system can never be overlooked completely and therefore cannot be answered for by the designer or user. It is for this reason that Floridi suggests that we broaden the concept of 'moral agency' and refrain from judging intentions.[65]

Starting from the question who or what a 'moral agent' is, Floridi argues that definitions must be looked at in their particular context:[66] A car mechanic looks at a car from a different point of view than an ethicist. To refer to these different points of view, Floridi uses the technical term 'level of abstraction.' At different levels of abstraction, different observables are relevant. For example, an ethicist delights in low pollutant emission, while a car mechanic is pleased by an unbroken V-belt.[67]

In order to define 'agent' properly, Floridi suggests a higher level of abstraction than is usually adopted. Candidates for 'agents' should no longer be examined for intentionality or other mental abilities; instead, they should be observed from a more distant perspective, appearing only vaguely as 'systems.' To be called 'agents,' systems have to be interactive, autonomous, and adaptive.[68]

According to Floridi, whether, for example, a computer program checking CVs is considered an 'agent' depends on the granularity of the level of abstraction employed: if only the incoming CVs and their outgoing evaluation are regarded as 'observables,' but the algorithm itself is hidden, the recruitment program appears interactive, autonomous, and adaptive, consequently, as an 'agent': 'interactive,' because it begins to work in reaction to an external input; 'autonomous,' because it arranges the many applications automatically – as in a black box –; and 'adaptive,' because it learns on the basis of the data records.[69]

From 'agent' to 'moral agent' takes only a small step: for Floridi, all 'agents' whose actions have morally qualifiable consequences are 'moral agents.'[70] Consequently, the recruitment program is not only an 'agent,' but also a 'moral agent,' because its selection is sexually discriminatory.

[62] Cf. Luciano Floridi and Jeff W. Sanders, 'Artificial Evil and the Foundation of Computer Ethics', *Ethics and Information Technology* 3 (2001), pp. 55-66, at pp. 57, 64f.
[63] Cf. Luciano Floridi and Jeff W. Sanders, 'On the Morality of Artificial Agents', *Minds and Machines* 14 (2004), pp. 349-379, at pp. 350f.
[64] Cf. Luciano Floridi, 'Faultless Responsibility: On the Nature and Allocation of Moral Responsibility for Distributed Moral Actions', *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 374 (2016), Issue 2083, at p. 4.
[65] Cf. ibid., p. 3f.
[66] Cf. Floridi and Sanders, 'On the Morality of Artificial Agents', pp. 352f.
[67] Cf. Luciano Floridi, 'Levels of Abstraction and the Turing Test', *Kybernetes* 39 (2010), pp. 423-440, at p. 426; Floridi and Sanders, 'On the Morality of Artificial Agents', p. 354.
[68] Cf. Floridi and Sanders, 'On the Morality of Artificial Agents', pp. 357f.; Floridi, 'Levels of Abstraction', p. 432.
[69] Cf. Floridi and Sanders, 'On the Morality of Artificial Agents', p. 362; Floridi, 'Levels of Abstraction', p. 432.
[70] Cf. Floridi and Sanders, 'On the Morality of Artificial Agents', p. 364.

However, the program is not morally responsible for its consequences, as responsibility requires intention,[71] but intention does not matter at the level of abstraction chosen for 'agency.' According to Floridi, 'moral agents' without intentions are not morally responsible for their actions but accountable.[72] If artificial 'moral agents' cause damage – by analogy with sanctions on people – they can be modified, disconnected from the data network, or completely deleted or destroyed.[73]

Floridi finally concludes that his understanding of 'moral agency' and 'accountability' sufficiently clarifies the ethical questions of human-computer interaction: 'The great advantage is a better grasp of the moral discourse in non-human contexts.'[74]

This positive self-evaluation of Floridi has to be questioned:

First, the AI debate is – according to Floridi – about attributing responsibility. If we stick to this assumption, we cannot see how the existence of non-responsible 'moral agents' can help in the search for a culprit.

Second, Floridi's reference to non-human 'moral' sources of good and evil of all kinds is nothing new in itself: a serious illness, a large avalanche, a chainsaw, a rabid dog, or falling roof tiles can all cause human suffering. However, despite the damage, we would never speak of a 'moral' avalanche, chainsaw, disease, dog, or tile.

By calling computer systems 'moral,' we can neither describe their mode of action better (causality), nor come closer to resolving moral issues (evaluation of an action or attribution of responsibility).

It can perhaps be said that the novelty of Floridi's approach lies not so much in qualifying the impacts of computer systems as 'moral' but in perceiving them as 'agents' at a certain level of abstraction. However, would that take us any further descriptively or normatively? This raises three thoughts: first, the necessity of making computer systems 'accountable' (i.e. that they have to be reprogrammed or even switched off if deficient) may be realized without there being any need of calling them 'moral agents.' While we may call our computer names when it does not do what we want it to, we do not do so because we seriously believe it will somehow impress our computer. Second, not all links in a causal chain need to be called 'moral agents' in order to become the object of ethical thought. Even in the standard ethics scolded by Floridi, a moral evaluation of an action or the attribution of responsibility is only possible after a precise and sufficient description of the causal connections. Third, it must also be criticized that if something goes wrong, at the level of abstraction favored by Floridi, the question of responsibility can no longer be posed for AI as a 'moral agent,' since Floridi abstracts from human intention, and computer systems are accountable but not morally responsible. In this way, ethically questionable incentive structures emerge, where the responsible party can be excused prematurely.

Thus, the impression is reinforced that the term 'moral agents' in Floridi's explanatory model contributes nothing toward gaining a better descriptive and normative understanding of human-computer interaction. It can thus be dismissed without consequences, since 'moral agent' or 'moral agency' is an empty concept if separated from responsibility.

[71] Cf. ibid., p. 365.
[72] Cf. ibid., pp. 351, 376.
[73] Cf. ibid., pp. 372f.
[74] Ibid., p. 376.

D. G. Johnson: Triadic agency

Deborah Johnson struggles to find a happy medium between two extremes: one position undermines human responsibility to the extent that computer systems are referred to as 'moral agents,' and Johnson explicitly criticizes Floridi's approach. Representatives of the other position, on the other hand, misjudge the moral quality of machine behavior since they regard technology as extra-moral.

In the course of a larger searching movement, Johnson developed the so-called 'Triadic agency' model. According to Johnson, a state is caused neither by man nor by the computer system alone, but by a differentiated interaction. Basically, 'agency' means a 'capability to act.' Johnson distinguishes between three forms of agency:

(1) 'causal agency': things have a causal effect;[75]

(2) 'intentional agency': people act intentionally; their intention causes the action;[76]

(3) 'triadic agency': these forms of 'agency' relate to each other and are more than the sum of their individual parts. When people cooperate with computer systems, then:

a. the user wants to achieve a certain goal – in our case the Amazon HR department wants an efficient and effective personnel selection –and delegates this task to the designers;

b. the designer project team creates the recruitment program;

c. with the help of this program the initial goal is achieved.[77]

In the 'triadic agency' model, responsibility is attributed only to those who are able to act intentionally. Since AI has no intention, it bears no responsibility for its causal effectiveness. Only humans can be 'moral agents' due to their intentional capacity. People therefore remain responsible, even if they delegate increasingly complex tasks to AI. In the search for the responsible person(s), it has to be asked in the direction of the designer or user until a person (or a group of persons) is found. However, an answer to the question of how much responsibility each person bears cannot be found without also considering the technological component.

By differentiating between three modes of action, Johnson first succeeds in maintaining the ontological difference between man and machine in terms of action theory. This differentiation is not essentialist, since it does not refer to fixed descriptive characteristics, but to certain abilities. Secondly, although only human beings can be responsible, their responsibility can only be clarified if all components of action are considered. Because of the descriptive and normative significance of machine behavior, Johnson does not want to renounce the agency attribution.

However, Johnson's inclusive use of the term 'agency' gives rise to misunderstandings and side scenes, since one term refers to human beings, computer systems, and human-computer interaction. Johnson strives to name the difference and interrelationship between man and computer systems, but she shrinks from taking the final step and continues to call computer systems 'agents.' Unlike Floridi's use of the term, Johnson's 'agency' is not meaningless but misleading. It would have been more

[75] Cf. Markus Schlosser, 'Agency', *The Stanford Encyclopedia of Philosophy* (2015), online at https://plato.stanford.edu/entries/agency/ (accessed 2019-11-15).

[76] Cf. ibid.

[77] Cf. Deborah G. Johnson and Mario Verdicchio, 'AI, Agency and Responsibility: The VW Fraud Case and Beyond', *AI & SOCIETY* (2018), online at https://doi.org/10.1007/s00146-017-0781-9 (accessed 2019-11-15), p. 4.

beneficial to use different terms such as 'factor,' 'cause,' or 'actor' in order to emphasize the specific descriptive and normative contribution of computer systems.

P.-P. Verbeek: Hybrid agency

Peter-Paul Verbeek's 'mediation theory' is based on Don Ihde's postphenomenological approach and Bruno Latour's 'actor-network theory.'[78] Verbeek emphasizes the joint causality of man and technology. Hence, technology actively mediates between human beings and their environment.[79] It does so on two levels: hermeneutically, by influencing human perception of the world, and pragmatically, in partaking in human action.[80]

Returning to our example of a recruitment program, the question of how the human resources department perceives the applicants – as deficient or positive – is decisively mediated by technology (hermeneutical mediation), and the final recruitment decision is pragmatically mediated. It is neither determined by, nor can it be made completely independently of, technology.

Consequently, according to Verbeek, moral decisions and actions are joint products of human beings and technology;[81] morality is 'hybrid,' and 'moral agency' is a mixture ('composite moral agency').[82] No thing or living being possesses 'moral agency' by itself. Rather, 'moral agency' results from complex technical-human interaction; it does not form the basis for an action but emerges from it.[83]

Verbeek goes so far as to describe even the actors themselves as the result of interaction.[84] Nevertheless, Verbeek's theorem of a hybrid 'moral agency' does not mean that people cannot bear responsibility. In particular, designers of computer systems bear great responsibility because technology shapes the way of being in the world, and thus the human being himself. Verbeek shows the ethical dimensions with sentences such as 'Designers materialize morality'[85] and 'Designing technology is designing human beings.'[86]

Against this background, we would like to ask whether Verbeek's 'moral agency' attribution helps us to understand human-computer interaction better both descriptively and ethical-normatively. The strength of Verbeek's postphenomenological-constructivist mediation theory undoubtedly lies in the fact that it acknowledges the complexity of human-computer interaction. Verbeek's approach is particularly successful in reflecting

[78] Cf. Peter-Paul Verbeek. 'Materializing Morality. Design Ethics and Technological Mediation', *Science, Technology, & Human Values* 31 (2006), pp. 361-380, at pp. 362f.; Peter-Paul Verbeek, *Moralizing Technology. Understanding and Designing the Morality of Things* (Chicago: Univ. of Chicago Press, 2011) pp. 33, 45-47, 52.
[79] Cf. Verbeek, 'Materializing Morality', p. 364; Peter-Paul Verbeek, 'Some Misunderstandings About the Moral Significance of Technology', in *The Moral Status of Technical Artefacts*, edited by Peter Kroes and Peter-Paul Verbeek (Dordrecht: Springer, 2014), pp. 75-88, at pp. 77f.
[80] Cf. Verbeek, 'Materializing Morality', pp. 364, 368.
[81] Cf. Verbeek, 'Some Misunderstandings', p. 78.
[82] Ibid., pp. 77f.
[83] Cf. ibid., pp. 75, 80; Peter-Paul Verbeek, 'Designing the Morality of Things: The Ethics of Behaviour-Guiding Technology', in *Designing in Ethics*, edited by Jeroen van den Hoven, Seumas Miller and Thomas Pogge (New York: Cambridge Univ. Press, 2017), pp. 78-94, at p. 84.
[84] Cf. Peter-Paul Verbeek, 'Beyond Interaction: A Short Introduction to Mediation Theory', *Interactions* 22 (2015), pp. 26-31, at p. 28.
[85] Verbeek, 'Beyond Interaction', p. 31 (cf. Verbeek, 'Materializing Morality', pp. 361, 369, 379; Verbeek, 'Designing the Morality of Things', p. 88).
[86] Verbeek, 'Beyond Interaction', p. 28.

reality. If we accept that technology creates reality in terms of its interplay with human beings, and if this awareness replaces both obsession with, as well as forgetfulness about, technology, then much is gained for the debate about the responsible use of technology in both a descriptive and normative sense. This is true even if mediation is not a specific characteristic of technology alone.

However, with regard to Verbeek's understanding of 'moral agency,' there are important inquiries to make:

Unlike Floridi, Verbeek considers intentionality and freedom as part of the term 'moral agency,' albeit in a mediated, hybrid form. However, intentionality and freedom do not constitute 'moral agency'. Instead, and much like 'moral agency' itself, this only results from a complex human-computer interaction.

The strength of the postphenomenological-constructivist view of reality turns into a weakness as soon as we want to attribute agency or responsibility to individual, concrete entities. In Verbeek's mediation theory, 'moral agency,' intention, freedom, and thus responsibility can no longer be attributed to individuals, since they always emerge from an overall structure. Ultimately, in Verbeek's theory of mediation, the individual and his actions cannot be conceived without technical influences or mediation. Human beings and computer systems are 'actants' – only as a mixture are they also 'agents.'

Verbeek's two concerns – reconstructing the understanding of human-computer interaction and attributing moral responsibility – could also be fulfilled if the human actors remained 'moral agents.' For the realization that human capacity to act is always mediated is nothing new from a philosophical point of view. However, in order to avoid a circular conclusion in the attribution of 'moral agency' and moral responsibility, the freedom of human actors must be regarded as taking precedence. This is because interaction does not have its origin in itself but is a consequence of the human ability to reflect, decide, and act freely.

Conclusion

This study has revealed the opportunities and risks of applying the concept of 'moral agency' to human-computer interaction. Ultimately, the risks of agency attribution to computational behavior are disproportionate to the benefits of such language practice.

From a descriptive and ethical-normative point of view, this practice proves to be both unnecessary and risky. Floridi's use of 'moral agents' for computer systems is redundant. Exclusive features for human or social contexts (e.g. 'intentionality' or 'responsibility'), which should be preserved, come out of sight.

Verbeek offers a comprehensive and promising understanding of human-computer interaction. However, his 'moral agent' attribution is circular or leads to an infinite regression, thus making it objectionable. This is illustrated by the fact that it is difficult to identify a specific human capacity or actor for responsibility.

Johnson's results are consistent in view of their ontological and action-theoretical premises. She also conceptually differentiates the contribution of each component and is thus able to provide an almost accurate understanding of human-computer interaction. However, the 'agency' attribution gives rise to misunderstandings. At the same time, there is a serious risk that the extensive use of 'moral agents' undermines the question of responsibility.

Consequently, an appropriate differentiation between humans and computers should also be conceptually discernible. In this way, human-computer interaction can not only be described more precisely but the ethical-normative structure can also be elaborated more clearly.

Alexis Fritz, Wiebke Brandt, Henner Gimpel and Sarah Bayer
alexis.fritz@ku.de, wiebke.brandt@ku.de,
henner.gimpel@fit.fraunhofer.de, sarah.bayer@fim-rc.de

Bibliography

Akrich, Madeleine and Bruno Latour. 'A Summary of a Convenient Vocabulary for the Semiotics of Human and Nonhuman Assemblies', in *Shaping Technology/ Building Society. Studies in Sociotechnical Change*, edited by Wiebe E. Bijker and John Law. Cambridge, Mass.: The MIT Press, 1992, pp. 259-264.

Anderson, Michael and Susan Leigh Anderson. 'Machine Ethics. Creating an Ethical Intelligent Agent', *AI Magazine* 28:4 (2007), pp. 15-26.

Belliger, Andreá and David J. Krieger. 'Einführung in die Akteur-Netzwerk-Theorie', in *ANThology. Ein einführendes Handbuch zur Akteur-Netzwerk-Theorie*, edited by Andréa Belliger and David J. Krieger. Bielefeld: transcript, 2006, pp. 13-50.

Biran, Or and Kathleen McKeown. 'Human-Centric Justification of Machine Learning Predictions', *Proceedings of International Joint Conferences on Artificial Intelligence* (2017), pp. 1461-1467.

Budnik, Christian. 'Handlungsindividuation', in *Handbuch Handlungstheorie. Grundlagen, Kontexte, Perspektiven*, edited by Michael Kühler and Markus Rüther. Stuttgart: J. B. Metzler Verlag, 2016, pp. 60-68.

Callon, Michel. 'Einige Elemente einer Soziologie der Übersetzung: Die Domestikation der Kammmuscheln und der Fischer der S. Brieuc-Bucht', in *ANThology. Ein einführendes Handbuch zur Akteur-Netzwerk-Theorie*, edited by Andréa Belliger and David J. Krieger. Bielefeld: transcript, 2006, pp. 135-174.

[original English version:

Callon, Michel. 'Some Elements of a Sociology of Translation: Domestication of the Scallops and the Fishermen of St. Brieuc Bay', in *Power, Action and Belief: A New Sociology of Knowledge?*, edited by John Law. London: Routledge, 1986, pp. 196-233.]

Callon, Michel. 'Techno-ökonomische Netzwerke und Irreversibilität', in *ANThology. Ein einführendes Handbuch zur Akteur-Netzwerk-Theorie*, edited by Andréa Belliger and David J. Krieger. Bielefeld: transcript, 2006, pp. 309-342.

[original English version:

Callon, Michel. 'Techno-Economic Networks and Irreversibility', in *A Sociology of Monsters? Essays on Power, Technology and Domination,* edited by John Law. London/ New York: Routledge, 1991, pp. 132-161.]

Carpenter, Julia. 'Google's Algorithm Shows Prestigious Job Ads to Men, But Not to Women. Here's Why That Should Worry You', *The Washington Post* (July 6, 2015), online at https://www.washingtonpost.com/news/the-intersect/wp/2015/07/06/googles-algorithm-shows-prestigious-job-ads-to-men-but-not-to-women-heres-why-that-should-worry-you/ (accessed 2019-11-10).

Castelvecchi, Davide. 'Can we open the black box of AI?', *Nature* 538:7623 (2016), pp. 20-23.

Corbett-Davies, Sam, Emma Pierson, Avi Feller and Sharad Goel. 'A Computer Program Used for Bail and Sentencing Decisions was Labeled Biased Against Blacks. It's Actually Not That Clear', *The Washington Post* (October 17, 2016), online at www.washingtonpost.com/news/monkey-cage/wp/2016/10/17/can-an-algorithm-be-racist-our-analysis-is-more-cautious-than-propublicas (accessed 2019-11-10).

Crnkovic, Gordana Dodig and Baran Çürüklü. 'Robots: Ethical by Design', *Ethics and Information Technology* 14:1 (2012), pp. 61-71.

Davidson, Donald. 'Handlungen, Gründe und Ursachen', in *Gründe und Zwecke. Texte zur aktuellen Handlungstheorie*, edited by Christoph Horn and Guido Löhrer. Berlin: Suhrkamp, 2010, pp. 46-69.

Dressel, Julia and Hany Farid. 'The Accuracy, Fairness, and Limits of Predicting Recidivism', *Science Advances* 4:1 (2018).

Flores, Anthony W., Kristin Bechtel and Christopher T. Lowenkamp. 'False Positives, False Negatives, and False Analyses: A Rejoinder to 'Machine Bias: There's Software Used Across the Country to Predict Future Criminals. And It's Biased Against Blacks'', *Federal Probation Journal* 80:2 (2016), pp. 38-46.

Floridi, Luciano and Jeff W. Sanders. 'Artificial Evil and the Foundation of Computer Ethics', *Ethics and Information Technology* 3 (2001), pp. 55-66.

Floridi, Luciano and Jeff W. Sanders. 'On the Morality of Artificial Agents', *Minds and Machines* 14 (2004), pp. 349-379.

Floridi, Luciano. 'Faultless Responsibility: On the Nature and Allocation of Moral Responsibility for Distributed Moral Actions', *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 374 (2016), Issue 2083.

Floridi, Luciano. 'Levels of Abstraction and the Turing Test', *Kybernetes* 39 (2010), pp. 423-440.

Fong, Ruth C. and Andrea Vedaldi. 'Interpretable Explanations of Black Boxes by Meaningful Perturbation', *Proceedings of the IEEE International Conference on Computer Vision* (2017), pp. 3429-3437.

Häußling, Roger. *Techniksoziologie. Eine Einführung*. Opladen, Toronto: Verlag Barbara Budrich, 2019.

Hern, Alex. 'Google's Solution to Accidental Algorithmic Racism: Ban Gorillas', *The Guardian* (January 12, 2018), online at https://www.theguardian.com/technology/2018/jan/12/google-racism-ban-gorilla-black-people (accessed 2019-11-10).

Horn, Christoph and Guido Löhrer. 'Einleitung: Die Wiederentdeckung teleologischer Handlungserklärungen', in *Gründe und Zwecke. Texte zur aktuellen Handlungstheorie*, edited by Christoph Horn and Guido Löhrer. Berlin: Suhrkamp, 2010, pp. 7-45.

Johnson, Deborah G. and Mario Verdicchio. 'AI, Agency and Responsibility: The VW Fraud Case and Beyond', *AI & SOCIETY* (2018), online at https://doi.org/10.1007/s00146-017-0781-9 (accessed 2019-11-15).

Johnson, Jim. 'Mixing Humans and Nonhumans Together: The Sociology of a Door-Closer', *Social Problems* 35:3 (1988), pp. 298-310.

Kamp, Georg. 'Basishandlungen', in *Handbuch Handlungstheorie. Grundlagen, Kontexte, Perspektiven*, edited by Michael Kühler and Markus Rüther. Stuttgart: J. B. Metzler Verlag, 2016, pp. 69-77.

Kaplan, Andreas and Michael Haenlein. 'Siri, Siri, in My Hand: Who's the Fairest in the Land? On the Interpretations, Illustrations, and Implications of Artificial Intelligence', *Business Horizons* 62:1 (2019), pp. 15–25.

Lacave, Carmen and Francisco J. Díez. 'A Review of Explanation Methods for Bayesian Networks', *The Knowledge Engineering Review* 17:2 (2002), pp. 107-127.

Latour, Bruno. 'Social Theory and the Study of Computerized Work Sites', in *Information Technology and Changes in Organizational Work*, edited by W. J. Orlinokowsky and Geoff Walsham. London: Chapman and Hall, 1996, pp. 295-307.

Latour, Bruno. 'Technology is Society Made Durable', in *A Sociology of Monsters? Essays on Power, Technology and Domination*, edited by John Law. London/ New York: Routledge, 1991, pp. 103-131.

Latour, Bruno. *Pandora's Hope. Essays on the Reality of Science Studies*. Cambridge, Mass.: Harvard Univ. Press, 1999.

Latour, Bruno. *Reassembling the Social. An Introduction to Actor-Network-Theory*. Oxford: Oxford University Press, 2007.

Latour, Bruno. *Wir sind nie modern gewesen. Versuch einer symmetrischen Anthropologie*. Berlin: Akad.-Verl., 1995.

[original English version:

Latour, Bruno. *We Have Never Been Modern.* Cambridge, Mass.: Harvard Univ. Press, 1993.]

Mitchell, Tom M. *Machine Learning*. Boston, Mass.: WBC/McGraw-Hill, 1997.

Montavon, Grégoire, Wojciech Samek and Klaus-Robert Müller. 'Methods for Interpreting and Understanding Deep Neural Networks', *Digital Signal Processing* 73 (2018), pp. 1-15.

Quitterer, Josef. 'Basishandlungen und die Naturalisierung von Handlungserklärungen', in *Soziologische Handlungstheorie. Einheit oder Vielfalt*, edited by Andreas Balog and Manfred Gabriel. Opladen: Westdeutscher Verlag, 1998, pp. 105-122.

Rammert, Werner. *Technik – Handeln – Wissen. Zu einer pragmatistischen Technik- und Sozialtheorie*. Wiesbaden: Springer VS, 2016 [2007].

Reuters. 'Amazon Ditched AI Recruiting Tool that Favored Men for Technical Jobs', *The Guardian* (October 11, 2018), online at https://www.theguardian.com/technology/2018/oct/10/amazon-hiring-ai-gender-bias-recruiting-engine (accessed 2019-11-10).

Ricken, Friedo. *Allgemeine Ethik*. Stuttgart: W. Kohlhammer, 2013 [1983].

Runggaldier, Edmund. *Was sind Handlungen? Eine philosophische Auseinandersetzung mit dem Naturalismus*. Stuttgart: W. Kohlhammer, 1996.

Russell, Stuart J. and Peter Norvig. *Artificial Intelligence. A Modern Approach*. Boston: Pearson, 2016.

Schlosser, Markus. 'Agency', *The Stanford Encyclopedia of Philosophy* (2015), online at https://plato.stanford.edu/entries/agency/ (accessed 2019-11-15).

Schulz-Schaeffer, Ingo. *Sozialtheorie der Technik*. Frankfurt am Main: Campus-Verlag, 2000.

Sehon, Scott R. 'Abweichende Kausalketten und die Irreduzibilität telologischer Erklärungen', in *Gründe und Zwecke. Texte zur aktuellen Handlungstheorie*, edited by Christoph Horn and Guido Löhrer. Berlin: Suhrkamp, 2010, pp. 85-111.

Verbeek, Peter-Paul. 'Beyond Interaction: A Short Introduction to Mediation Theory', *Interactions* 22 (2015), pp. 26-31.

Verbeek, Peter-Paul. 'Designing the Morality of Things: The Ethics of Behaviour-Guiding Technology', in *Designing in Ethics*, edited by Jeroen van den Hoven, Seumas Miller and Thomas Pogge. New York: Cambridge Univ. Press, 2017, pp. 78-94.

Verbeek, Peter-Paul. 'Materializing Morality. Design Ethics and Technological Mediation', *Science, Technology, & Human Values* 31 (2006), pp. 361-380.

Verbeek, Peter-Paul. 'Some Misunderstandings About the Moral Significance of Technology', in *The Moral Status of Technical Artefacts*, edited by Peter Kroes and Peter-Paul Verbeek. Dordrecht: Springer, 2014, pp. 75-88.

Verbeek, Peter-Paul. *Moralizing Technology. Understanding and Designing the Morality of Things*. Chicago: Univ. of Chicago Press, 2011.

Wieser, Matthias. *Das Netzwerk von Bruno Latour. Die Akteur-Netzwerk-Theorie zwischen Science & Technology Studies und poststrukturalistischer Soziologie*. Bielefeld: transcript, 2012.

Zhu, Jichen, Antonios Liapis, Sebastian Risi, Rafael Bidarra and G. Michael Youngblood. 'Explainable AI for Designers: A Human-Centered Perspective on Mixed-Initiative Co-Creation', *IEEE Conference on Computational Intelligence and Games* (2018), pp. 1-8.

Violence, Shame, and Moral Agency – An Exploration of Krista K. Thomason's Position

Jan-Olav Henriksen

Krista Thomason's account of shame explains the link between shame and violence as something that arises out of a tension between our identity and our self-conception: those things about which we feel shame are part of our identities, but they are not part of our self-conception. She sees violence as an attempt to regain agency and control and overcome shame. Although this is an important trait in shame, to explain violence as a response to the loss of agency is not sufficient. Furthermore, it cannot explain serious self-harm as the result of shame, since such reactions undermined the agency she holds that violence attempts to reclaim. Hence, these features need to be incorporated into a wider account of shame that sees it as a response to the interruption of intentional projects and attempts for coherent agency.

Violence is among the more problematic moral topics humans face. However, the phenomenon of violence also contributes to making problematic some of the prevalent definitions of shame, such as those advocated by, e.g., Deonna et al.,[1] Manion,[2] and Kekes,[3] all of whom, along different lines, argue for shame as a valuable device for human improvement, albeit not without important qualifications. These contributions all take their point of departure in an understanding of shame as a reaction to the failure to live up to standards, norms, or ideals that we have.[4] The problem with this approach to shame is that it does not seem to be able to take into account the empirical observation that agents may respond to shame with violence, aggression, or self-aggression. Krista Thomason, therefore, argues that this definition, especially in its moral form, is flawed.[5] There are many valuable elements in her analysis and her proposal for an alternative. However, her own proposal also raises some issues that require further discussion. The following discussion is nevertheless not an attempt to vindicate the moral definition but is an attempt

[1] Julien A. Deonna, Raffaele Rodogno, and Fabrice Teroni, *In Defense of Shame: The Faces of an Emotion* (New York: Oxford University Press, 2011).
[2] Jennifer Manion, 'The Moral Relevance of Shame', *American Philosophical Quarterly* 39:1 (2002) pp. 73–90.
[3] John Kekes, 'Shame and Moral Progress', *Midwest Studies In Philosophy* 13:1 (1988) pp. 282–296.
[4] John Deigh, 'Shame and Self-Esteem: A Critique', *Ethics* 93:2 (1983), p. 225 traces this moral definition back to Rawls, in John Rawls, *A Theory of Justice* (Cambridge, Mass.: Belknap Press, 1972), pp. 440-446. He also argues that similar definitions are pervasive in the literature. It should be noted, though, that Rawls sees shame as more than a moral phenomenon, as he distinguishes between natural and moral shame. Hence, Rawls can say that "Shame is sometimes a moral feeling" (ibid, 443).
[5] See Krista K. Thomason, 'Shame, Violence, and Morality', *Philosophy and Phenomenological Research* 91:1 (2015), pp. 1–24. Thomason's argues her position in a more expanded context in Krista K. Thomason, *Naked : The Dark Side of Shame and Moral Life* (New York: Oxford University Press, 2018).

to develop a broader and more nuanced understanding of the phenomenon of shame – which is an enterprise to which Thomason has offered valuable insights.

Thomason's Critique and Her Proposal for an Alternative Position

To act violently is not an obvious or immediately intelligible response to the painful feeling of failing to live up to an ideal. If such ideals or norms are of relevance to the one experiencing shame, then such a reaction seems to contradict or obliterate the previous presuppositions for agency. In other words: the moral definition of shame (which is mainly based on categories that must be considered moral, cf. standards, norms, ideals) cannot explain why such reactions take place. Thomason, therefore, argues that we need a definition of shame that is not based on moral concepts alone, but which nevertheless allows us to address the morally problematic feature of violence as a reaction to failure. Her main claim is that "shame arises out of a tension between our identity and our self-conception: those things about which we feel shame are part of our identities, but they are not part of our self-conception."[6] Let us consider her position more in detail.

According to Thomason, a philosophical account of moral emotions, among which shame is often included, faces two challenges. The first challenge has to do with explaining how emotion has moral value and what role it plays in the moral life. The other challenge is to provide a good conceptual analysis of shame that also can account how we actually experience it.[7] The second challenge is accordingly not only philosophical but also empirical: a definition is called for that must make sense of the empirical data we have on shame and the experiences of it. This point is where the moral definition of shame fails.

Thomason concludes this line of reasoning with a statement about the empirical falsification of the moral definition of shame:

> If shame is the painful feeling of not living up to one's values, it does not make sense that agents would respond to that feeling by doing something morally bad. What is more, doing something violent alleviates feelings of shame. If the traditional view is right, this experience is impossible: doing something morally wrong should make agents feel more shame rather than less.[8]

Moreover, Thomason refers to several examples in the literature where people respond to shame by doing something that is violent to themselves or others. Thus, they do something morally wrong in response to shame. Now, since the moral definition sees shame as a painful response to a failure to embody the values we care about, the definition is challenged. According to it, shame should cause us to act with restraint. Alternatively, as we can see in the cases that argue for shame as instigating self-improvement: the moral definition implies the expectation that one attempts to overcome shame by living up to one's ideals and values (or those of others) in the future.[9] However, in the cases Thomason describes, shame inspires the opposite of such self-improvement. The moral definition of shame cannot explain why people act in such a manner. Furthermore, and even more

[6] Thomason, 'Shame, Violence, and Morality', p. 1.
[7] Ibid. p. 2.
[8] Ibid. p. 2.
[9] Cf. Manion, 'The Moral Relevance of Shame', p. 84.

serious, it fails to address the fact that immoral acts can make those who experience shame feel better. In other words: immorality sometimes alleviates shame.[10]

There are different approaches to such instances of alleviating shame by acts of immorality that present us with possible strategies for explaining them without having to give up the moral definition. First, one can claim that shame is sometimes irrational, and accordingly, in exceptional cases, irrationality serves as an explanation. Second, one can also argue that shame, in some cases, is not properly focused. It is properly focused when "(1) we hold ourselves responsible for our failure and (2) when the norm to which we respond is a legitimate one."[11] Third, irrational shame, defined as shame that leads to acts that are incomprehensible, can also be explained by Gabriele Taylor's notion of "false shame," which occurs when we have standards or norms imposed upon us for a brief period, and which are opposed to genuine shame, the latter being the moral kind that occurs when we fail to live up to our ideals.[12] And finally, one can classify cases of reactive and immoral shame in shame-prone individuals with a maladaptive self-image, because "shame-prone individuals are more apt to respond with aggression than those who are not, but this is an issue with shame-proneness and not with shame."[13] Thomason nevertheless finds no reason for comprehending violent responses to feelings of shame as irrational.[14] That some shame-prone individuals respond to shame with aggression does not mean that anyone who responds to shame with aggression is shame-prone. No empirical data suggest that this may be the case.[15]

Accordingly, Thomason rejects attempts to explain the link between shame and violent responses within the frames of the moral definition. Accounts of shame based on this definition cannot explain why agents are tempted to respond to shame by doing something wrong.[16] She suggests an alternative account or explanation of the relationship between shame and violence that widens the scope and does not see shame as based on ideals and values. Her suggestion is especially fruitful since it therefore also points to the wider conditions for agency and to how shame may be a response to the intentions, orientations, and desires that lie behind the agency that is *interrupted* when it occurs. As we shall see in the following, this interruption is a central element in the alternative into which we try to integrate Thomason's analysis of shame.

According to Thomason, shame "arises when we feel some aspect of our identities defines us."[17] Thus, it is the globalization of one aspect of us that comes to dominate our inner realm of self-experience. In order to make this definition work, she has to make a distinction between identity and self-conception: "Those things about which we feel shame

[10] Cf. Thomason, 'Shame, Violence, and Morality', p. 6.
[11] Ibid. p. 7.
[12] Ibid.p. 7. Cf. Gabriele Taylor, 'Shame, Integrity, and Self-Respect', *Dignity, Character, and Self-Respect* edited by Robin Dillon (New York: Routledge, 1995).
[13] Ibid. p. 7.
[14] Cf. Ibid. p. 7-8. She also points to how these approaches can in fact contribute further to shame, and to people feeling shame about feeling shame: These explanatory approaches "encourage us to find fault with ashamed people. Because the traditional way of understanding shame is about failing to live up to values, we are forced to claim that agents who experience shame about their faces have misguided values and false beliefs. A different account of shame could explain them in a way that does not require attributing mistaken values to agents who feel this way." Ibid, p. 12.
[15] Ibid. p. 8.
[16] Ibid. p. 10.
[17] Ibid. p. 2.

are part of our identities, but they are not part of our self-conception."[18] In other words, she contributes further to nuancing the role of shame in the complex architecture of the self. She offers the following example as an illustration:

> … an agent feels shame when some aspect of her identity becomes prominent or revealed in the shameful moment and that she feels that this thing defines her as a whole. That is, in episodes of shame she feels defined by, reduced to, or totalized by some feature of herself. I take this defining feature of shame to be necessary rather than sufficient. In other words, someone may find herself in these circumstances and feel something other than shame. My contention is that when an agent reports feeling shame, this feature will be present in the experience. Similarly, if an agent does not yet feel shame, but fears it, it is because she fears that some aspect of herself will define her. [19]

Accordingly, shame is the result of experiences of feeling defined, reduced to, or totalized by some feature of ourselves.[20] This is the reason why she sees experiences of shame as something that involves a tension between our identity and our self-conception. A self-conception is the "self-image," i.e., "how we represent to ourselves the person we take ourselves to be." On the other hand, "our identities extend beyond what we represent to ourselves. An agent's identity is constituted by who she is in a broader sense and can include things that fall outside of her self-conception."[21] In her definition of shame, Thomason thereby sheds light on John Deigh's distinction between authorship and ownership,[22] because shame makes it impossible to disown the feature in question:

> Shame is the result of our *inability to disavow* that aspect of ourselves by which we feel defined explains why shame makes us feel so powerless. The thing that causes me shame both overshadows me and yet is me.[23]

The metaphor "overshadow" can be linked to the previously used notion of *interruption.* Thomason uses it in the characterization of shame as experiencing one's lack of agential control over the feature that causes shame: "Shame arises in response to those aspects of ourselves over which we have very limited control."[24] It can be features related to our bodies, our intelligence, our grace (or lack thereof), our families, and our socioeconomic status, all of which are things over which we have little influence:

[18] Ibid. Cf. Thomason, *Naked : The Dark Side of Shame and Moral Life*, pp. 101f., where she distinguishes between being defined and feeling defined by one aspect of one's identity.
[19] Thomason, 'Shame, Violence, and Morality', p. 11.
[20] Ibid. p. 12.
[21] Ibid. p. 12. On her use of the notion self.-conception, see especially Thomason, *Naked : The Dark Side of Shame and Moral Life*, pp. 92ff. Thomason seems to build on G. Taylor, when she speaks of the agent experiencing shame as "becoming aware of the discrepancy between her own assumption about her state or action and a possible detached observer-description of this state or action, and of her further being aware that she ought not to be in a position where she could be so seen, where such a description at least appears to fit." See Gabriele Taylor, *Pride, Shame and Guilt: Emotions of Self-Assessment* (Oxford: Clarendon Press, 1985), p. 66.
[22] Cf. Deigh.
[23] Thomason, 'Shame, Violence, and Morality', p.13. My italics.
[24] Ibid. p. 14

> The shame that we feel about these aspects of our identities does not stem from the fact that we falsely believe we are responsible for them and thus failing to live up to ideals. It stems from the fact that they compete with our self-conception in comprising who we are.[25]

Thomason's conception also provides a means for understanding how shame can be correlated to issues like race and gender, which, similarly, is beyond one's ability to control.[26] Such shame may not be due to the person's feeling of failure, but because of their sex or skin color, as they feel overshadowed by this aspect of themselves:

> Women and people of color are often thought of as a group rather than as individuals and others attribute thoughts, feelings, and behaviors to them in light of their sex or their skin color. Feeling as though one's identity can be "read off" of one's skin color or sex understandably makes one feel totalized by one's skin color or sex.[27]

Accordingly, Thomason presents us with a phenomenology of shame that can account for feelings of being made small: "The feeling of smallness is the feeling of our self-conception being dwarfed by the aspect of our identities that inspires our shame."[28] That which causes shame thus overshadows us, or, as we would say, interrupts us and our self-conception in the way it is articulating itself in agency.

Thomason argues that her model makes it unnecessary to divide shame into kinds: All shame is the result of the lack of coherence between self-conception and identity.[29] She can explain cases of moral shame without reference to a failure to live up to ideals. Shame emerges because someone, despite how she represented her moral character to herself, clearly was capable of doing something she thought she would never do. Shame interrupts or disturbs one's self-conception.[30]

How can this understanding of shame offer a better account of the relationship between shame and violence than the moral definition? Thomason argues that "we respond to shame with violence because it allows us to feel once again defined by our self-conception rather than those aspects of ourselves that fall outside of it." Thus, violent acts should be seen as a protest reaction – I am more than my face, my arms, my failure. It is, in her view, not the destructive element in the violent act that is its main aim, but the attempt to regain control: "Violence is the attempt to regain control, which shame itself has caused one to feel that is lost."[31] She elaborates:

> Our bodies, our sexuality, and our socioeconomic statuses are all rich targets for shame that are a part of our identity even though we do not choose them. Shame makes us feel that we are not in control of who we are: parts of my identity define me independently of how I want to define myself. One of the ways of alleviating shame is to do something that regains a sense of control. We try to hide, cover ourselves, or get away from the situation, and these actions can help us regain feelings of control

[25] Thomason, 'Shame, Violence, and Morality', p. 14.
[26] Cf. Ibid. pp. 14f.
[27] Ibid. p. 15.
[28] Ibid. p. 15.
[29] Cf. Ibid. p. 16. Cf. Thomason, *Naked : The Dark Side of Shame and Moral Life*, pp. 87f. On her use of the notion self.-conception, see especially pp. 92ff.
[30] Thomason, 'Shame, Violence, and Morality', p. 16.
[31] Cf. Ibid. p. 17.

> because we remove from sight the thing we experience as shameful. Violence, anger, and aggression can accomplish the same goal. At first, this looks puzzling because it seems that we might be equally reduced to or totalized by our acts of violence or aggression as much as our faces or bodies.[32]

Even though she sees the violence in question primarily as an act of self-assertion, it is warranted to ask why one cannot regain control and assert oneself in other and less destructive ways? An obvious example of this is how acts of violence may affect or even be directed against those who are weak or not able to protect themselves. Thomason seems to downplay the severe content of acts of shame-caused violence.[33] At least, one would think that violence was performed against the one who made one feel ashamed. However, that is not always the case.

Against this backdrop, shame that results in violence becomes possible to understand as rational. It restores a sense of agency:

> My sense of myself as an agent is closely connected to my self-conception. That is, one of the primary ways I think of myself is as an agent: one who chooses acts and makes decisions. Since my sense of my own agency is a large part of my self-conception, when I am seen as an agent, I feel as though my self-conception (not the parts of my identity that fall outside of it) is determining who I am.[34]

Thus, shame's violence is also a protest against becoming "reduced to some feature of our identity that we experience as fixed." The act of violence instigates the one who performs it as something else and more than what he is in his shame. That is the rationale for performing the act. The response from others – even a negative one, implies that the person to whom they are responding is more than the possessor of some shameful feature.[35] Moreover:

> becoming the object of resentment by doing something violent helps us to regain the feeling of control we lose in shame because we once again feel that our self-conception determines who we are. Others surely respond negatively to me as the violent agent, but they are no longer seeing me as an object of amusement or fascination. What we seek in shame is not approval, but recognition […]: Violence gains us that recognition because in asserting our agency, we assert our self-conception.[36]

Discussion and an alternative proposal

People might respond to shame with acts that are violent – against themselves or others. Examples of self-aggression are perhaps most obvious in cases where individuals commit suicide because of shame – as in cases where politicians have been outed for sexual abuse, or Japanese workers have been exposed for doing a bad job resulting in disgrace for their company. Examples of aggression against others I will return to below. There are several

[32] Ibid. p. 17.
[33] Cf. on Breivik below.
[34] Thomason, 'Shame, Violence, and Morality', p. 18.
[35] Ibid. p. 18.
[36] Ibid. p. 19.

elements to point to and discuss in Thomason's analysis of the causes of shame. First, we need to ask why violence stands forth as a reasonable way of assuring one's agency? Violence is not only destructive, but is often a mode of acting that is almost guaranteed to diminish the agent in the eyes of others, and thereby, it may cause even more shame. One could easily think of other ways of responding to shame that was not (so) destructive, be it other types of protest, laughing, or simply doing something that shows that you were not accepting being defined only by this or that trait – as in the running for public office if you are a woman or a person of color. Furthermore, since not everyone responds to shame by acting violently, the variation in responses may also suggest that this mode of response is not considered a good solution for everyone. Hence, one should distinguish between what makes acts of violence possible to understand as the least rational response for some candidates, whereas for others it may be the most reasonable choice, i.e., what appears to be founded on good reasons or warrants.

This leads us to a related, second, comment, namely that it is possible to see a moral interest in the struggle for recognition that the violent act implies, according to Thomason. However, a *negative* recognition like the one achieved by a violent act (which is mostly morally condemned by others) is, in the end, most likely going to end up in a new rejection and more shame, due to how the agent is then again defined by the violent act. The most obvious recent example of this is the Norwegian terrorist Anders Behring Breivik, whose acts can be understood as the result of narcissistic shame and rage.[37] After his deeds, however, hardly anyone can relate to him without thinking of the shameful acts he performed. It was nevertheless a long discussion about his sanity and to what extent he could be considered as a responsible agent. Thus, it does not seem obvious that violence is a rational response to shame – simply because it is likely to engender more shame, and not overcome it. However, it is still possible to see shame as causing different strategies that articulate struggles for recognition, among which violence can be one.[38]

Thirdly, in the description of self-asserting violence referred above, Thomason seems to emphasize the response of others to these acts as crucial for the experience of overcoming the shame-defining features in the agent. It may explain some instances of violence, but do they also explain self-inflicting violence like suicide? Is it not more plausible to see violence as one painful act that is done in order to numb an experience of another pain, without ascribing too much rationality to it?

Thomason's conception of shame does not lead her to argue for the elimination of shame altogether, still she sees shamelessness (which for her is the alternative) as something that: "reveals an agent's failure to recognize the limitations of her own self-conception".[39] In a more positive assessment, she sees shame as valuable because it can open us up to the perspectives of others: "it means that we do not take our own points of view as the only important ones."[40] It is a way of allowing us to overcome the limitations of our self-conceptions, by disturbing them: "The more authoritative I think my self-

[37] For a careful analysis of different aspects of Breivik's actions, see Arne Johan Vetlesen, *Studier I Ondskap* (Oslo: Universitetsforl., 2014). (Eng.: *Studies in Evil*). As Thomason herself points to, there is much empirical evidence for the connection between narcissism, shame and violence. However, "It is still not clear why shame-prone individuals would be more disposed to violence than others." Thomason, *Naked : The Dark Side of Shame and Moral Life*, p. 58.
[38] Cf. Axel Honneth, *The Struggle for Recognition : The Moral Grammar of Social Conflicts* (Cambridge: Polity Press, 1995), p. 177.
[39] Thomason, 'Shame, Violence, and Morality', p. 20.
[40] Ibid., p. 21.

conception is, the more prone I am to overlook things that do not fit with it."[41] She continues:

> A liability to shame prevents us from taking the way we see ourselves to be the primary authority in our self-estimation. Feelings of shame arise when we feel defined by some aspect of our identity that is not part of how we see ourselves. Even though that part of our identity is not part of our self-conception, we feel shame because we still acknowledge it as ours.[42]

We can phrase Thomason's intention here more negatively: We need shame when we are too high on ourselves. Our all-too-prevalent tendency to evaluate ourselves positively is the reason why we need shame to prevent self-inflation: "not because it is morally good to judge ourselves lowly or poorly, but because a liability to it requires that we recognize that we are not always the people we take ourselves to be."[43] Accordingly, she takes issue with conceptions of shame that sees it as an emotion of self-protection, although she does so without offering any discussion of the positions that argue thus.[44]

Thomason's proposal nevertheless contributes to a more comprehensive understanding of the phenomenon of shame, which the following description tries to summarize: Shame is rooted in the specific relational mode of being in the world that humans exist as since they are intentional beings. Shame is a *composite* phenomenon that involves various combinations of psychological, social, and embodied experiences of self. This point speaks in favor of seeing shame as more than an emotional reaction to *one* set of conditions. Shame is the result of a diversity of types of interplay between different realms of experience in which an agent participates, be it internal psychological, be it socio-cultural, or be it conditions that have to do with one's body. Let us try to elaborate this from a phenomenological point of view: Because human beings are constituted by their relations and are intentional at the same time, they are directed towards others, towards the world, and involved in different kinds of projects. The notion "projects" is important here, since it captures the intentional character of the self as someone who projects herself into the world and towards different objects or aims that she wants to achieve or accomplish. Intentionality is here understood as an orientation towards something and has to do with how the self manifests an interest that is directed and shaped by the relation to this something.[45] This intentional and projective character of being (which Heidegger calls Da-sein, being-towards), is not based only on intellectual deliberations: it may also be rooted in instinctual elements (as in the infant seeking the breast for food) or in desires that emerge as the result of interactions with others (as in Girard's mimetic desire, which

[41] Ibid. p. 21.
[42] Ibid.
[43] Ibid., p. 22.
[44] Ibid. p. 21-22.
[45] Another way to express this intentionality is by means of the notion "interest". In her book on shame Probyn comes close to the description I develop here, especially with regard to relationality and interruption: "Interest constitutes lines of connection between people and ideas. It describes a kind of affective investment we have in others. When, for different reasons, that investment is questioned, and interest interrupted, we feel deprived. Crucially, that's when we feel shame. That little moment of disappointment— 'oh, but I was interested'—is amplified into shame or a deep disappointment in ourselves. Shame marks the break in connection. We have to care about something or someone to feel ashamed when that care and connection— our interest—is not reciprocated." Probyn, Elspeth. *Blush: Faces of Shame*, Minneapolis: University of Minnesota Press, 2005, p. 13.

implies, e.g., that an infant wants to have what another has, simply because the other has it).[46] There is a personal *investment* in these projects, and I hold that this personal investment in a necessary component for shame to appear. I therefore also hold that the instinctual and/or desire-based intentions not only are an expression of the embodied self but that the embodied mode of intentionality that is articulated in the self's projects and projections are of crucial importance for understanding shame. This would imply that to see shame as manifesting a tension between our self-conception and our identity, or as the result of feeling defined or overshadowed by one aspect of one's identity, as Thomason's account suggests, is not sufficient. The intentional projective mode of being implies that the self understands herself in relation to something that is of importance or value to her, something she wants to achieve. What she seeks to achieve is not simply external to her or of instrumental value but is linked to one's sense of self. Thus, she invests herself in these projects, and the projects become expressions of her intentions: she may want to be fed, sexually satisfied, recognized as the bearer of a specific status, admired, considered as skilled, worthy of recognition, etc. Often, such projects turn out to be successful. Then shame is not an option. The self-image or self-conception is maintained and confirmed in agency.

Shame occurs when there is a rupture in, or interruption of, these intentions and desires and the projects they engender – in a way that makes it impossible to enact agency in a way that appears as unproblematic to the agent. Thereby, shame manifests itself in the lack of ability to fulfill the intended project or achieve the desired aim in a coherent manner.[47] Part of this implies a temporary loss of the capacities for agency. Coherence means here that there is an undisturbed consonance between intentions and desires, actions, values, and the projected result of these. The rupture implies that the intended project is blocked or impeded. This can happen in a wide variety of ways, spanning from moral rebuke to ridicule because of hair color. The problematization of the intentional project and the investment therein is not necessarily mediated by the intervention of others: it may be that the person in question realizes that she is not competent to fulfill the intentions or come to see that the project implies a way of appearing that is not desirable

[46] When Merleau-Ponty understands consciousness as a kind of bodily understanding, it can be related to the idea about intentionality's relation to shame that I sketch here. He argues that our exploratory and goal-directed movement itself constitutes a way of being conscious of things and a form of understanding of what is perceived that is not derived from activities of conceptual categorization and inference. In other words: intentionality does not require intellectual deliberation. Siewert, Charles, "Consciousness and Intentionality", *The Stanford Encyclopedia of Philosophy:* Spring 2017 Edition, Edited by Edward N. Zalta. Online at https://plato.stanford.edu/–archives/spr2017/entries/consciousness-intentionality/ (accessed Jan. 21, 2020).

[47] On the basis of his affect theory, Silvan Tomkins describes shame as "inevitable for any human being insofar as desire outruns fulfillment sufficiently to attenuate interest without destroying it. The most general sources of shame are the varieties of barriers to the varieties of objects of excitement or enjoyment, which reduce positive affect sufficiently to activate shame, but not so completely that the original object is renounced: "I want, but—" is one essential condition for the activation of shame. Clearly not all barriers suspend the individual between longing and despair. Many barriers either completely reduce interest so that the object is renounced, or heighten interest so that the barrier is removed or overcome. Indeed, shame itself may eventually also prompt either renunciation or counteraction inasmuch as successful renunciation or counteraction will reduce the feeling of shame. We are saying only that whatever the eventual outcome of the arousal of shame may be, shame is activated by the incomplete reduction of interest–excitement or enjoyment—joy, rather than by the heightening of interest or joy or by the complete reduction of interest or joy." See Silvan S. Tomkins, *Affect Imagery Consciousness [Electronic Resource] : The Complete Edition*, *Ebook Central* (New York: Springer Pub., 2008). Book 2, Vol. 1, p. 388.

after all. However, it may also be that others react to the project in ways that engender shame, as when the infant is rejected in her intention to be fed, or when one realizes that the project one is investing in is considered by others to be morally repugnant.

The interruption of the intentional project that engenders shame is, therefore, more than an experience of failure to achieve the desired good. Accordingly, shame not is connected exclusively to the experience of being unable to reach the desired good. It may also be constituted by other things than the split between self-conception and identity. It may comprise an experience of failure or lack of ability to act in ways that can lead to the desired result, or it may be an experience of the desire or intention itself as failed, or as considered by others as objectionable. The *frustration* of the desire that leads to the intentional projects is among the elements that allow us to see shame as an embodied phenomenon: shame could not occur if there was not an agent that had intentions fueled by a desire denied for the assumed good. It can be argued that such frustration is the cause of violent responses and what results in aggression against a (perceived or imagined) interrupter of the coherent agency.

Thomason's understanding of shame can nevertheless be incorporated into this sketch because she so clearly underscores the loss of control and the need for overcoming a mere moral definition of shame. She also underscores the role of the self-conception or self-image. Against the backdrop of the sketch above, it is also possible to account for violent reactions engendered by shame, without having to declare them as rational or assess them from a moral point of view. Shame is experienced as an interruption in which the experience of failure is globalized when it takes over the self. The interruption manifests the distinction between intention/project and realization, in a way that may, but need not, be manifested in the split between self-conception and identity (Thomason). The concomitant lack of control furthermore manifests itself in the lack of coherent agency. As a result, the self withdraws from the project because it experiences the vulnerable character of its own investment – and this is something that scholars may see as the self-protective feature of shame, but we do better in assuming that such protective moves are results of shame and not a feature that belongs to shame itself.

Jan-Olav Henriksen, MF Norwegian School of Theology, Religion and Society
jan.o.henriksen@mf.no

Bibliography

Deigh, John. "Shame and Self-Esteem: A Critique.", *Ethics* 93:2 (1983), pp. 225-45.

Deonna, Julien A., Raffaele Rodogno, and Fabrice Teroni. *In Defense of Shame: The Faces of an Emotion*. New York: Oxford University Press, 2011.

Honneth, Axel. *The Struggle for Recognition*: *The Moral Grammar of Social Conflicts*. Cambridge: Polity Press, 1995.

Kekes, John. "Shame and Moral Progress." *Midwest Studies in Philosophy* 13:1 (1988), pp. 282-96.

Manion, Jennifer. "The Moral Relevance of Shame." American Philosophical Quarterly 39, no. 1 (2002): 73-90.

Probyn, Elspeth. *Blush: Faces of Shame*. Minneapolis: University of Minnesota Press, 2005.

Rawls, John. *A Theory of Justice*. Cambridge, Mass.: Belknap Press, 1972.

Siewert, Charles, "Consciousness and Intentionality", *The Stanford Encyclopedia of Philosophy*: Spring 2017 Edition, Edited by Edward N. Zalta. Online at https://plato.stanford.edu/–archives/spr2017/entries/consciousness-intentionality/ (accessed Jan. 21, 2020).

Taylor, Gabriele. *Pride, Shame, and Guilt: Emotions of Self-Assessment*. Oxford: Clarendon Press, 1985.

Thomason, Krista K. "Shame, Violence, and Morality." *Philosophy and Phenomenological Research* 91:1 (2015): 1-24.

Thomason, Krista K. *Naked: The Dark Side of Shame and Moral Life*. New York: Oxford University Press, 2018.

Tomkins, Silvan S. *Affect Imagery Consciousness [Electronic Resource]: The Complete Edition*. Ebook Central. New York: Springer Pub., 2008.

Vetlesen, Arne Johan. *Studier I Ondskap*. Oslo: Universitetsforlaget, 2014.

Distributive Energy Justice and the Common Good

Anders Melin

Recently, philosophers and social scientists have shown increased interest in questions of social, global, and intergenerational distributive justice related to energy production and consumption. However, so far there have been only a few attempts to analyse questions of distributive energy justice from a religious point of view, which should be considered a lack since religions are an important basis of morality for a large part of the world's population. In this article, I analyse issues of distributive energy justice from a Christian theological viewpoint by employing the Catholic common good tradition as a theoretical framework. First, I present and argue for a global and ecological interpretation of the Catholic common good tradition. Then I analyse the implications of such an interpretation on questions of distributive energy justice, focusing on the view of property rights within the Catholic common good tradition. I conclude that, in comparison with Nussbaum's liberal capabilities approach, the common good tradition provides stronger reasons for individuals and groups in more economically developed countries to share their resources and knowledge with individuals and groups in less economically developed countries.

Introduction

Energy is essential for human life, especially in today's economically developed countries, where the population depends on easy access to energy. The science of economics defines energy as 'the capacity to perform work, useful for human beings, thanks to changes in the structure of matter or its position in space. These changes are not free, but imply some cost or effort'.[1] Energy is created using limited resources, which gives rise to questions of national, global, and intergenerational distributive justice. Moreover, questions of justice also arise due to the negative impacts of energy production and consumption, which are often unequally distributed.

Recently, there has been an increased interest in questions of distributive energy justice among both philosophers and social scientists.[2] However, so far there has been only

[1] Astrid Kander, Paolo Malanima, and Paul Warde, *Power to the People: Energy in Europe over the Last Five Centuries* (Princeton: Princeton University Press 2013), p. 18.

[2] See, for example, Benjamin K. Sovacool, *Energy & Ethics: Justice and the Global Energy Challenge*, (New York: Palgrave Macmillan 2013); Rosie Day, Gordon Walker, and Neil Simcock, 'Conceptualising energy use and energy policy using a capabilities framework', *Energy Policy* 93

a few attempts to analyse questions of distributive energy justice from a religious point of view,[3] which should be considered a lack since religions are an important basis of morality for a large part of the world's population.[4] In this article, I analyse issues of distributive energy justice from the perspective of Christian theology. Since religions vary, it seems necessary to focus on one specific religious tradition. The common good tradition, especially its contemporary Catholic interpretations, is used as a theoretical framework since it is an important tradition within Christian social thought. Since the Catholic common good tradition rests on a contentious metaphysical worldview, it is questionable whether it can be regarded as an ethical theory that should be universally accepted.[5] However, the common good tradition is at least relevant for many Christian individuals and communities and possibly some other religious believers.[6]

First, I present and argue for a global and ecological interpretation of the Catholic common good tradition based on a critical reading of some recent contributions to the debate. Then, I discuss the implications of such an interpretation on questions of distributive energy justice, especially regarding the view of property rights within the Catholic common good tradition. In order to clarify the practical relevance of this view of property rights, I compare the Catholic common good tradition with a justice theory that has a more liberal view of property rights and is influential in today's philosophical debate. I have chosen Martha C. Nussbaum's capabilities approach since it has both interesting similarities and differences with the Catholic common good tradition. On the one hand, Nussbaum advocates an Aristotelian view of humans as social beings that closely resembles the one in the Catholic common good tradition, and on the other hand, she describes her approach as a form of political liberalism and gives entitlements a more primary role than duties.

(2016); and Behnam Taebi, Sabine Roeser, *The Ethics of Nuclear Energy: Risk, Justice, and Democracy in the post-Fukushima Era*, (Cambridge: Cambridge University Press 2015). Our duties to non-human life forms are sometimes discussed within the framework of justice, see, Brian Baxter, *A Theory of Ecological Justice* (New York: Routledge 2005). However, due to the space limitations, I do not include the moral relationship to non-humans in my analysis here.

[3] Some examples are the following: Pontifical Council for Justice and Peace, *Energy, Justice, and Peace: A Reflection on Energy in the Current Context of Development and Environmental Protection*, (New York: Paulist Press, 2014); Sigurd Bergmann, *Religion, Space, and the Environment* (New Brunswick, London: Transaction Publishers, 2014), pp. 315-322; and Larry L. Rasmussen, 'Energy: the Challenges to and from Religion', *Zygon*, 46:4 (December 2011).

[4] William Schweiker (ed.), *The Blackwell Companion to Religious Ethics* (Malden, MA: Blackwell 2008).

[5] David S. Oderberg, "The Metaphysical Foundations of Natural Law" in *Natural Moral Law in Contemporary Society*, edited by Holger Zaborowski (Washington, D.C.: Catholic University of America Press 2010)

[6] Daniel P. Scheid argues that the common good tradition shares many characteristics with ethical concepts within Hinduism, Buddhism, and American Indian Traditions. See Daniel P. Scheid, *The Cosmic Common Good: Religious Grounds for Ecological Ethics* (Oxford: Oxford University Press, 2016).

A Global and Ecological Interpretation of the Common Good

The common good tradition has a long history within Christianity, especially within Catholicism, although it is also found in the theologies of the Protestant Reformers and in contemporary Protestant ethics.[7] The notion of the common good has several different meanings. It is sometimes used to denote the noble and immaterial ends that humans should strive for, such as justice. Steven A. Long, for example, describes the common good as primarily a theological and moral principle. In order for something to be a genuine common good, it should participate in the ultimate common good of the universe that is God. According to the Catholic understanding, as formulated by Thomas Aquinas, all things are directed to God as their end.[8] There is a teleological hierarchy of common good that ends in the celestial city of the saints.[9]

Other interpretations of the notion of the common good are more focused on the material well-being of humans. Some interpretations limit the common good to those goods that can only be held or enjoyed in common, such as clean air, whereas other interpretations include all natural resources that humans need for their survival and well-being, even those that individuals can enjoy in isolation.[10] John Hart, for example, offers two meanings for the term: (1) the collective well-being of a community and (2) the goods of the earth that to some extent should be shared.[11] Hart claims that natural resources should be considered common goods when they are needed for the subsistence and well-being of the members of the community; the basic needs of all its members should be met. In addition, we should also show concern for future generations.[12] A similar interpretation of the Catholic common good tradition is put forward by Daniel P. Scheid, who argues that the common good encompasses those material goods that every individual needs, such as food and clothing.[13] According to Scheid, the belief in the inherent dignity of all humans – based on the theological doctrine that they are created in the image of God – is an important aspect of the Catholic common good tradition.[14] Moreover, the mainstream Catholic view, which is influenced by Aristotelian anthropology, regards the human person as a unit of body and soul; therefore, it is necessary to recognize that humans have both bodily and spiritual needs.[15]

[7] See, for example, Benjamin Myers, 'Rights, Resistance and the Common Good: Calvin's Political Theology', *Uniting Church Studies,* 17:1 (2011); Wanda Deifelt, 'Seeking the Common Good: Lutheran Contributions to Global Citizenship', *Intersections*, 29 (2009); and Marcia Pally, *The New Evangelicals: Expanding the Vision of the Common Good* (Rapids, MI: Eerdmans, 2011).
[8] Steven A. Long, 'Understanding the Common Good', *Nova et Vetera*, 16:4 (2018), p. 1143.
[9] Long, 'Understanding the Common Good', p. 1148.
[10] Waheed Hussain, 'The Common Good' in *Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta (published 2018-02-26), John Hart, *Sacramental Commons: Christian Ecological Ethics* (Lanham, MA: Rowman & Littlefield, 2006).
[11] Hart, *Sacramental Commons,* p. 147.
[12] Ibid., pp. 149-152.
[13] Daniel P. Scheid, *The Cosmic Common Good: Religious Grounds for Ecological Ethics* (Oxford: Oxford University Press, 2016), p. 16.
[14] Scheid, *The Cosmic Common Good*, pp. 15-16.
[15] Joseph Koterski, 'Human Nature from a Catholic Perspective', *American Journal of Economics and Sociology*, 71 (4) 2012, pp. 810-812.

Since this article is concerned with the material welfare of humans in relation to energy production and consumption, I focus on material well-being in my interpretation of the concept of the common good. I employ a wide interpretation that encompasses both goods that members of a society share and goods that individuals can enjoy in isolation and that are necessary for their individual subsistence and well-being. This interpretation is based on the belief in God as the proper owner of all of earth's resources. From a theological point of view, these resources are God's gift to all of humanity (see below).

One central element in both classical and contemporary interpretations of the Catholic common good tradition is the belief that promoting the common good requires prioritizing the needs of individuals and the community over individual preferences.[16] Although the individual is important, he or she has both civic and Christian responsibilities since he or she exists in social relationships. The individual has a duty to work for the common good, even at his or her own expense.[17] While more individualistic views of society see the good of a community primarily as the sum of its members' interests, the Catholic common good tradition puts more emphasis on cooperation and shared social objectives.[18]

Recently, the Christian common good tradition has been called into question for arguably promoting the self-interest of certain social groups. For example, in the history of Christianity, women have often been encouraged to a greater extent than men to neglect their individual welfare for the sake of the good of the family or community.[19] However, modern interpreters of the Catholic common good tradition consider this critique and emphasize respect for individual human dignity and human rights as an important part of the tradition.[20]

The Catholic common good tradition differs from the natural right tradition, which is an important point of departure for contemporary liberal philosophers. While the former considers humans social by nature and created to live in communities, the latter sees humans as pre-political individuals that have certain rights (often understood as negative rights). The natural rights tradition also sees society as based on a contract between individuals, who are not by necessity social. Further, in the Catholic common good tradition, the individual's duties to the community have a higher priority than his or her rights, which are regarded as means necessary to fulfil duties. The natural right tradition

[16] Hart, *Sacramental Commons*, pp. 149-151; George J. Lavere, 'The problem of the common good in Saint Augustine's *Civitas Terrena*', *Augustinian Studies*, 14 (1983); Richard A. Crofts, 'The Common Good in the Political Theory of Thomas Aquinas', *The Thomist: A Speculative Quarterly Review*, 37:1 (January 1973).

[17] Lavere, 'The problem of the common good in Saint Augustine's *Civitas Terrena*'; Crofts, 'The Common Good in the Political Theory of Thomas Aquinas'.

[18] John Langan, S.J. 'Common Good' in *A New Dictionary of Christian Ethics*, edited by John Macquarrie, James Childress (London: SCM Press, 1986), p. 102.

[19] For a feminist critique of the common good tradition, see, for example, Susanne DeCrane, *Aquinas, Feminism, and the Common Good*, (Washington, D.C.: Georgetown University Press, 2004).

[20] See, for example, David Hollenbach, *The Common Good and Christian Ethics* (Cambridge: Cambridge University Press, 2002); Eric Mount, Jr., *Covenant, Community and the Common Good: An Interpretation of Christian Ethics* (Cleveland, OH: The Pilgrim Press, 1999).

instead emphasizes individual rights.[21] Of course, it is a contentious issue which of the two traditions is the most tenable and convincing. However, it can be argued that the common good tradition fits better with the Christian tradition as a whole since it is more in line with the Biblical message. In both the Old and the New Testament, covenant is a central concept and a point of departure for moral rules. The Bible stresses that humans are created to live in community with one another and with God.[22]

One element of the Catholic common good tradition that is especially relevant to distributive justice is its view on property rights. The ethics of property within this tradition rests on two pillars: the affirmation of private property and the belief in the universal destination of goods. This double basis is found already in the writings of Thomas Aquinas, and it is further developed in some papal encyclicals and other important documents from the Catholic Church from the end of the nineteenth century onwards.[23] On the one hand, the Catholic common good tradition affirms the importance of private property. Since humans are created as not only spiritual but also physical beings, they need certain external goods for their survival. Humans have a duty of self-preservation and, therefore, the right to such goods. Moreover, as the image of God, humans have sovereignty over the earth and authorization to place necessary goods in their service, which is generally accomplished through work. In this way, humans acquire property.[24] Furthermore, the right to property is regarded as necessary for the autonomy of persons and their ability to function in society – it is essential for human freedom.[25]

On the other hand, the natural right to private property is not absolute in the Catholic common good tradition; it is 'subordinate to the universal destination of goods', that is, the common right of all people to usage of goods. Property rights should always be exercised for the common good.[26] However, given the human condition, the Catholic tradition argues that common property and lack of private property rights would lead to sloth and disorder. Private property is therefore the best means for achieving the common good.[27] Nevertheless, if proprietors forget their social duties, expropriation can sometimes be justified.[28]

[21] J. Bryan Hehir, 'The modern Catholic Church and human rights', in *Christianity and Human Rights: An Introduction*, edited by John Witte, Jr., Frank S. Alexander (Cambridge: Cambridge University Press, 2010), pp. 118-119.

[22] Mount, *Covenant, Community, and the Common Good*, pp. 14-15.

[23] Manfred Spieker, 'The Universal Destination of Goods: The Ethics of Property in the Theory of a Christian Society', *Journal of Markets & Morality*, 8 (2), Fall 2005, pp. 334-337; B. Andrew Lustig, 'Property and Justice in the Modern Encyclical Literature', *The Harvard Theological Review*, 83:4 (Oct 1990); Thomas Aquinas, *Summa Theologica* (translated by fathers of the English Dominican Province), (Notre Dame, IN: Ave Maria Press, 1948), 2a-2ae, q. 32, a. 5, ad. 2; Leo XIII, *Rerum Novarum*, 1891; Paul VI, *Gaudium et Spes*, 1965; John Paul II, *Laborem Exercens*, 1981; and John Paul II, *Centesimus Annus*, 1991.

[24] Spieker, 'The Universal Destination of Goods', p. 336; John Paul II, *Laborem Exercens,* § 12; John Paul II, *Centesimus Annus*, § 31.

[25] Paul VI, *Gaudium et Spes*, § 71.

[26] Spieker, 'The Universal Destination of Goods', pp. 336-337; John Paul II, *Centesimus Annus,* § 30; Pontifical Council for Justice and Peace, *Compendium of the Social Doctrine of the Church*, 2004, § 328-329.

[27] Spieker, 'The Universal Destination of Goods', p. 337.

[28] Ibid., p. 341.

In today's theological debate, the principle of universal destination of goods is also applied on the global level. Traditionally, the nation-state was conceived of as the boundary for the common good. However, after World War II, an international or global interpretation of the common good concept has become increasingly prevalent.[29] Such an interpretation is put forward in several papal encyclicals,[30] as well as by individual Catholic theologians.[31] For example, the contemporary theologian David Hollenbach argues for a conception of the common good that includes the whole world. He claims that in the current state of affairs, the well-being of individuals in different parts of the world is increasingly interdependent. Actions conducted in one country may affect both the economic status of individuals in other countries and the environmental conditions under which they live. Hollenbach argues that current trends of globalization call for a widened understanding of the community of which we are a part.[32] The realization of the global common good requires respecting the civil and political rights, as well as the social and economic rights, of all humans on the planet.[33] Another important aspect of the Catholic common good tradition that is relevant for issues of global justice is the preferential option for the poor, which demands special concern for underprivileged groups in less economically developed countries.[34]

I agree with Hollenbach that a contemporary interpretation of the Catholic common good tradition needs to recognize that we also have duties to individuals living in other parts of the world. However, the increased interaction today between individuals in different countries and continents is neither a necessary nor a sufficient reason for concluding that we have duties to citizens in other nation-states. Since the Catholic common good tradition is not a contractarian theory, social cooperation with people in other parts of the world is not a precondition for having duties to them. Moreover, the mere fact that we consume products manufactured by people in other countries is not a sufficient reason for concluding that we have such duties. From a rational self-interest standpoint, we have limited reasons to be concerned about humans in other countries. Instead, from the Catholic common good perspective, our duties to people in other parts of the world follow from the belief that all humans have dignity since they are created in the image of God.

In addition, contemporary Catholic social teaching often includes an ecological interpretation of the common good tradition, which treats the environment as a common resource for all humans.[35] As seen above, contemporary theologians such as Hart and Scheid argue that the earth as a whole with all its natural resources is a common good that should be preserved for the sake of humanity, both current and future humans. In light of

[29] William A. Barbieri, Jr., 'Beyond the Nations: The Expansion of the Common Good in Catholic Social Thought', *The Review of Politics*, 63:4 (Autumn, 2001).

[30] See, for example, John XXIII, *Pacem in Terris*, 1963; John Paul II, *Sollicitudo Rei Socialis,* 1987; and Francis, *Laudato Si'*, 2015.

[31] See, for instance, Jacques Maritain, *Man and the State* (Chicago: University of Chicago Press, 1951) and John Courtney Murray, *We Hold These Truths: Catholic Reflections on the American Proposition* (New York: Sheed and Ward, 1960)

[32] Hollenbach, *The Common Good and Christian Ethics*, pp. 212-217.

[33] Ibid., pp. 220-227.

[34] See, for example, Francis, *Evangelii Gaudium*, § 194.

[35] See, for example, Pontifical Council for Justice and Peace, *Compendium of the Social Doctrine of the Church*, § 466-467; Francis, *Laudato Si´*.

the contemporary scientific understanding of humanity's dependency on nature, it seems necessary to include natural resources among those goods that humans need for their well-being and to acknowledge the ways in which environmental destruction threatens the welfare of humans.

To conclude, the theoretical framework of this article is based on a global and ecological interpretation of the Catholic common good tradition, focused on the material well-being of humans. Accordingly, the protection of the environment is a precondition for the well-being of communities and their individuals. The resources of the earth are regarded as parts of the common good. Therefore, property rights are not seen as absolute but as subordinate to the material well-being of all humanity. Out of respect for their human dignity, both current and future humans should be given the opportunity to satisfy their material needs. As we will see in the next section, this interpretation of the Catholic common good tradition has important implications for distributive energy justice.

Implications for Energy Justice

The Catholic common good tradition views property rights differently than the liberal theories of justice that dominate today's philosophical debate.[36] As previously mentioned, the Catholic common good tradition is based on the view that God is the proper proprietor of the earth's resources, which makes property rights subordinate to the common good.

A report on energy and justice by The Pontifical Council for Justice and Peace explicitly relates the question of property rights to questions of energy production and consumption. It affirms the view that energy should be regarded as a common good – a gift from God to the whole of humanity. Therefore, every human is called to share energy resources with those who need it.[37] According to the report, the view that energy and other natural resources are a common good leads to a demand for redistribution, both at the national and the global level.[38] Moreover, it is important to transfer scientific and technological knowledge to less economically developed countries.[39]

The Pontifical Council for Justice and Peace seems right in pointing out that the Catholic common good tradition leads to a demand for redistribution of energy sources at both the national and global level. As stated in the scientific debate on energy justice, there are large inequalities in energy consumption, both within and between countries. While members of the richest segments of the population in some of the more economically developed countries drive around in SUVs and travel to Thailand or Hawaii on vacation, members of the poorest segments can barely afford to pay the electricity bill. Energy poverty is a common phenomenon also in many of the more economically developed

[36] Some other positions within contemporary political philosophy, such as communitarianism and Marxism, have a view of property rights that more closely resembles the common good tradition than liberalism. However, since they are less influential in today's debate and since I do not have space here for a more detailed comparison of the common good tradition with these forms of political philosophy, I limit this study to a comparison with a liberal view of property rights.
[37] Pontifical Council for Justice and Peace, *Energy, Justice, and Peace*, pp. 2-4.
[38] Ibid., p. 24.
[39] Ibid., p. 65.

countries.[40] The differences are even more apparent if one compares the energy consumption in more and less economically developed countries. In the latter group of countries, around 1.1 billion live without electricity and 2.8 billion lack access to clean cooking facilities.[41] A large part of the world's population is dependent on wood, charcoal, and dung for cooking, which causes a very high number of premature deaths each year due to indoor air pollution.[42]

From the Catholic common good perspective, countries cannot claim an exclusive right to certain resources just because they happen to own them. Individuals in less economically developed countries are entitled to a higher energy consumption than they currently have to live a dignified life, and more economically developed countries are required to share some of their resources, both material resources and technological expertise.

Moreover, according to the Catholic common good tradition, all humans have the right to environmental conditions that make it possible to live a dignified life. Therefore, we should take into account the environmental consequences of energy production and consumption – in terms of their unequal distribution between countries as well as their effect on future generations. For instance, more economically developed countries have higher greenhouse gas (GHG) emissions. At the same time, the less economically developed countries often suffer more from the negative environmental effects of global warming – such as increased frequencies of heat waves, droughts, and floods – partly because they have lower capacity to adapt to climate change. They are also likely to be more seriously affected by climate change in the near future. The Intergovernmental Panel on Climate Change's fifth assessment report states that in the coming decades global warming is likely to decrease food security and make poverty reduction more difficult in the developing world.[43] We also need to consider that our GHG emissions will have consequences for the ecosystems several thousand years into the future and will therefore have a significant negative impact on future humans.[44]

[40] Harriet Thomson, Carolyn Snell, and Stefan Bouzarovski, 'Health, Well-Being and Energy Poverty in Europe: A Comparaive Study of 32 European Countries', *International Journal of Environmental Research and Public Health*, 14:6 (2017).

[41] The webpage of the International Energy Agency, https://www.iea.org/energyaccess/ (accessed 2018-06-20).

[42] Benjamin K. Sovacool and Michael H. Dworkin, *Global Energy Justice: Problems, Principles, and Practices*, (Cambridge, UK: Cambridge University Press, 2014), pp. 228-229.

[43] IPCC, 'Summary for policymakers', in *Climate Change 2014: Impacts, Adaptation, and Vulnerability. Par A: Global and Sectoral Aspects. Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change,* edited by C.B. Field, V.R. Barros, D.J. Dokken, K.J. Mach, M.D. Mastrandea, T.E. Bilir, M. Chatterjee, K.L. Ebi, Y.O. Estrada, R.C. Genova, B. Girma, E.S. Kissel, A.N. Levy, S. MacCracken, P.R. Mastrandea, and L.L. White, (Cambridge, UK and New York, NY: Cambridge University Press, 2014).

[44] M. Colllins, R. Knutti, J. Arblaster, J.-L. Dufresne, T. Fichefet, P. Friedlingstein, X. Gao, W.J. Gutowski, T. Johns, G. Krinner, M. Shongwe, C. Tebaldi, A.J. Weaver, and M. Wehner, 'Long-term Climate Change: Projecctions, Commitments and Irreversibility', in *Climate Change 2013: The Physical Science Basis. Contributions of Working Group 1 to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, edited by T.F. Stocker, D. Qin, G.-K. Plattner, M. Tignor, S.K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex, and P.M. Midgley (Cambridge: Cambridge University Press, 2013).

The Catholic common good tradition obligates us to develop an energy system that is just from a social, global, and intergenerational perspective. To live a dignified life, both current and future humans in different parts of the world need access to energy as well as decent environmental conditions. We must phase out fossil fuels, or at least drastically reduce their use, before the middle of this century in order to save large parts of the planet from becoming uninhabitable in the future. Fossil fuels will have to be replaced by renewables, such as wind and solar power as well as biofuels; however, according to common predictions, it will be difficult to replace fossil fuels completely with renewables. It is likely that we will have to reduce our energy consumption.[45] Since the earth's population is expected to continue increasing, although at a less rapid pace,[46] the global per capita consumption of energy in 2050 needs to be significantly lower than the present. Simultaneously, individuals in less economically developed countries need to increase significantly their use of energy to live a dignified life. Accordingly, most individuals in the more economically developed countries have to reduce drastically their energy use. This requires a radical shift of life style – and perhaps also the current economic and political systems – which is likely to be unpopular and difficult to accomplish. Nevertheless, it is warranted from the perspective of the Catholic common good tradition.

Nussbaum's capabilities approach can also justify certain forms of resource redistribution. However, since it regards entitlements as more fundamental than responsibilities, it is questionable whether it can achieve the same results as the Catholic common good tradition. Nevertheless, the capabilities approach is rather similar to the Catholic common good tradition since both theories draw inspiration from Aristotle's view of humans and society.[47] Nussbaum's point of departure is an Aristotelian conception of individuals as social by nature.[48] At the same time, she states that her capabilities approach is a form of political liberalism in accordance with John Rawls's definition; namely, the list of capabilities should be based on an 'overlapping consensus' between different 'comprehensive doctrines' – that is, different philosophical or religious worldviews.[49] Furthermore, entitlements are more fundamental than duties in her approach (see below), which gives it a clearly liberal character.[50]

The key element of Nussbaum's theory of justice is her list of capabilities. She argues that the minimal requirement of justice is that each individual reaches a threshold level of the following ten capabilities: (1) Life, (2) Bodily Health, (3) Bodily Integrity, (4) Senses, Imagination, and Thought, (5) Emotions, (6) Practical Reason, (7) Affiliation, (8) Other Species, (9) Play, and (10) Control over One's Environment (Political and Material).

[45] Richard Heinberg and David Fridley, *Our Renewable Future: Laying the Path for One Hundred Percent Clean Energy*, (Washington, D.C.: IslandPress, 2016), pp. 7-9.
[46] United Nations, Department of Economic and Social Affairs, Population Division, *World Population Prospects: the 2017 Revision, Key Findings and Advance Tables*, Working Paper No. ESA/P/WP/248 (UN: New York, 2017).
[47] Mary M. Keys, *Aquinas, Aristotle and the Promise of the Common Good,* (Cambridge: Cambridge University Press, 2008); Martha C. Nussbaum, *Frontiers of Justice: Disability, Nationality, Species Membership*, (Cambridge, MA, London, UK: The Belknap Press of Harvard University Press, 2006), pp. 273-274.
[48] Nussbaum, *Frontiers of Justice*, pp. 273-274.
[49] Martha C. Nussbaum, *Women and Human Development: The Capabilities Approach*, (Cambridge: Cambridge University Press, 2000), p. 5.
[50] Nussbaum, *Frontiers of Justice*, pp. 279-281.

Nussbaum sees the different capabilities as separate components, so a lack of one of them cannot be compensated for by a high level of another.[51] Nussbaum's list of capabilities is closely related to her concept of dignity since she considers the possession of these capabilities necessary for living with human dignity. In order to determine which capabilities are essential, we should find a way of living and acting that is compatible with human dignity.[52]

Further, in Nussbaum's most detailed discussion of dignity in chapter 14 of *Human Dignity and Bioethics: Essays Commissioned by the President's Council on Bioethics*, she endorses an Aristotelian-Marxian view of dignity. Nussbaum states that such a conception recognizes that humans are vulnerable and needy beings who require help from others.[53] Simultaneously, she recognizes that any conception of dignity used to form the basis of political principles should not be linked to any specific philosophical and religious worldview. Therefore, it cannot be based on an Aristotelian view of human flourishing.[54] In *Creating Capabilities*, Nussbaum puts forward another conception of human dignity that seems to be more compatible with her framework of political liberalism. She describes the following characteristics as important for her view of dignity: (1) respect, beings with dignity should be respected by others; (2) agency, people's ability to manage their own lives should be promoted; and (3) equality, all individuals are worthy of the same respect.[55]

As for Nussbaum's view on property rights, she does not include the ability to hold property in her earlier formulation of her list of capabilities. However, in *Women and Human Development*, she describes the ability to hold property as part of the capability to have control over one's environment. Moreover, she states that it is not only a question of formal ability but also a question of real opportunity.[56] In other words, everyone should have the right to actually own a certain amount of property. However, in *Creating Capabilities*, Nussbaum has omitted the latter statement, so it is unclear whether she still maintains this view.[57] For the sake of the argument, I assume that she does.

Concerning Nussbaum's view of global justice, she claims in *Frontiers of Justice* that our world is not minimally just unless all the people of the world reach a certain threshold of the ten capabilities.[58] According to Nussbaum, entitlements are more fundamental than duties, which are derived from entitlements. All humans deserve to live a dignified life, and this moral claim justifies a collective obligation for rich individuals and countries to help poor individuals and countries.[59] Nussbaum argues that all humans have equal dignity, which justifies equal political rights for all citizens. However, the fact that all humans have equal dignity does not justify equal distribution of property. Nussbaum

[51] Nussbaum, *Women and Human Development*, pp. 78-81.
[52] Nussbaum, *Frontiers of Justice*, pp. 161-162.
[53] Martha C. Nussbaum, 'Chapter 14: Human Dignity and Political Entitlements' in *Human Dignity and Bioethics: Essays Commisioned by the President's Council on Bioethics*, (Washington, 2008), p. 3.
[54] Nussbaum, 'Human Dignity and Political Entitlements', p. 4.
[55] Martha C. Nussbaum, *Creating Capabilities: The Human Development Approach* (Cambridge, MA: The Belknap Press of Harvard University Press, 2011), pp. 29-31.
[56] Nussbaum, *Women and Human Development*, p. 80. See also Rutger Claassen, 'The Capability to Hold Property', *Journal of Human Development and Capabilities,* 16:2 (2015), pp. 222-223.
[57] Nussbaum, *Creating Capabilities*, p. 34.
[58] Nussbaum, *Frontiers of Justice*, p. 281.
[59] Ibid., pp. 279-281.

argues that a dignified human life only requires a certain threshold of material property, not an equal amount of property for all citizens.[60]

Nussbaum argues that rich nations ought to give a substantial part of their GDP to poor countries. This demand follows from the principle of respecting human dignity.[61] Further, Nussbaum claims that we need some system of global governance. This system should include a tax on the more economically developed countries to promote the development of less economically developed parts of the world.[62] In addition, Nussbaum states that the focus for both institutions and individuals should be on the groups in society with an especially low quality of life, as measured by the capabilities list.[63]

Both the common good tradition and Nussbaum's capabilities approach argue for a redistribution of resources. Nussbaum's view that a dignified human life requires a certain threshold of property can justify a redistribution, at least to some extent. However, the two approaches differ since Nussbaum regards entitlements as more fundamental than duties. Her approach does not address whether (or to what extent) we are justified in infringing upon the property rights of individuals in rich parts of the world in order to realize the capabilities of individuals in poor parts. Nussbaum argues for a global tax that transfers resources from the rich parts of the world to the poor parts, but it is unclear if such a tax can be justified since she regards the capability to have control over one's environment as equally important as the capabilities of life and bodily health. In contrast, the Catholic common good tradition subordinates property rights to the common good.

Furthermore, unlike the Catholic common good tradition, the capabilities approach is intended to be only a minimal theory of justice that can be supported by different comprehensive doctrines.[64] It is not concerned with how one should handle questions of distribution above the threshold level for the different capabilities. Although the Catholic common good tradition emphasizes the satisfaction of needs, it can still justify redistributions between individuals whose basic needs are satisfied if the current distribution is considered unjust. For example, an unjust distribution could be a consequence of unjust conditions, such as the exploitation of one country by another.

Furthermore, there is a certain inconsistency in Nussbaum's description of human persons and human dignity. In some contexts, she argues for an Aristotelian conception of the human person and human dignity, but her commitment to political liberalism seems to force her to accept a 'thinner' conception. In *Creating Capabilities*, Nussbaum puts forward a view of dignity that seems more metaphysically neutral – centred on the concepts of respect, agency, and equality. However, there are many different interpretations of what these concepts mean. An interpretation that is compatible with different comprehensive doctrines is bound to be rather diluted. Since Nussbaum's view of dignity justifies her list of capabilities, her unclear conception of dignity makes it doubtful whether her view on property rights can be justified.

[60] Nussbaum, *Frontiers of Justice*, p. 293.
[61] Ibid., pp. 316-317.
[62] Ibid., pp. 319-320.
[63] Ibid., pp. 320-323.
[64] Nussbaum, *Women and Human Development*, p. 75.

Conclusions

I argue for energy distributions based on a global and ecological interpretation of the Catholic common good tradition. According to such an interpretation, the moral belief that all humans are created in the image of God requires us to also show concern for groups and individuals in other parts of the world. Moreover, in light of the current ecological awareness, we ought to regard natural resources as an important part of the common good that we need to protect.

I conclude that the Catholic common good tradition demands an energy system that is just from a social, global, and intergenerational point of view. Concern for future generations requires a significant reduction of the use of fossil fuels, which is likely to lead to a reduction of the global energy consumption per capita. At the same time, many individuals in the poor parts of the world need to increase their energy use to live a dignified life, which demands a significant reduction of energy consumption in the rich parts of the world.

In comparison with Nussbaum's capabilities approach – which is one of the most influential liberal theories of justice within the current philosophical debate – the Catholic common good tradition provides us with stronger reasons to redistribute energy and other natural resources between people in different parts of the world. Nussbaum's prioritization of entitlements instead of duties restricts the possibility of justifying limitations to individual property rights for the sake of redistributing resources.[65]

Anders Melin, Malmö University
anders.melin@mau.se

Bibliography

Aquinas, Thomas. *Summa Theologica* (translated by fathers of the English Dominican Province). Notre Dame, IN: Ave Maria Press, 1948.

Barbieri, William A., Jr. 'Beyond the Nations: The Expansion of the Common Good in Catholic Social Thought', *The Review of Politics*, 63:4 (Autumn, 2001), pp. 723-754.

Baxter, Brian. *A Theory of Ecological Justice.* New York: Routledge, 2005.

Bergmann, Sigurd. *Religion, Space, and the Environment*. New Brunswick, London: Transaction Publishers, 2014.

Claassen, Rutger. 'The Capability to Hold Property', *Journal of Human Development and Capabilities,* 16:2 (2015), pp. 220-236.

Colllins, M., R. Knutti, J. Arblaster, J.-L. Dufresne, T. Fichefet, P. Friedlingstein, X. Gao, W.J. Gutowski, T. Johns, G. Krinner, M. Shongwe, C. Tebaldi, A.J. Weaver and M. Wehner, 'Long-term Climate Change: Projections, Commitments and Irreversibility', in *Climate Change 2013: The Physical Science Basis. Contributions of Working Group 1 to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, edited by T.F. Stocker,

[65] This article has been written as part of a research project on energy and justice, funded by the Swedish Research Council, Dnr 421-2013-781.

D. Qin, G.-K. Plattner, M. Tignor, S.K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex, P.M. Midgley. Cambridge: Cambridge University Press 2013.

Crofts, Richard A. 'The Common Good in the Political Theory of Thomas Aquinas', *The Thomist: A Speculative Quarterly Review*, 37:1 (January 1973), pp. 155-173.

Day, Rosie, Gordon Walker, Neil Simcock, 'Conceptualising energy use and energy policy using a capabilities framework', *Energy Policy* 93 (2016), pp. 255-264.

DeCrane, Susanne M. *Aquinas, Feminism, and the Common Good*. Washington, D.C.: Georgetown University Press, 2004.

Deifelt, Wanda. 'Seeking the Common Good: Lutheran Contributions to Global Citizenship', *Intersections*, 29 (2009), pp. 22-27.

Francis. *Evangelii Gaudium*. 2013.

Francis. *Laudato Si*. 2015.

Hart, John. *Sacramental Commons: Christian Ecological Ethics*. Lanham, MA: Rowman & Littlefield 2006.

Hehir, J. Bryan. 'The modern Catholic Church and human rights' in *Christianity and Human Rights: An Introduction*, edited by John Witte, Jr., Frank S. Alexander. Cambridge: Cambridge University Press 2010, pp. 113-134.

Heinberg, Richard and David Fridley. *Our Renewable Future: Laying the Path for One Hundred Percent Clean Energy*. Washington, D.C.: IslandPress 2016.

Hollenbach, David. *The Common Good and Christian Ethics*. Cambridge: Cambridge University Press 2002.

Hussain, Waheed. 'The Common Good' in *Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta (published 2018-02-26).

IPPC, 2014: Summary for policymakers, In: *Climate Change 2014: Impacts, Adaptation, and Vulnerability. Par A: Global and Sectoral Aspects. Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change,* edited by C.B. Field, V.R. Barros, D.J. Dokken, K.J. Mach, M.D. Mastrandea, T.E. Bilir, M. Chatterjee, K.L. Ebi, Y.O. Estrada, R.C. Genova, B. Girma, E.S. Kissel, A.N. Levy, S. MacCracken, P.R. Mastrandea, and L.L. White. Cambridge, UK and New York: Cambridge University Press.

John XXIII. *Pacem in Terris,* 1963.

John Paul II. *Laborem Exercens,* 1981.

John Paul II. *Centesimus Annus*, 1991.

John Paul II. *Sollicitudo Rei Socialis*, 1987.

Kander, Astrid, Paolo Malanima, Paul Warde. *Power to the People: Energy in Europe over the Last Five Centuries*. Princeton: Princeton University Press, 2013.

Keys, Mary M. *Aquinas, Aristotle and the Promise of the Common Good.* Cambridge: Cambridge University Press, 2008.

Koterski, Joseph. 'Human Nature from a Catholic Perspective', *American Journal of Economics and Sociology*, 71:4 (2012), pp. 809-839.

Langan, John, S.J. 'Common Good', in *A New Dictionary of Christian Ethics*, edited by John Macquarrie, James Childress. London: SCM Press, 1986, p. 102.

Lavere, George J. 'The problem of the common good in Saint Augustine's *Civitas Terrena*', *Augustinian Studies*, 14 (1983), pp. 1-10.

Leo XIII. *Rerum Novarum*. 1891.

Long, Steven A. 'Understanding the Common Good'. *Nova et Vetera*, 16:4 (2018), pp. 1135-1152.

Lustig, B. Andrew. 'Property and Justice in the Modern Encyclical Literature', *The Harvard Theological Review*, 83:4 (Oct 1990), pp. 415-446.

Maritain, Jacques. *Man and the State*. Chicago: University of Chicago Press 1951.

Mount, Eric, Jr. *Covenant, Community, and the Common Good: An Interpretation of Christian Ethics*. Cleveland, OH: Pilgrim Press, 1999.

Murray, John Courtney. *We Hold These Truths: Catholic Reflections on the American Proposition*. New York: Sheed and Ward 1960.

Myers, Benjamin. 'Rights, Resistance and the Common Good: Calvin's Political Theology', *Uniting Church Studies,*17:1 (2011), pp. 33-43.

Nussbaum, Martha C. *Women and Human Development: The Capabilities Approach*, Cambridge, UK: Cambridge University Press 2000.

Nussbaum, Martha C. *Frontiers of Justice: Disability, Nationality, Species Membership*. Cambridge, MA, London, UK: The Belknap Press of Harvard University Press 2006.

Nussbaum, Martha C. 'Chapter 14: Human Dignity and Political Entitlements', in *Human Dignity and Bioethics: Essays Commisioned by the President's Council on Bioethics*. Washington, D.C.: 2008.

Nussbaum, Martha C. *Creating Capabilities: The Human Development Approach*. Cambridge, MA: The Belknap Press of Harvard University Press 2011.

Oderberg, David S. "The Metaphysical Foundations of Natural Law" in *Natural Moral Law in Contemporary Society*, edited by Holger Zaborowski. Washington, D.C.: Catholic University of America Press, 2010, pp. 44-75.

Pally, Marcia. *The New Evangelicals: Expanding the Vision of the Common Good*. Rapids, MI: Eerdmans, 2011

Paul VI. *Gaudium et Spes*. 1965.

Pontifical Council for Justice and Peace, *Compendium of the Social Doctrine of the Church*, 2004.

Pontifical Council for Justice and Peace. *Energy, Justice, and Peace: A Reflection on Energy in the Current Context of Development and Environmental Protection*. New York: Paulist Press 2014.

Rasmussen, Larry L. 'Energy: the Challenges to and from Religion', *Zygon*, 46:4 (December 2011), pp. 985-1002.

Scheid, Daniel P. *The Cosmic Common Good: Religious Grounds for Ecological Ethics*. Oxford: Oxford University Press 2016.

Schweiker, William (ed.). *The Blackwell Companion to Religious Ethics*. Malden, MA: Blackwell 2008.

Sovacool, Benjamin K. *Energy & Ethics: Justice and the Global Energy Challenge*. New York: Palgrave Macmillan 2013.

Sovacool, Benjamin K., Michael H. Dworkin. *Global Energy Justice: Problems, Principles, and Practices*. Cambridge, UK: Cambridge University Press, 2014.

Spieker, Manfried. 'The Universal Destination of Goods: The Ethics of Property in the Theory of a Christian Society', *Journal of Markets & Morality*, 8:2 (Fall 2005), pp. 333-354.

Taebi, Behnam and Sabine Roeser (eds.). *The Ethics of Nuclear Energy: Risk, Justice, and Democracy in the post-Fukushima Era*. Cambridge: Cambridge University Press 2015.

The webpage of the International Energy Agency, https://www.iea.org/energyaccess/ (accessed 2018-06-20).

Thomson, Harriet, Carolyn Snell, Stefan Bouzarovski. 'Health, Well-Being and Energy Poverty in Europe: A Comparative Study of 32 European Countries', *International Journal of Environmental Research and Public Health*, 14, 584, 2017, pp. 1-20.

United Nations, Department of Economic and Social Affairs, Population Division. *World Population Prospects: the 2017 Revision, Key Findings and Advance Tables*, Working Paper No. ESA/P/WP/248. UN: New York 2017.

'What Do We Do with the Art of Monstrous Men?' Betrayal and the Feminist Ethics of Aesthetic Involvement

Sarah Stewart-Kroeker

The #MeToo movement has put a spotlight on sexual harassment and abuse in a number of industries, notably the arts. It has raised a set of questions about how to receive the artistic works of the accused, particularly when such work has been beloved or formative for an individual, and collectively when it has cultural significance and influence. Claire Dederer, writing in The Paris Review, posed the question bluntly in her piece, "What Do We Do with the Art of Monstrous Men?" This question, and the range of (often perplexed) responses to it, reveal the lack of adequate resources to evaluate responses to an artist's actions that may bear on our aesthetic valuations of the artist's work and that may be experienced as quite intimately personal. What do we do with the sense of betrayal that may follow on the discovery of an artist's bad behavior? What are the implications of consuming of such art? What concepts and norms might help to guide reflection? These questions bear on the ethical significance of love and appreciation for artworks and artists, and, more broadly, the ethical consumption of artworks. This paper responds to these questions in two ways: first, it develops an account of "aesthetic involvement" to elaborate the sense of betrayal that may follow accusations or revelations of sexual harassment and abuse. Second, it proposes a feminist ethics of aesthetic involvement in response to such betrayals and to dilemmas about the individual and collective ethical consumption of artworks.

In fall 2017, a wave of accusations against "Shitty Media Men" – the title of an anonymous, crowdsourced spreadsheet naming sexual abusers and harassers in the industry[1] – spurred a series of high-profile revelations, firings, and criminal charges, as well as a massive social media campaign, #MeToo, taking up the affirmation at the heart of a movement begun by Tarana Burke more than a decade prior to help young women of color who had suffered sexual abuse. The spotlight on problems of sexual harassment and abuse has since

[1] Moira Donegan, "I Started the Media Men List. My name is Moira Donegan." *The Cut*, January 10, 2018. Online at https://www.thecut.com/2018/01/moira-donegan-i-started-the-media-men-list.html (accessed 2018-11-8).

broadened to other industries, including academia.[2] The moment has been called "a reckoning."

In the arts, this reckoning poses a particular set of questions about how to receive the artistic work of the accused, particularly when such work has been beloved or formative for an individual, and collectively when it has cultural significance and influence. Indeed, the question of engaging the creative output of artists whose ethical actions we find lacking may be the "central pop-cultural question" of the current moment.[3] The debate is being waged in newspaper columns, essays, innumerable think-pieces and hot takes, casual and not-so-casual conversation (for this reason, I draw significantly on popular materials in the following discussion). Claire Dederer, writing in *The Paris Review*, posed the question bluntly in the title of her piece, "What Do We Do with the Art of Monstrous Men?"[4] But this raises a set of subsequent questions over which she puzzles more than a little ambivalently: Ought we try to separate the art from the artist, the maker from the made? … Or do we believe genius gets special dispensation, a behavioral hall pass? And how does our answer change from situation to situation? … Do we withhold our support if the person is alive and therefore might benefit financially from our consumption of their work? Do we vote with our wallets? If so, is it okay to stream, say, a Roman Polanski movie for free? Can we, um, watch it at a friend's house?[5]

Dederer's question (echoed by others[6]), and ambivalent response, reveal the lack of adequate resources to evaluate betrayals that respond to an artist's actions and that may

[2] In the field of Christian ethics, for example, there is a years-long debate about how to receive the legacy of John Howard Yoder. See Rachel Waltner Goossen, "Defanging the Beast: Mennonite Responses to John Howard Yoder's Sexual Abuse," *Mennonite Quarterly Review* 89 (January 2015), pp. 7-80; David Cramer, Jenny Howell, Jonathan Tran, Paul Martens, "Scandalizing John Howard Yoder," *The Other Journal: An Intersection of Theology & Culture*, July 7, 2014. Online at https://theotherjournal.com/2014/07/07/scandalizing-john-howard-yoder/ (accessed 2018-11-8); Stanley Hauerwas, "In Defence of 'Our Respectable Culture': Trying to Make Sense of John Howard Yoder's Sexual Abuse," ABC Religion and Ethics, October 18, 2017. Online at http://www.abc.net.au/religion/articles/2017/10/18/4751367.htm (accessed 2018-11-8); Hilary Scarsella, "Not Making Sense: Why Stanley Hauerwas's Response to Yoder's Sexual Abuse Misses the Mark," ABC Religion and Ethics, Updated 1 Dec 2017 (First posted 2017-11-30). Online at http://www.abc.net.au/religion/articles/2017/11/30/4774014.htm (accessed 2018-11-8). In philosophy, the debate about the reception of Heidegger in light of his anti-Semitism has taken on new intensity with the recent publication of the Black Notebooks. See for example Ingo Farin and Jeff Malpas (eds), *Reading Heidegger's Black Notebooks 1931-194*1 (Cambridge: MIT Press, 2016); Andrew Mitchell and Peter Trawny (eds), *Heidegger's Black Notebooks: Responses to Anti-Semitism* (Columbia University Press, 2017).

[3] Daniel Kolitz, "Academics Explain David Foster Wallace To Me," *The Outline*, July 25, 2018, https://theoutline.com/post/5543/david-foster-wallace-conference-profile?zd=1&zi=hoxjdain (accessed 2018-11-6).

[4] Claire Dederer, "What Do We Do With the Art of Monstrous Men?," *The Paris Review*, November 20, 2017. Online at https://www.theparisreview.org/blog/2017/11/20/art-monstrous-men/ (accessed 2018-11-8).

[5] Dederer, "Monstrous Men."

[6] Sarah Lyall and David Itzkoff, "Charlie Rose, Louis C.K., Kevin Spacey: Rebuked. Now What Do We Do With Their Work?" *The New York Times*, November 24, 2017. Online at https://www.nytimes.com/2017/11/24/arts/charlie-rose-kevin-spacey-louis-ck-art.html (accessed 2018-11-8); Amanda Hess, "How the Myth of the Artistic Genius Excuses the Abuse of Women," *The New York Times*, November 10, 2017. Online at

bear on our aesthetic valuations of the artist's work and that may be experienced as quite intimately personal. What do we do with the sense of *betrayal* that may follow on the discovery of an artist's bad behavior? What are the implications of consuming of such art? What concepts and norms might help to guide reflection? As Kolitz puts it, what do you say (and do) "when you're told, over and over, that the work you love is tainted, and that loving it taints you too?"[7] Underlying these are deep and ambiguous questions about how to construe the relationship between art and life – both the artist's and the viewer-lover's – and the ethics of consuming art in light of those relationships. In other words, there are questions about both the ethical significance of the love for and involvement with artworks and artists, and, more broadly, questions about the ethical consumption of artworks.[8] At stake in both sets of questions are concerns about the ways in which we are formed by the images and works we consume (a concern that is heightened with respect to works we love and hold in high regard), and also how consuming and cherishing particular works contributes to moral formation.[9]

Though the potential scope of this set of issues is large, I will focus on artists who have been accused of sexual harassment and abuse in light of the feminist "reckoning" in the arts spurred by the #MeToo movement. My starting point is the articulated dilemma that these accusations create for those who have highly regarded specific artworks and consequently been formed by them and found themselves "involved" with them insofar as they have been personally significant, which is why my discussion draws significantly on two first-personal accounts by people grappling with how to relate to artworks that have been formative and beloved in their lives. Because the dilemmas arise not simply out of judgments of aesthetic value (though these are at play) but from a kind of involvement with the artwork that entails both aesthetic regard and personally formative significance, I use the term "aesthetic involvement" to denote this relationship.[10] Both of the accounts of

https://www.nytimes.com/2017/11/10/arts/sexual-harassment-art-hollywood.html (accessed 2018-11-8); Svetlana Mintcheva, "Caravaggio Killed a Man. Should We Therefore Censor His Art?" *The Guardian*, February 3, 2018. Online at https://www.theguardian.com/commentisfree/2018/feb/03/caravaggio-killed-a-man-censor-art?CMP=share_btn_fb (accessed 2018-11-8); A.O. Scott, "My Woody Allen Problem," *The New York Times*, January 31 2018. Online at https://www.nytimes.com/2018/01/31/movies/woody-allen.html (accessed 2018-11-8).

[7] Kolitz, "Academics Explain."

[8] Note that some of these questions about consumption or economic and cultural support may apply to other "arts" (scholarship, for example) as well as to other kinds of figures (public intellectuals, politicians, religious leaders, and so forth). I restrict my focus here, however, to the discussion of artists and artworks.

[9] By moral formation, I mean ways in which we are shaped as moral agents: affectively, dispositionally, and intellectually, according to particular values, ideals, and principles, both consciously and unconsciously. An underlying premise of this paper, which I will not defend here, is that this formation operates aesthetically insofar as what we find appealing, attractive, and compelling – as well as what we find repulsive or disruptive – accompanies and influences our valuations of people, things, and actions (and vice versa). This is not to say that aesthetic valuations always track moral ones, or vice versa – but that in the broad and pervasive shaping of an agent, moral and aesthetic formation intersect in important ways.

[10] Strong forms of aesthetic regard for works often compel attention to those works such that they generally entail significant attachments. Certainly, it is possible to make positive judgments of aesthetic value that do not entail such formative importance. And it is also possible to maintain an

aesthetic involvement I consider in the beginning of this paper also tie their individual attachment to certain works and artists to the broader cultural importance these works and their artists enjoy.

While the following discussion touches on longstanding questions about the relationship between art and the artist's biography, the effects of art on the viewer, and the relationship between aesthetics and ethics more broadly, it does so in the context of addressing the specific set of questions around regard for and consumption of artworks that have been raised in the #MeToo reckoning.[11] I will, in what follows, account for how one can feel betrayed by an ethically "tainted" artist and their work, and offer a preliminary sketch of a feminist ethics of aesthetic involvement in response to such betrayals.[12] I explore this first, with respect to individual regard for artworks that are aesthetically valued as well as beloved (individual aesthetic involvement). Second, I explore betrayals with respect to cultural aesthetic involvement: that is, artworks that enjoy a broad cultural regard for their aesthetic value and hold an important place in the "common artistic

attachment to artworks that have been personally significant in some way without considering them to be highly aesthetically valuable in their own right. Art may be formative without being highly regarded aesthetically, insofar as exposure to and consumption of art presumably has some effect on us whether or not we hold it in high regard. "Aesthetic involvement," however, designates a judgment of both aesthetic regard and personal significance, which means that the artwork's formative power at issue (and entails a particular set of dynamics with regards to betrayal).

[11] Regarding the philosophical questions around ethics and aesthetics more broadly, I am particularly indebted to Berys Gaut's account of "ethicism" in *Art, Emotions and Ethics* (Oxford: Oxford University Press, 2007). The literature on the topic is vast; for further recent works, see R.W. Beardsmore, *Art and Morality* (London: Macmillan, 1971); Noël Carroll, "Art and Ethical Criticism: An Overview of Recent Directions of Research," *Ethics* vol. 110/2 (2000): pp. 350-387; Diané Collinson, "'Ethics and Aesthetics Are One,'" *The British Journal of Aesthetics* vol. 25/3 (1985): pp. 255-272; Stephen Davies (ed.), *Art and its Messages: Meaning, Morality, and Society* (University Park: Pennsylvania State University Press, 1997); Mary Devereaux, "Moral Judgments and Works of Art: The Case of Narrative Literature," *Journal of Aesthetics and Art Criticism* vol. 62 (2004): pp. 3-11; Arnold Isenberg, "Ethical and Aesthetic Criticism," in *Aesthetics and the Theory of Criticism: Selected Essays of Arnold Isenberg*, ed. William Callaghan et al. (Chicago: University of Chicago Press, 1973); Daniel Jacobson, "Ethical Criticism and the Vice of Moderation," in Matthew Kieran (ed.), *Contemporary Debates in Aesthetics and the Philosophy of Art* (Oxford: Blackwell, 2006); Matthew Kieran, "Art, Imagination, and the Cultivation of Morals," *Journal of Aesthetics and Art Criticism* vol. 54 (1996): pp. 337-351; Jerrold Levinson (ed.), *Aesthetics and Ethics: Essays at the Intersection* (Cambridge: Cambridge University Press, 1998); Hallvard Lillehammer, "Values of Art and the Ethical Question," *The British Journal of Aesthetics* vol. 48/4 (2008): pp. 376-394; Jennifer A. McMahon, *Art and Ethics in a Material World: Kant's Pragmatist Legacy* (New York: Routledge, 2014); Panos Paris, The 'Moralism' in Immoralism: A Critique of Immoralism in Aesthetics," *The British Journal of Aesthetics* vol. 591/2 (January 2019): pp. 13–33; Richard Posner, "Against Ethical Criticism," *Philosophy and Literature* vol. 21 (1997): pp. 1-27 and "Against Ethical Criticism: Part Two," *Philosophy and Literature* vol. 22 (1998): pp. 394-412; Robert Stecker, "The Interaction of Ethical and Aesthetic Value," *British Journal of Aesthetics* vol. 45 (2005): pp. 138-150.

[12] In this sense, I am trying to wed Dederer's evocation of "ethical thoughts" and "moral feelings," which she distinguishes: "we tell ourselves we're having *ethical thoughts* when really what we're having is *moral feelings*." But ethics must account for moral feelings; here, I offer an ethical framework for moral feelings of betrayal by art.

record," and thus may be said to be culturally formative, if variably formative for individuals.

Art's Formative Qualities

The relationship between a viewer,[13] a work of art, and its maker creates a relational triad that may entail particular forms of intimacy.[14] Certainly, a degree of separability should be maintained between these three "terms" of the triad, not least because there is a historical-cultural contingency to how this triad is construed; where the artist's personal expressiveness and authenticity is prized, this impacts the triad significantly.[15] But the questions about receiving the art of the so-called "monstrous" supposes that in the contemporary context, there are ethically significant links that require attention. The questions regarding the works of "monstrous men" emerge in response to revealed biographical details that create dilemmas about both the consumption of their works as such and the aesthetic regard that we have for them: these questions range from economic support, to cultural influence, to the ethical attitudes manifest in a work, to the impact such works may have on one's own perceptions and attitudes. Elaine Scarry and Alexander Nehamas have both offered compelling descriptions of the formative relationships people may have with art (and beautiful people/things more broadly[16]). Though neither of them delves into the relationship between a viewer, a work of art, and the artist's biography directly, I suggest that if we take seriously their emphasis on art's formative and relational

[13] I use this term for expediency's sake since most of the works under discussion here are visual artworks, but I take the scope of this discussion to extend to non-visual arts as well.

[14] Wayne Booth, for example, uses a friendship model to describe these relations. Wayne Booth, *The Company We Keep: An Ethics of Fiction* (Berkeley and Los Angeles: University of California Press, 1988). Gaut rightly contests this as a general model for the relationship a viewer or reader may have to the "manifest author" but also acknowledges that it may apply in certain cases. See Gaut, *Art*, pp. 109-114.

[15] Certain questions might still apply to a medieval master painting icons of saints, for example, but surely not to the same extent as to artists whose works assert a more directly personal expressivity. On Gaut's model, expressive artistic acts impart a "manifest author" that may and may not be directly continuous with the author's real-life character. Gaut, *Art*, pp. 67-82. As Gaut notes, "people vary in their attitudes in different contexts and also can fashion delusive or fictional personae in their work. However, the view also allows that the artist's personality as manifested outside his work may be relevant, since it is the same person who acts in both contexts. The test must be whether, in the light of one's knowledge of the artist's attitudes outside his work, one can detect in the work traces of these attitudes." Gaut, *Art*, p. 74.

[16] While Nehamas and Scarry are both developing accounts of beauty, neither of their accounts rest on a notion of beauty as a formal aesthetic property. Both of their accounts give a central place to affective and cognitive responses to things we find beautiful (rather than a specific aesthetic property we might call "the beautiful"), impulses toward learning that follow upon this response, and the attachments these responses generate. This makes their accounts amenable to extrapolation beyond responses to beauty alone, but to all responses to what we find aesthetically compelling such that we hold it in high regard. My interest here is not restricted to aesthetic properties or artworks with regards to beauty specifically, but to the formative and relational qualities both Nehamas and Scarry describe, which I take to be relevant for aesthetic responses broadly and not only responses to beauty as such.

qualities, abstracting art from artist from the viewer in this relational triad is not straightforward.

According to Nehamas, we are drawn to things and people we find beautiful in some way because they beckon with a promise of happiness.[17] This promise consists in the idea that by devoting a part of one's life to someone or thing we find attractive and compelling, one will come to know both the object and oneself in new and unforeseeable ways. This is transformative; it actively refashions the lover's desires and standards and understanding.[18] As a result, when one enters a relationship with such people or things, one gives them power over oneself "emotionally, ethically, and intellectually, trusting [them] not to exploit it. By becoming vulnerable in this way, I put my identity at serious risk because I have no way of telling how our relationship will ultimately affect me and whether it will be for good or bad."[19] If love is propelled by this sense transformative promise, as Nehamas claims, this necessarily implies an aesthetic judgment that also shapes the self – it opens onto aesthetic involvement. Indeed, Dederer describes the impact of repeatedly viewing Roman Polanski films in direct consumptive terms: "I ate them. They became part of me, the way something loved does."[20]

A.O. Scott's recent confessional, "My Woody Allen Problem," (*The New York Times*, 2018-1-31) illustrates this intimately formative dynamic in relation to one of the disputed figures of the #MeToo reckoning. Woody Allen has been accused of sexually assaulting his adoptive daughter, Dylan Farrow, in what has been and continues to be a highly contentious and highly mediatized conflict.[21] A number of actors and critics have weighed in on the dispute and on the attitude viewers should have towards Allen. In this first section, I focus on Dederer's and Scott's accounts of their reckoning with Allen's work, a formative figure for both of them. With regards to the accusations, Scott declares his assessment from the start: he thinks Allen is guilty. In response to an earlier scandal, Allen's affair with and subsequent marriage to Soon-Yi Previn (his long-term partner Mia Farrow's adopted daughter), Scott notes that he had then argued that Allen's personal life should be distinguished from the work. This contrasts with Dederer's account, in which she describes the Soon-Yi Previn revelation as the moment of "terrible betrayal".

[17] Alexander Nehamas, *Only a Promise of Happiness: The Place of Beauty in a World of Art* (Princeton: Princeton University Press, 2007). Nehamas construes beauty's appeal broadly; he suggests we may be attracted (for example) to people or works that are conventionally or outwardly ugly, just as we may be repulsed by people or works that are conventionally or outwardly beautiful. In this sense, the propulsion is attraction to and love for an object, which necessarily entails finding it beautiful in some way. This emphasis on the relational qualities of attraction and love makes his account particularly relevant to aesthetic involvement broadly.

[18] Nehamas, *Promise*, pp. 57-63.

[19] Nehamas, *Promise*, p. 57.

[20] Dederer, "Monstrous Men."

[21] The search results for recent coverage in the New York Times alone yield dozens of articles. A number of these cover the flurry of accusations and responses back and forth, primarily between Ronan and Dylan Farrow on the one hand, and on the other, Moses Farrow and Soon-Yi Previn (Allen's wife and Mia Farrow's adopted daughter, who was a young girl when Allen and Farrow began their relationship). For a recap and timeline of events, see Sopan Deb and Deborah Leiderman, "Woody Allen, Mia Farrow, Soon-Yi Previn, Dylan Farrow: A Timeline," *The New York Times*, January 31, 2018, https://www.nytimes.com/2018/01/31/movies/woody-allen-mia-farrow-dylan-farrow-a-timeline.html (accessed 2018-11-2). A further number address the debate over the status of the accusations, how they should be received, and Woody Allen's artistic merits.

The position to which Scott previously adhered, that the artist's personal life should be distinguished from the work, is the standard objection to the art/biography/viewer triad as ethically significant.[22] This is more difficult to maintain with regards to an artist like Allen whose work is particularly personal: "saturated with his personality, his preoccupations, his biography and his tastes."[23] More importantly, Scott describes how powerful the image of Woody Allen as a male lead – as a sex symbol, or at least as the persistent object of women's sexual and romantic attentions – was for him as an adolescent:

> The man himself was a plausible definition of sexy. The achievement of his early movies, culminating in "Annie Hall" (his seventh feature as a director) was to turn a scrawny, bookish, self-conscious nebbish into a player. … The aspects of his temperament held up for mockery — the hyper-intellectualism, the snobbery, the irreducible Jewishness — doubled as weapons of seduction. His self-deprecation was a tactic, a feint, a rope-a-dope, and he was plagued less by the frustration of his desires than by their fulfillment. As soon as the heart got what it wanted, it wanted something else. What impressionable, heterosexual, unathletic adolescent boy would not want a piece of that action?[24]

Interestingly, Dederer too describes a strong identification with Allen – "I *felt like* Woody Allen. …I felt closer to him than seems reasonable for a little girl to feel about a grown-up male filmmaker."[25] This capacity to produce identification in his audience is an aspect of his particular genius, she suggests – and perhaps precisely artists with this particular talent are the ones who will elicit the strongest sense of betrayal.

In his recent re-assessment of Allen's work, Scott bluntly rejects the separation of art from artist: "the notion that art belongs to a zone of human experience somehow distinct from other human experiences is both conceptually incoherent and intellectually crippling. Art belongs to life, and anyone — critic, creator or fan — who has devoted his or her life to art knows as much."[26] Allen's persona is intimately woven into his work, which makes the aesthetic triad particularly obvious. And because of the ways in which Scott values them, the works are intimately woven into his own life. He describes watching Allen's movies – all of them, some repeatedly, from a young age:

> I could, I suppose, declare that I won't watch any more of his movies. But I can hardly unwatch the ones I've seen, which is all of them, at least half more than once. And even if I could, by some feat of cinephilic sophistry, separate those movies from Mr. Allen's life, I can't possibly separate them from mine.[27]

Scott's attraction to a figure unconventionally attractive to women in whom he saw himself held a kind of transformative appeal ("What impressionable, heterosexual, unathletic

[22] This is one feature of what Gaut terms the "autonomist" stance on the relationship between ethics and aesthetics. For the full account, see Gaut, *Art,* pp. 67-89.
[23] Scott, "My Woody Allen Problem." Allen, who both directs and stars in many of his films, is a particularly clear case, but as Gaut shows, any continuity between the real-life author and the manifest author may be significant in this sense.
[24] Scott, "My Woody Allen Problem."
[25] Dederer, "Monstrous Men."
[26] Scott, "My Woody Allen Problem."
[27] Scott, "My Woody Allen Problem."

adolescent boy would not want a piece of that action?"). Woody Allen made the nerd into a player; and not just a player, but one who liked and attracted young – sometimes very young – women, whom he would (try to) educate, and whose vulnerability and relative ignorance was often portrayed as part of their (often transient) charm.[28] Scott notes that for all of Woody Allen's self-deprecation, his character exhibits:

> …a powerful sense of entitlement. The Woody Allen figure in a Woody Allen movie is almost always in transit from one woman to another, impelled by a dialectic of enchantment, disappointment and reawakened desire. The rejected women appear shrewish, needy, shallow or boring. Their replacements, at least temporarily, are earnest, sensuous, generous and, more often than not, younger and less worldly than their predecessors. For a very long time, this was taken not as a self-serving fantasy but as a token of honesty, or freedom from sentimental conceptions of domestic love. … What I find most ethically troubling about Mr. Allen's work at present is the extent to which I and so many of my colleagues have ignored or minimized its uglier aspects. A sensibility that seemed sweet, skeptical and self-scrutinizing may have been cruel, cynical and self-justifying all along.[29]

This epitomizes the way in which, as Berys Gaut notes, after learning about the biographical facts (or learning more about the biographical facts), one may "see attitudes in the [artworks] with greater focus and clarity than she previously had done."[30] What is particularly relevant for ethical criticism of artworks is whether an artist's "reprehensible attitudes infect their work."[31] This is a matter not just of individual concern insofar as Allen's works were personally formative for Scott, but because Allen's works are "part of the common artistic record, which is another way of saying that they inform the memories and experiences of a great many people. I don't mean this as a defense, but an acknowledgment of *betrayal and shame*."[32] When aesthetic involvement concerns a body of work, as it does in this case – particularly one that is not only significant to the individual but has also had an important place in what Scott calls the "common artistic record" – the relational triad between viewer, art, and artist is further heightened, as is the sense of betrayal that may then follow.[33]

[28] *Manhattan* is undoubtedly the prime offender in this regard, the one eliciting the most re-evaluation. In addition to Dederer and Scott, see Steve Kurutz, "How Do You Solve a Problem Like Manhattan?" *The New York Times*, March 1 2018, https://nyti.ms/2FHasqw (accessed 2018-11-2).
[29] Scott, "My Woody Allen Problem."
[30] Gaut, *Art*, p. 74.
[31] Gaut, *Art*, p. 77.
[32] Scott, "My Woody Allen Problem," emphasis mine.
[33] While the starting point for this essay are individual accounts of feelings of betrayal relative to personally formative artworks and their artist, the individual dilemma is not strictly separable from a collective discussion about artworks and artists who have a place in the "common artistic record," and therefore have a formative role in the broader culture – even if they will not have the same value for every individual. For this reason, individuals who may not have felt themselves to have been specifically or especially formed by a high personal regard for Woody Allen (for example) may still be invested (if not to the same degree) in assessing his works and the ways they choose to relate to them.

Error, Betrayal, and Reassessment

Nehamas briefly addresses a kind of betrayal that follows on disillusionment with the beloved or admired object, drawing on Proust's character Swann from *In Search of Lost Time*:

> You can't even be certain that you will eventually consider what you find through the pursuit of beauty to have been worth your while. Perhaps you will feel about it as Swann came to feel about Odette after all the years he devoted to her: "To think that I have wasted years of my life, that I wanted to die, that I felt my deepest love, for a woman who did not appeal to me, who was not even my type!"[34]

For Nehamas, such disappointment is the risk of love: one pursues the beloved not knowing where it will lead, and it may lead down paths one eventually judges shameful. The betrayal is significant not only because the object no longer appears beautiful or compelling as it had previously, but because when one has made oneself vulnerable to transformation by a beloved and admired object, has given it a place in one's life and identity, the implications are not only evaluative (a reconsidered appraisal of its aesthetic value) but personal.

Elaine Scarry offers a more detailed account of how one may feel betrayed by an object or person one judged beautiful (or, I would say more broadly, held in high aesthetic regard). She describes a feeling of betrayal that results from errors of over-attribution: the feeling one has when a person or object one had regarded as beautiful (or aesthetically compelling) no longer appears so. Errors of over-attribution have "the peculiarity that when the person or thing ceases to appear beautiful, it often incites the perceiver to repudiate, scorn, or even denounce the object as an invalid candidate or carrier of beauty."[35] The object may be "turned upon, as though it has enacted a betrayal."[36] In these passages, Scarry is describing the "temptation to scorn the innocent object for ceasing to be beautiful,"[37] but the dynamics are applicable beyond judgments of beauty as such to aesthetic involvements more broadly, just as the dynamics Nehamas describes apply to compelling, beloved objects broadly, since he thinks attraction and love involve a judgment of beauty in some sense.

Scarry clearly deems the repudiation of the reconsidered object excessive, in the case of an otherwise "innocent" object – but what if the beloved, admired object was made by one deemed criminal or abusive? What is the nature of the betrayal and what kind of repudiation and denunciation is appropriate? Scarry's discussion of aesthetic regard, error, and betrayal provides some preliminary pathways for reflection, but in the context of the current debate, her distinction between errors of over- and under-attribution of beauty is too limited a model. Dederer's question "what do we do with the art of monstrous men?" articulates a dilemma about how an artist's ethical conduct bears on one's aesthetic regard for and consumption of their works, but the sense of betrayal arises specifically when one is aesthetically involved with the work.[38]

[34] Nehamas, *Promise*, p. 130.

[35] Scarry, *On Beauty and Being Just* (Princeton: Princeton University Press, 2001), p. 49.

[36] Scarry, *On Beauty*, p. 51.

[37] Scarry, *On Beauty*, p. 49.

[38] I note that the ethical dimensions of attributing aesthetic value are various and complex; a comprehensive treatment of this topic far exceeds the scope of this paper.

One response, as Scarry details, is to revise one's regard for a work or an artist as an error of over-attribution: one considered an artist's work compelling, beautiful, or important, but in fact it is not. Insofar as an artist's ethical merits (or defects) may be expressed in the work in aesthetically relevant ways, such reconsiderations may indeed be pertinent (as in Gaut's claim above that biographical details may allow for a more focused and clear assessment of the works).[39] But Dederer's and Scott's reflections show that revisiting the quality of one's aesthetic involvement is more complicated than mere revisions of aesthetic judgment. As Scott notes, he cannot "unwatch" Allen's movies. Though he recognizes that not watching Allen's movies is a legitimate response, he suggests that rather than attempting the impossible task of expunging Allen's impact on him, he may need to start all over again. I take him to mean that given Allen's imprint on him, and his sense of the possibly ugly and de-formative effects on him, a better response may be to re-watch his films with a new critical perspective. Such reassessment involves more reflexivity than Scarry describes in her description of responding to errors of over-attribution.

Attributive errors may be more variable than simple errors of over- or under-attribution. We might better name them errors of *mis*attribution. Revising such errors may imply a steadfast attribution of aesthetic value in some respects while drastically revising that attribution in others. For example, one might still regard Allen's *Manhattan* as a cinematic landmark and admire its opening montage (as Dederer describes[40]), while unfavorably reassessing its uncritical portrait of the romantic relationship between a 42-year-old Isaac (played by Allen) and a 17-year-old Tracy (Mariel Hemingway), as well as the unflattering gap between straightforward, sex-loving Tracy and the "brittle" anxious full-grown women in the movie.[41] Or, to take a different and more egregious example, one might still on some level positively acknowledge the raw intensity of Marlon Brando's and Marie Schneider's performances in Bertolucci's *Last Tango in Paris* while reviling the deception that underlies the anal rape scene, which according to Schneider had not been included in the original script, and according to Bertolucci had been in the script but without the detail that butter would be used as a lubricant. Bertolucci admits that he and Brando deliberately concealed the detail about using butter from Schneider prior to filming the scene expressly for the purpose of eliciting a spontaneous response from her on film.[42] Once informed by its staging, we see in the scene a continuity between the fictional violation depicted and the actual violation of the actress's autonomy by withholding information about the acts, involving her body, that would be depicted in the scene – and, further, a continuity between a deeply problematic attitude towards the woman depicted in the film (manifest in the rape) and a deeply problematic attitude toward the actress playing her part (manifest in the deliberate withholding of critical information regarding the depiction of acts to which her body would be subject).

[39] As Gaut's model for "ethicism" suggests, the claim that certain ethical virtues or insights are aesthetically relevant does not entail the claim that all ethical virtues are aesthetically relevant nor that ethical considerations are exclusively or overridingly important for the aesthetic assessment of the work. See in particular Gaut, *Art*, pp. 57-66.

[40] Dederer, "Monstrous Men."

[41] Dederer, "Monstrous Men."

[42] Elahe Izadi, "Why the 'Last Tango in Paris' rape scene is generating such an outcry now," *The Washington Post*, December 5 2016, https://wapo.st/2g0vcLt?tid=ss_mail&utm_term=.1b18e8f8fcec (accessed 2018-11-2).

But such reassessments of works in light of relevant biographical details regarding the real-life artist do not only regard the works as such (as Gaut describes), they may also be deeply personal. The sense of betrayal reflects not simply an incomplete or misconstrued interpretation of a person or object's value but its intimately formative effects on a person in terms of aesthetic involvement, and, finally, the impact of this valuation and one's involvement with a work on those who may have suffered disregard at the artist's hands. Scott does not enter into extended detail of how he assesses Allen's impact on him, but he notes that he found Allen's status as a bookish sex symbol aspirational ("a masculine ideal"). Evidently, Allen shaped his sense of male and female desire, desirability, and of romantic relational dynamics. More broadly, the Woody Allen figure exhibits a sense of sexual entitlement while at the same time portraying the female rejects and replacements in ways that suit the character's wants – to reject, to replace. Dederer, for her part, connects Allen's lack of "moral shading" on "middle-aged men fucking teenage girls" to a "moral disregard for anything but the self" encapsulated in his infamous response to a question about Soon-Yi Previn – "the heart wants what it wants."[43]

The point I want to emphasize here is not about whether Scott's reassessment of Allen following the renewed attention to the allegations of molesting Dylan Farrow is definitive in some generalizable way. Critical evaluations are always subject to contestation and revision. The point I want to emphasize is that Scott expresses his reassessment of Allen's works, and the sense of betrayal it entails, as also *self-indicting*. Scott thus makes the move Dederer sees as critically and uncharacteristically absent in Allen's work on this point: "Woody Allen's usual genius is one of self-indictment, and here is his one film where that self-indictment falters…" But Dederer's point is broader; she questions the self-serving nature of the failure in critical reception: "…and also he fucks a teenager, and *that's* the film that gets called a masterpiece? What exactly are these guys defending? Is it the film? Or something else?"[44] Allen's failure of self-indictment mirrors the viewer's defense of Allen's artistry (which, Dederer implies, is itself a failure of self-indictment or at least self-examination). Scott, too, implies concern over the de-formative effects of Allen's place in the "common artistic record" inscribed by the critical adulation Allen received while critics ignored or minimized the "uglier aspects" of the films' portrayal of women and male-female relationships. Scott makes a move from reflection on his individual formation to the collective implications of minimizing – or even adulating – certain works' de-formative elements.

A similar concern has been expressed regarding Louis C.K.'s comedy works – that what was often received as self-parodying may in fact have been self-serving.[45] C.K. has admitted to the repeated sexual harassment of female colleagues by (non-consensually)

[43] Dederer, "Monstrous Men."

[44] Dederer, "Monstrous Men."

[45] "The stories told by the women raise sharp questions about the anecdotes that Louis C.K. tells in his own comedy. He rose to fame in part by appearing to be candid about his flaws and sexual hang-ups, discussing and miming masturbation extensively in his act — an exaggerated riff that some of the women feel may have served as a cover for real misconduct." Melena Ryzik, Cara Buckley, and Jodi Kantor, "Louis C.K. Is Accused by 5 Women of Sexual Misconduct," *The New York Times*, November 9, 2017, https://nyti.ms/2ho0aE0 (accessed 2018-11-6). See also Joe Berkowitz, "In A Post-Weinstein World, Louis CK's Movie Is a Total Disaster," *Fast Company*, October 20 2017, https://www.fastcompany.com/40483261/in-a-post-weinstein-world-louis-cks-movie-is-a-total-disaster (accessed 2018-11-6).

masturbating in their presence. His comedy work includes extended discussion and miming of masturbation and his 2017 film, *I Love You, Daddy*, includes a male character who pretends to masturbate in front of a female character. C.K.'s character is also depicted dismissing sexual abuse allegations against an Allen-esque character. Precisely in translating personal wrongdoing into comedic content, does C.K. minimize attitudes and acts that are both sexually and professionally degrading? Comedy, of course, can undermine the objects, acts, or attitudes it depicts by satirizing their flaws. This can be a powerful form of critique. But it can also normalize its objects through levity and trivialization.[46] Certainly once one knows about Louis C.K.'s harassment, it is difficult to disassociate his masturbation bits from his actions towards his female colleagues and the impact on them. Reassessing his work, and one's own aesthetic involvement with it, in light of his admissions of sexual harassment may elicit feelings of shame and betrayal.

Reevaluating the art of those who are discovered to have committed morally reprehensible actions is particularly difficult when one has been shaped by that work, because it implies an evaluation not only of one's aesthetic regard for it but of the ways in which one has been shaped by it (one's aesthetic involvement with the work and/or artist). The sense of betrayal is visceral because the object's promise does not merely cease: if that object has taken its place in one's life, its perceived promise has shaped that life in ways one did not expect and is unhappy to discover. This bears on one's possible implication in the "uglier aspects" of such works. The betrayal regards the unforeseen or undiagnosed formative implications of aesthetic involvement. This betrayal requires a reconsideration not only of the object itself but of its place in one's life and its impact on one's self.

The Problem with Monsters

Dederer's evocation of "monstrous men" is problematic for the reflexive dimension of reassessment on a few levels. First, because not only men commit acts of sexual harassment and assault (though the disparately gendered nature of sexual harassment and assault should not be occluded). Second, because calling them these men "monstrous" risks simply setting them apart – cloistering them as exceptionally vicious. As Kate Manne notes in her recent book on the logic of misogyny, demonizing rapists obscures the prevalence of sexual assault. Monsters "are unintelligible, uncanny, and they are outwardly frightening." But rapists "are human, all too human, and they are very much among us. The idea of rapists as monsters exonerates by caricature."[47] As Manne argues, acknowledging and working to address the pervasiveness of sexual harassment and assault requires acknowledging the "*banality* of misogyny."[48] But to take the point further, rendering sexually deformed patterns of interaction monstrous in the exceptional sense evades a reflexive examination of one's own implication in misogyny more broadly. Framing the question in terms of

[46] "It is often through humor (say through irony or satire) that people can keep making sexist and racist utterances." Sarah Ahmed, *Living a Feminist Life* (Durham: Duke University Press, 2017), p. 261.

[47] Kate Manne, *Down Girl: The Logic of Misogyny* (Oxford: Oxford University Press, 2018), p. 198.

[48] Manne, *Down Girl*, p. 211.

"monstrous men" risks precluding the reflexive and collective reassessment that Scott, for example, describes. And Dederer is clearly attuned to this danger:

> Something in us—in me—chimes to that awfulness, recognizes it in myself, is horrified by that recognition, and then thrills to the drama of loudly denouncing the monster in question. The psychic theater of the public condemnation of monsters can be seen as a kind of elaborate misdirection: nothing to see here. I'm no monster. Meanwhile, hey, you might want to take a closer look at *that* guy over there.[49]

But Dederer's response to this problem opens onto a third problem: that she then conflates monstrosity with the artistic impulse as such. She identifies monstrosity in the act and conditions of writing itself. There's a selfishness that creativity requires, she notes – there are clear echoes here of her accusation of Allen's "monstrous disregard" for others. Dederer worries that she's not quite monstrous *enough* for great art – but she knows she's monster enough to finish.[50] The heart wants what it wants and the artist does what it takes to finish their work. But with this move, Dederer asserts the wrong kind of banality. This banality erases the very moral shading whose absence she noted in Allen, though from a different vantage. Dederer identifies the statutory rape of an adolescent girl in *Manhattan* as monstrous in the attitude of normalizing nonchalance towards that act; in her own shift to reflexivity, the everyday judgments a working artist parent/spouse makes about the use of their time become monstrous. In the first instance, the monstrous becomes banal in a way that effaces its monstrosity. In the second instance, the banal becomes monstrous with similar consequences – but only because Dederer targets the wrong object for critique. Had she focused (as she does elsewhere) on the ways in which Allen's latent misogyny is not just a feature of his particular moral failings but represents and reinforces our moral failings both as viewers or appreciators of his work but also as a collective in which he has occupied a place of cultural authority that rests on aesthetic regard for and involvement with his work, the connection between monstrous banality and banal monstrosity would have been clearer. Allen may be exceptional in certain regards, but in other regards he is not and neither are the male-female relations he portrays. This is the point that needs making, and that Dederer misses by shifting monstrosity to artistic ambition. I am sympathetic to the reflexivity Dederer displays but worry that the monster language either cloisters the "bad guys" as exceptional cases or makes monsters of us all in ways that ultimately preclude meaningful critical assessment in both cases.

Naming the process Dederer and Scott are engaged in as discerning the contours of possible ethical *deformation* in their practices of aesthetic involvement is a better alternative. Deformation admits of a spectrum of gravity and specificity without losing the importance of banality in the sense that we are all imperfectly formed, subject to distortions great and small. If aesthetic involvement is or may be in certain respects ethically formative, then the problem Dederer and Scott are describing has to do with ethically-deformed artists/artworks and the aesthetic involvement viewers develop with them. If aesthetic involvement is formative, deformation refers to the ways in which aesthetic formation may be for the worse: how it may misshape a person's affections, sensibilities, dispositions, their sense of what is attractive and repulsive, their sense of what is right or

[49] Dederer, "Monstrous Men."

[50] "Finishers are always monsters." Dederer, "Monstrous Men."

good or at the very least, permissible, excusable. Deformation cannot only be identified in the artist and/or their works, but in the perceiver who (and collectively, the culture that) bestows regard upon them. Though Scarry and Nehamas address the possibility of error, betrayal, and deformation (being "shaped for the worse" as Nehamas says) to some extent, neither really addresses the problem of one's possible implication in ethically problematic social and cultural structures of aesthetic involvement. They do not address how to respond to aesthetic betrayals both personal and social beyond simply repudiating the object, which risks eliding the reflexive examination of how the betrayal may reveal one's implication in possibly distorting forms and structures of aesthetic involvement. The formative sources that shape aesthetic involvement can be implicated in distortions in one's view of the world or of relationships with others: our aesthetic involvement may thus render one complicit in perpetuating such views or in the consequences they have in relationships with others. Returning to the Louis C.K. example, continuing to treat his jokes with levity risks trivializing his (and similar) actions, with real consequences for women who have such experiences. Of course, one risks being labelled a killjoy for refusing to laugh at such jokes – but that, in fact, is precisely the point, and one of comedy's potentially uglier aspects.[51]

An Interrogative Feminist Ethics of Aesthetic Involvement

Once one knows about an artist's reprehensible actions or ideologies, how one evaluates the ethical stakes of engaging that art/ist requires looking both at how an artist's vices may infect (or not) the work, their relationships within the artistic community, and those who spend time, attention, acclaim, and money within that web of relationships. One can understand the deformation that creates a sense of betrayal only by looking at how viewers assess the aesthetic involvement they may have had with artists and/or their works and its formative implications.

To be clear, I take it that when it comes to the wrongdoing the artist commits directly against specific persons (people harassed, molested, assaulted, or otherwise degraded, manipulated, or harmed), the artist is fully responsible and culpable for his or her actions toward these people and is accountable to the legal and moral mechanisms that pertain. Of course, a strong collective aesthetic involvement with an artist and/or their works creates a position of power that can undermine precisely the enforcement of these mechanisms for accountability – this is no small difficulty, and points to the ways in which individual and collective responsibility are connected. But nevertheless, understanding a culpable artist's relationship to their art and to the viewer of that art is more ambiguous. The question "what do we do with the accused artist's art" requires us to ask how we evaluate art and our ethical responsibility in consuming it in light of an artist's actions, how we evaluate the art/ist's formative effects on its viewers/us, and how we evaluate the possible effects of our engagement and appreciation on others. Understanding the ethical stakes requires all three of these dimensions. I note that thus far, I have not suggested what accusations against an artist imply for the viewer's actions; I have not suggested that one should necessarily stop watching Louis C.K. or *Manhattan* or *Last Tango in Paris*. Indeed, part of my aim in this paper is precisely to refuse a linear, self-evident, or categorical path

[51] Ahmed, *Feminist Life*, pp. 261-2.

from sexual harassment or abuse (allegations or convictions) to either boycotting all works associated with allegations or disassociating the art from the artists on principle. Rather, I want to suggest that discerning responses to the many and various cases that may be implicated by allegations of moral reprehensibility (sexual and otherwise) will not be resolved by appeal to a single normative principle, or as Gaut puts it, "a general theoretical edict."[52] Responses can only be discerned by considering a range of questions about both the work at hand and the relationships involved. I note that the relevant questions here concern both aesthetic assessment of the work and one's aesthetic involvement with it as well broader questions about consumption of that work as such:

1) *Regarding an artist's actions*: What is the nature of the artist's actions? Have the artist's actions or has the artist's character deformed the work? How? How have they damaged or undermined the artistic community of which they are a part?
2) *Regarding the art/ist's relationship to or effects on viewers*: What mark has the artist's work left on me or on others close to me? How does knowing about the artist's actions affect my assessment of the work and/or my willingness to engage the work? What mark has this art/ist left on the "common artistic record"? How does that mark inform memory and experience? What does it lift up, what does it critique, what does it leave unsaid or undisturbed?
3) *Regarding the effects of engagement or consumption on those who have been harmed*: How might my engagement affect those who have suffered or might suffer from the artist's actions? How might my engagement continue to support (or not) an artist's position of power and influence in ways that might reinforce impunity?
4) *Regarding questions to address if/when one engages the art/ist*: Whose voices do I need to listen to before or as I continue to engage this work? How should I, if I do, work with this artist's work? How should the artists' actions influence the substance and focus of my engagement?

The argument for or against continuing to engage and how to engage an artist's work (if one so chooses) develops through such a process: an interrogative approach to the feminist ethics of aesthetic involvement and consumption. Any one person's judgment on these matters may subsequently be subject to contestation by others or to reassessment on one's own terms. Such a process will yield varying arguments; the arguments will be as variable as the responses to the questions themselves, for an art/ist's impact on one viewer or another will not be straightforward, predictable, or uniform. Responses to and interpretations of specific works of art and the artists who make them vary widely, and where one sees an offense, another sees a critique of precisely what the other takes as offensive. This range of interpretation is essential. Answerability for one's position and contestability of that position work together. For this reason, I take the process of discernment I am sketching as first of all a personal and contextual one, one that individuals undertake, that allow them then to make judgments that can be articulated to others. Dederer, too, is critical of the "we" – she rejects the false, imposing authority it implies. The real question, she suggests, is "can *I* love the art but hate the artist?"[53] Part of

[52] Gaut, *Art*, p. 74.

[53] Dederer, "Monstrous Men." But the "we" in Dederer's title is nevertheless telling, because these debates have issued in questions about whether the response to an artist's wrongdoing should

my argument for the approach I am proposing rests on the claim that in most cases, "we" should not restrict or condemn ongoing engagement with an artist's work as such (though we certainly might criticize ongoing engagement that fails to account for the kinds of considerations I have just outlined). Here, I focus on judgments that one must make in contexts where few works are inaccessible as such and debate about the ethics of viewership, consumption, and aesthetic involvement is active.

Before proceeding, let me note that there are a number of arguments to be made for not engaging an artist's work without fully responding to the set of questions I have detailed: one is not invested in the works at issue[54]; one finds the prospect of engaging traumatic or triggering; one prefers to bestow one's attention on other artists, particularly artists who may have been sidelined or who have suffered as a result of this artist's actions. I take it that there is a moral responsibility of "lateral regard" that stands regardless of the decision to engage a reprehensible artist's work. Lateral regard is a term I'm borrowing from Scarry but applying differently in light of concerns about male and racial domination in particular fields. Lateral regard entails expanding the range of one's aesthetic attention – the feminist twist is to specify that there is a particular obligation to expand this regard to the often underappreciated work of women and minorities. This responsibility inheres in a feminist ethics of aesthetic involvement whether or not one chooses to continue to engage with work by dominant (and dominantly male) artists accused of sexual harassment or abuse, though such responsibility is heightened if one chooses to do so. There are two reasons for this. First, lateral regard acts as a counterweight to the ways in which continuing to engage the works of certain artists – even critically – continues to enforce their importance in a kind of cultural canon, or collective artistic record as Scott puts it. Second, lateral regard contributes to honing one's critical perspective, which requires serious and sustained attention to the voices of those most negatively affected by structures of domination. Lateral regard is fundamental to diversifying the voices and works that receive attention, acclaim, and engagement where imbalances persist.

require a more collective response about removing an artist's work from a public or private space: should "we" make decisions about the accessibility of an artist's work that will override to an extent an individual's judgments about engagement? Think of Netflix's removal of Louis C.K.'s work from their platform or post-WWII German prohibitions on screening Leni Riefenstahl's *Triumph of the Will*, the cinematographically acclaimed and equally reviled documentary film chronicling the 1934 Nazi party congress. How are such collective judgments reached and enforced? How might they be contested? Questions around collective judgments on prohibiting access to certain works exceed the scope of this paper; generally speaking, I take censorship to be a very weighty act requiring a heavy burden of argumentation.

[54] The refusal to read David Foster Wallace, for example, crops up in more than one recent piece. See Amy Hungerford, "On Not Reading," *The Chronicle of Higher Education*, September 11 2016, https://www.chronicle.com/article/On-Refusing-to-Read/237717 (accessed 2018-11-6); Deirdre Coyle, "Men Recommend David Foster Wallace to Me," *Electric Lit*, April 17, 2017, https://electricliterature.com/men-recommend-david-foster-wallace-to-me-7889a9dc6f03 (accessed 2018-11-6). While I fall into the non-reading group on this particular author, I admire and respect Clare Hayes-Brady's articulation of how she approaches reading Foster Wallace in light of his domestic abuse and overt misogyny. See "David Foster Wallace in the #MeToo Era: A Conversation with Clare Hayes-Brady," *Los Angeles Review of Books*, September 10, 2018, https://lareviewofbooks.org/article/david-foster-wallace-in-the-metoo-era-a-conversation-with-clare-hayes-brady/ (accessed 2018-11-8).

Returning to the prior point: there are good reasons simply not to engage an artist's work without going to extensive justificatory lengths. Critical and diagnostic work is essential; but it is not the only work to be done, and not everyone needs to take it on. But there are a number of reasons or cases in which one will need to weigh the questions and concerns I have detailed: when one has a relationship to the art/ist in question, whether at a deep personal level, a professional level, or because one cares about the social and cultural concerns at issue; where one feels that certain (potentially) de-formative dimensions of the work require diagnosis, whether personally or collectively. The impetus for A.O. Scott to invest time in revisiting the films that shaped him with a critical perspective is distinct from a casual viewer or the uninitiated. One will have to weigh these considerations in light of the question of lateral regard (attending to the work of other artists) and economic/cultural support (particularly in the case of living artists). If one chooses to engage, one has a responsibility to consider how one's engagement might impact others negatively, and to consider how one might precisely through one's engagement address the de-formative qualities of the art/ist. At a minimum, one's engagement should account for and acknowledge possible tensions and difficulties in the work; the recently reconsidered approach to an exhibition on Casanova (previously titled "The Seduction of Europe") is an example of how #MeToo prompted a more critical appraisal of how to present certain works.[55] To go a step further, one might ask whether one's engagement might precisely seek to form readers, listeners, viewers in ways that create better kinds of attentiveness, that move towards some form of redress?

For illustrative purposes, I will run through how I might apply the questions to a couple of examples. I note that these reflect my own contestable responses. Let's begin with Allen, since he is a recurrent example in the accounts I have discussed. First, regarding the nature of his actions: I take the molestation allegations to be the most serious ones. Evaluating them in light of the investigation proceedings in 1992 and 1993 is not straightforward, as the proceedings are muddled by accusations on both sides: that Mia Farrow coached Dylan, seeking revenge for Allen's relationship with Soon-Yi; that the Yale-New Haven hospital report clearing Allen of abuse charges was seriously flawed. More recently, however, Dylan Farrow's own testimony made public in 2014 via the New York Times carries a great deal of weight.[56] The ensuing conflicting public responses by various family members should not obscure Farrow's voice. The difficulty in many sexual harassment and assault cases is that testimonies are often all one has to judge. Without drawing a clear judgment on Allen's guilt or innocence per se, I find it impossible to disassociate the accusations against Allen from my response to his work.

Second, regarding the relationship to and effects on viewers: as Dederer and Scott both describe, I cannot dissociate the "manifest Woody," whose relationships to and views of girls and women are ethically problematic, from the real-life accusations against him. I

[55] Kimberly Chrisman-Campbell, "When Casanova Met #MeToo," *The Boston Globe*, March 3, 2018, https://www.bostonglobe.com/ideas/2018/03/03/when-casanova-met-metoo/ALNHZdUQJJZ41KKylv8hqM/story.html (accessed 2018-11-6); Cynthia Durcanin, "Casanova as Case Study: How Should Art Museums Present Problematic Aspects of the Past ?," *Art News*, August 13, 2018, http://www.artnews.com/2018/08/13/casanova-case-study-art-museums-present-problematic-aspects-past/ (accessed 2018-11-6).

[56] Dylan Farrow, "An Open Letter from Dylan Farrow," *The New York Times*, February 1, 2014, https://kristof.blogs.nytimes.com/2014/02/01/an-open-letter-from-dylan-farrow/ (accessed 2018-11-7).

note that this is a judgment about Allen specifically, and not a generalizable point – discontinuities are certainly possible between the manifest artist and the real-life artist, between work and biography. Part of what gives rise to the aesthetic questions regarding Allen's work (and not just questions about the ethics of consumption) is that the question of continuity is salient. Further, Allen has been both critically acclaimed and successful with popular audiences, which has made him an influential cultural icon. That is, as an artist he has been highly regarded by many individuals and has received a significant level of cultural regard. As such, his characterizations of male-female relationships and attitudes toward women are well-positioned to influence perception: Dederer and Scott both describe the impact on themselves and beyond, in their social circles and the conversations they have with others. They both describe the impulse to ignore the "uglier aspects" of these and/or to defend and glamourize Allen as a "masculine ideal" and (at least some of) his films as aesthetic and cinematic masterpieces. This entails not only a critical failure to engage the work's uglier aspects but, further, contributes to normalizing those very aspects of male-female interactions and attitudes toward women.

Third, regarding the effects of consuming Allen's works on those who have been harmed, Dylan Farrow's closing appeal in her open letter carries quite a bit of force on this front: "So imagine your seven-year-old daughter being led into an attic by Woody Allen. Imagine she spends a lifetime stricken with nausea at the mention of his name. Imagine a world that celebrates her tormenter. Are you imagining that? Now, what's your favorite Woody Allen movie?"[57] Continuing to engage Allen uncritically reinforces Allen's position of cultural influence in ways that Farrow testifies as profoundly hurtful. In light of the consideration on the second set of questions, uncritical reinforcement of his status as a cultural and cinematic icon (a form of aesthetic involvement) also risks continuing to ignore the problematic aspects of his portrayals of women and relationships – and, more broadly, similar portrayals or attitudes more broadly beyond Allen's œuvre alone (for if we defend Allen on this front, we will probably also fail to examine similar attitudes and portrayals in other art whether or not the artists in question are in any way personally culpable for reprehensible acts). Beyond questions about engagement on a cultural level there is the issue of economic support entailed by consumption, which is difficult to extract from the question of reinforcing aesthetic involvement in terms of his cultural status. On this point, the third and fourth sets of questions dovetail. I take any engagement of Allen as an artist and his works to imply a critical burden. That is to say, engagement of his films should take seriously the need to critically examine his portrayals of girls and women, male-female patterns of interaction, and male-female desires and desirability. This kind of assessment of his works is both ethically and aesthetically relevant. But this by no means implies that we should collectively refocus our attention on Allen and his works specifically (or that any attention to Allen's works must focus on these dimensions alone). On the one hand, awareness of the problematic dimensions of Allen's work should prompt attention to these same dimensions in other social-cultural artistic representations – not a narrow fixation on Allen alone. On the other hand, and more constructively, it is important to broaden attention to other artists, to practice the virtue of lateral regard. I won't say that I will never watch another Woody Allen film, but if and when I do, I will be attentive to the conditions of my viewing, the critical perspective I bring to it, and its relative place in my cinematic diet more broadly.

[57] Farrow, "Open Letter."

Briefly, I will address a few points of deviation from the Allen or C.K. cases, in which the artists' personal lives and the contents of their works are closely aligned. First of all, we might consider an artist accused of sexual harassment or abuse whose actions are not, or are not self-evidently, reflected in the art itself. Bill Cosby, for example, perturbs less because of a continuity between his personal life and his character than because of its discontinuity: the rapist who plays the loveable family man. Not all of the same questions around consuming *The Cosby Show* apply as they do to *Manhattan* (this reflects, in part, the difference between acting and directing). Nevertheless, the same questions about the extent to which Cosby's sexual assault record influences the consumption of his work (particularly insofar as it might entail economic support and cultural status) and the discussion of his significance as a cultural icon do apply. Second, historical context matters in evaluating an artist's actions even if it does not simply excuse actions deemed reprehensible by contemporary standards. For example, Paul Gauguin is one of the modern art figures who has attracted some of the most excoriating post-colonialist and feminist critique – and rightly so.[58] Gauguin abandoned his wife and five children in search of an idealized, "primitive" Eden in the French-colonized Polynesian islands. There he took three teenage brides (between the ages of 13 and 14), whom, in addition to painting, he may or may not have infected with syphilis.[59] That Gauguin's behavior would not have been perceived the same way by the standards of the time as it is by ours does not diminish the force of the contemporary moral evaluation one makes of his participation in colonial and sexual practices abhorrent on contemporary terms and the question of how to view, interpret, and respond to his works in light of that evaluation. It is, nevertheless, important to situate Gauguin in his context just as one's own judgment is situated in a particular context: both historical and contemporary context should inform how one understands and evaluates both his actions, his works, and influence (in all of its possible range, positive and negative).[60] This does not imply homogenous interpretations or valuations of his works – feminist interpretations of the works and their relationship to his life and his influence may vary widely, and are by no means bound to reach specific conclusions. Assessment of his work may attend both to the aesthetically relevant ethical dimensions of the works (for example, the attitudes implied towards its subjects; the character of both the "manifest artist" and the real-life artist) and to its other aesthetic qualities.

Part of the question that Gauguin's work opens up is the role of institutions such as museums and galleries in integrating critique. On this point, I argue that exhibitions and discussions of Gauguin's works should present post-colonialist and feminist accounts of imperial tourism and colonial exoticism just as they present accounts of his place in

[58] Paul van der Grijp, *Art and Exoticism: An Anthropology of the Yearning for Authenticity* (Berlin: LIT Verlag, 2009); Griselda Pollock, *Avant-Garde Gambits 1888-1893: Gender and the Color of Art History* (London: Thames & Hudson, 1993); Abigail Solomon-Godeau, "Going Native: Paul Gauguin and the Invention of Primitivist Modernism," in *Art in America* 77 (July 1989), pp. 118-29; Belinda Thomson (ed), *Gauguin: Maker of Myth* (London: Tate Modern, 2010).

[59] This dominant assertion has recently been qualified, as a set of teeth that are very likely Gauguin's have shown no evidence of mercury, the treatment for syphilis at the time, therefore implying either that he did not have syphilis or simply that he was not treated for it. William A Mueller and Caroline Boyle Turner, "Gauguin's Teeth," *Anthropology* 6 :1 (2018), DOI : 10.4172/2332-0915.1000198.

[60] See Gaut, *Art*, pp. 79-80, on historical context.

developments in modern art.[61] Moreover, art institutions also have a responsibility to practice lateral regard. The Guggenheim's exhibition of Hilma af Klint's work, long under-recognized for its innovative abstraction that preceded Kandinsky, Malevich, and Mondrian (typically credited as the innovators of non-representative art), is an example of how attending to underrepresented figures may invite a reevaluation of a received artistic canon and history.[62]

The institutional dimension raises questions about how to assess individual actions that reflect and contribute to collective socio-cultural structures. This touches questions about how art – regardless of an artist's personal life – reflects and participates in social and cultural structures informed by norms and values that may be distorted. This is an issue that I will not fully address, as I focus on artists who have been accused of morally and/or criminally reprehensible actions. But insofar as individual and collective moral-aesthetic formation and responsibility are intertwined, as I think they are, the boundaries of these discussions are porous. The formative dimension of aesthetic involvement and consumption raises questions about answerability and responsibility. This requires recognizing how both artists and viewers are formed in variously distorted social structures and that both artists and viewers ought to consider and respond to their implication – both attitudinal and active – in these distorted structures.

The meaning of individual acts is found in the nexus of practices, dispositions, and affections in which they are taken.[63] This means that human action is inextricable from the (distorted) social structures in which it is situated. One may thus be pervasively implicated in collectively distorted structures. This raises questions about the extent to which we are answerable and responsible for our participation in those structures, especially insofar as we may have been non-deliberately formed by them.[64] Regardless of how one frames the principle of answerability and responsibility on an ethical theoretical front, one ought to take seriously and respond to contestations of one's implicit assumptions and attitudes and the actions to which they lead, when such contestations arise. Insofar as aesthetic involvement may dovetail with moral formation, such contestations of one's implicit assumptions and attitudes may emerge in response to aesthetic betrayals of the sort I have discussed in this paper. Though I do not engage the question of collective judgments on prohibit access to certain works, I have touched on the heightened obligation that institutions and their members bear for considering how they present artists and their artworks, and the kinds of involvements they have with the works and foster in the public, insofar as they have a role in framing and direction attention to artists and works. I am not suggesting any particular default position by art institutions or art historians. But the

[61] This is a judgment regarding art institutions and critics who are engaging Gauguin given his place in a "common artistic record." This reflects the reality that one's position in a field relative to certain artists also influences one's consideration of whether or not to engage an artist.
[62] Caitlin Dover, "Who Was Hilma af Klint?: At the Guggenheim, Paintings by an Artist Ahead of Her Time," October 11, 2018, https://www.guggenheim.org/blogs/checklist/who-was-hilma-af-klint-at-the-guggenheim-paintings-by-an-artist-ahead-of-her-time (accessed 2018-11-8).
[63] Ryan Darr, "Social Sin and Social Wrongs: Moral Responsibility in a Structurally Disordered World," *Journal of the Society of Christian Ethics* 37:2 (Fall/Winter 2017), pp. 21-37.
[64] For one such account, see Ryan Darr's claim that "intersubjective answerability" means that "then we are responsible not only for that which is freely chosen or under voluntary control but also for anything for which it is possible to give reasons and which is, at least in principle, open to modification by rational considerations." Darr, "Social Wrongs," p. 29.

question of how we should respond to art by artists who have been accused of sexual wrongdoing arises when the art and the artist are already embedded in formative relationships of some kind: that is, in existing structures of aesthetic involvement. Reflexive consideration of one's responsibility falls more heavily on those who are involved with the artworks or artist in some regard, and institutions and persons involved in existing structures by virtue of their profession, expertise, or field of study are inevitably in such positions of heightened consideration.

Throughout this article, I have focused on the question of what *we* should do with the art of the accused. In doing so, I do not want to confuse the responsibility harassers and abusers bear for their actions and the responsibility viewers bear in evaluating their implication. But my purpose has been to articulate how to respond to wrongdoing as a viewer, and I have argued that this requires reflexive and social-structural examination. A reflexive move is necessary in cases where one discovers oneself to be implicated as a viewer, as an appreciator, as someone who becomes aesthetically involved with people and things and works and is shaped by them. Reassessing a work requires consideration not only of the object's aesthetic value but of the formative implications of aesthetic involvement, individually and collectively. It also means that ongoing engagement with artist's work can and should be recognized as a legitimate dimension of grappling with de-formation and its harms, and that this can be compatible with appreciation of and interpretive attention to the works. At the same time, I have tried to articulate conditions and considerations for engagement that address the responsibility one has for considering how and why one chooses to engage, in ways that will presumably inflect that engagement in important ways.

Sarah Stewart-Kroeker, University of Geneva
sarah.stewart-kroeker@unige.ch

Bibliography

Ahmed, Sarah. *Living a Feminist Life*. Durham: Duke University Press, 2017.

Berkowitz, Joe. "In A Post-Weinstein World, Louis CK's Movie Is a Total Disaster." *Fast Company*, October 20 2017. https://www.fastcompany.com/40483261/in-a-post-weinstein-world-louis-cks-movie-is-a-total-disaster. Accessed 2018-11-6.

Chrisman-Campbell, Kimberly. "When Casanova Met #MeToo." *The Boston Globe*, March 3, 2018. https://www.bostonglobe.com/ideas/2018/03/03/when-casanova-met-metoo/ALNHZdUQJJZ41KKylv8hqM/story.html. Acessed 2018-11-6.

Coyle, Deirdre. "Men Recommend David Foster Wallace to Me." *Electric Lit*, April 17, 2017. https://electricliterature.com/men-recommend-david-foster-wallace-to-me-7889a9dc6f03. Accessed 2018-11-6.

Cramer, David, Jenny Howell, Jonathan Tran, Paul Martens, "Scandalizing John Howard Yoder," *The Other Journal: An Intersection of Theology & Culture*, July 7, 2014. https://theotherjournal.com/2014/07/07/scandalizing-john-howard-yoder/. Accessed 2018-11-8.

Darr, Ryan. "Social Sin and Social Wrongs: Moral Responsibility in a Structurally Disordered World." *Journal of the Society of Christian Ethics* 37:2 (Fall/Winter 2017), pp. 21-37.

Deb, Sopan and Deborah Leiderman. "Woody Allen, Mia Farrow, Soon-Yi Previn, Dylan Farrow: A Timeline." *The New York Times*, January 31, 2018. https://www.nytimes.com/2018/01/31/movies/woody-allen-mia-farrow-dylan-farrow-a-timeline.html. Accessed 2018-11-2.

Dederer, Claire. "What Do We Do With the Art of Monstrous Men?" *The Paris Review*, November 20, 2017. https://www.theparisreview.org/blog/2017/11/20/art-monstrous-men/. Accessed 2018-11-8.

Donegan, Moira. "I Started the Media Men List. My name is Moira Donegan." *The Cut*, January 10, 2018. https://www.thecut.com/2018/01/moira-donegan-i-started-the-media-men-list.html. Accessed 2018-11-8.

Dover, Caitlin. "Who Was Hilma af Klint?: At the Guggenheim, Paintings by an Artist Ahead of Her Time." October 11, 2018, https://www.guggenheim.org/blogs/checklist/who-was-hilma-af-klint-at-the-guggenheim-paintings-by-an-artist-ahead-of-her-time. Accessed 2018-11-8.

Durcanin, Cynthia. "Casanova as Case Study: How Should Art Museums Present Problematic Aspects of the Past ?" *Art News*, August 13, 2018. http://www.artnews.com/2018/08/13/casanova-case-study-art-museums-present-problematic-aspects-past/. Accessed 2018-11-6.

Farin, Ingo and Jeff Malpas (eds). *Reading Heidegger's Black Notebooks 1931-194*1. Cambridge: MIT Press, 2016.

Farrow, Dylan. "An Open Letter from Dylan Farrow." *The New York Times*, February 1, 2014. https://kristof.blogs.nytimes.com/2014/02/01/an-open-letter-from-dylan-farrow/. Accessed 2018-11-7.

Gaut, Berys. *Art, Emotion and Ethics*. Oxford: Oxford University Press, 2007.

Goossen, Rachel Waltner. "Defanging the Beast: Mennonite Responses to John Howard Yoder's Sexual Abuse." *Mennonite Quarterly Review* 89 (January 2015), pp. 7-80.

Hauerwas, Stanley. "In Defence of 'Our Respectable Culture': Trying to Make Sense of John Howard Yoder's Sexual Abuse." ABC Religion and Ethics, October 18, 2017. http://www.abc.net.au/religion/articles/2017/10/18/4751367.htm. Accessed 2018-11-8.

Hess, Amanda. "How the Myth of the Artistic Genius Excuses the Abuse of Women." *The New York Times*, November 10, 2017. https://www.nytimes.com/2017/11/10/arts/sexual-harassment-art-hollywood.html. Accessed 2018-11-8.

Hungerford, Amy. "On Not Reading." *The Chronicle of Higher Education*, September 11 2016. https://www.chronicle.com/article/On-Refusing-to-Read/237717. Accessed 2018-11-6.

Izadi, Elahe. "Why the 'Last Tango in Paris' rape scene is generating such an outcry now." *The Washington Post*, December 5 2016. https://wapo.st/2g0vcLt?tid=ss_mail&utm_term=.1b18e8f8fcec. Accessed 2018-11-2.

Kolitz, Daniel. "Academics Explain David Foster Wallace To Me." *The Outline*, July 25, 2018, https://theoutline.com/post/5543/david-foster-wallace-conference-profile?zd=1&zi=hoxjdain. Accessed 2018-11-8.

Kurutz, Steve. "How Do You Solve a Problem Like Manhattan?" *The New York Times*, March 1 2018. https://nyti.ms/2FHasqw. Accessed 2018-11-2.

Lyall, Sarah and David Itzkoff. "Charlie Rose, Louis C.K., Kevin Spacey: Rebuked. Now What Do We Do With Their Work?" *The New York Times*, November 24, 2017.

https://www.nytimes.com/2017/11/24/arts/charlie-rose-kevin-spacey-louis-ck-art.html. Accessed 2018-11-8.

Manne, Kate. *Down Girl: The Logic of Misogyny*. Oxford: Oxford University Press, 2018.

Mintcheva, Svetlana. "Caravaggio killed a man. Should we therefore censor his art?" *The Guardian*, February 3, 2018. https://www.theguardian.com/commentisfree/2018/feb/03/caravaggio-killed-a-man-censor-art?CMP=share_btn_fb. Accessed 2018-11-8.

Mitchell, Andrew and Peter Trawny (eds). *Heidegger's Black Notebooks: Responses to Anti-Semitism*. Columbia University Press, 2017.

Mueller, William A. and Caroline Boyle Turner. "Gauguin's Teeth." *Anthropology* 6:1 (2018). DOI : 10.4172/2332-0915.1000198.

Nehamas, Alexander. *Only a Promise of Happiness: The Place of Beauty in a World of Art.* Princeton: Princeton University Press, 2007.

Paulson, Steve and Claire Hayes-Brady. "David Foster Wallace in the #MeToo Era: A Conversation with Clare Hayes-Brady." *Los Angeles Review of Books*, September 10, 2018. https://lareviewofbooks.org/article/david-foster-wallace-in-the-metoo-era-a-conversation-with-clare-hayes-brady/. Accessed 2018-11-8.

Pollock, Griselda. *Avant-Garde Gambits 1888-1893: Gender and the Color of Art History.* London: Thames & Hudson, 1993.

Ryzik, Melena, Cara Buckley, and Jodi Kantor. "Louis C.K. Is Accused by 5 Women of Sexual Misconduct." *The New York Times*, November 9, 2017. https://nyti.ms/2ho0aE0. Accessed 2018-11-6.

Rothfeld, Becca. "Can Sexual Predators Be Good Scholars?" *The Chronicle of Higher Education*, December 07 2017. https://www.chronicle.com/article/Can-Sexual-Predators-Be-Good/242002?cid=wcontentgrid_40_5. Accessed 2018-11-8.

Scarry, Elaine. *On Beauty and Being Just*. Princeton: Princeton University Press, 2001.

Scarsella, Hilary. "Not Making Sense: Why Stanley Hauerwas's Response to Yoder's Sexual Abuse Misses the Mark." ABC Religion and Ethics**,** Updated 1 Dec 2017 (First posted 30 Nov 2017). http://www.abc.net.au/religion/articles/2017/11/30/4774014.htm. Accessed 2018-11-8.

Scott, A.O. "My Woody Allen Problem." *The New York Times*, January 31 2018. https://www.nytimes.com/2018/01/31/movies/woody-allen.html. Accessed 2018-11-8.

Solomon-Godeau, Abigail. "Going Native: Paul Gauguin and the Invention of Primitivist Modernism." in *Art in America* 77 (July 1989), pp. 118-29.

Thomson, Belinda (ed). *Gauguin: Maker of Myth*. London: Tate Modern, 2010.

van der Grijp, Paul. *Art and Exoticism: An Anthropology of the Yearning for Authenticity.* Berlin: LIT Verlag, 2009.

On Some Moral Implications of Linguistic Narrativism Theory

Natan Elgabsi and Bennett Gilbert

In this essay we consider the moral claims of one branch of non-realist theory known as linguistic narrativism theory. By highlighting the moral implications of linguistic narrativism theory, we argue that the "moral vision" expressed by this theory can entail, at worst, undesirable moral agnosticism if not related to a transcendental and supra-personal normativity in our moral life. With its appeal to volitionism and intuitionism, the ethical sensitivity of this theory enters into difficulties brought about by several internal tensions as to what morality and moral judgements involve. We contend that the proponents of linguistic narrativism theory must strongly recognize and take responsibility for the "moral vison" their theory professes, in so far as they want to think of their theory as a morally responsible one.

1. Linguistic narrativism theory and moral life

Is theoretical investigation of the human being separable from explicating a vision of human life? If it is inseparable, how do proponents of theoretical perspectives reflect and explore the life-visions that they develop and advocate in theory? Usually we think that description and evaluation in virtually every type of inquiry are ontologically separable as a matter of basic principle. On this principle, both empirical descriptions of states of affairs and theoretical descriptions of conceptualized relationships avoid involvement with judgements of value, particularly with moral value.[1]

Philosopher Iris Murdoch, however, argues that evaluative features stand in any description of matters concerning the human being and that these features necessarily involve elements not fully explicable through concepts and propositions. Theoretical and empirical descriptions of human matters are, as she says, expressions of a certain "vision of life." They therefore commit us to ethical evaluation.[2]

Proponents of discursivist or linguistic narrativism theories of culture, such as Michel Foucault or Hayden White, would generally agree with Murdoch that descriptions and, even more so, narrations necessitate evaluations. White, for example, writes: "The important point is that although one can 'explain' any worldly phenomenon without

[1] In this essay we used morality/morals and ethics/ethical interchangeably, with no necessary normative distinction between the concepts.
[2] Iris Murdoch, 'Vision and Choice in Morality', in *Existentialists and Mystics: Writings on Philosophy and Literature*, p. 81; Iris Murdoch, *Metaphysics as a Guide to Morals* (London: Vintage, 2003), pp. 25-26.

assessing a value to it, it is impossible to describe anything without also assessing a value or set of values to it."[3]

For both Foucault and White, this means that each person assesses what is "good" and "evil" in particular instances of narrated stories on her own evaluative preferences. Each person endows each narration, so to speak, with such-and-such ethical or political meanings because narratives must always have contemporary ideological force.[4]

The difference between Murdoch on the one hand and Foucault and White on the other, however, is that for Murdoch the evaluative features of our descriptions or narrations do not necessarily express political or ideological sympathies, nor are they adequately explained by such sympathies. In her view, ethical evaluations are not restricted to the categories of "good" and "evil" but instead involve a range of less strong moral concepts that our ordinary decisions in life incorporate. "It is," she says, "in terms of the inner complexity of such concepts that we may display really deep differences of moral vision."[5] Thus, in any particular description at hand we will, according to Murdoch, have to investigate and understand just what "vision of life", or "moral vision", that description necessitates or entails. This is less a suspicious, ideology-critical undertaking than a close moral understanding of what someone's words and ideas mean through the way they are uttered.

In this investigation, we start by supporting Murdoch's notions that theoretical descriptions contain degrees of ethical evaluations that express "moral visions" regarding what our relationships with other persons ought to be and that the relationship between the ideological and the moral is not inescapable. This "moral vision" need not be (and in fact seldom is) an articulated part of the theoretical perspective itself, but it can be implied through the epistemological (and metaphysical) claims of the theory. In order to concretely emphasize the ethically evaluative character of theoretical descriptions, we continue by describing the subject of our investigation: a theory that, according to us, forwards normatively difficult moral claims. The theory is inspired by the linguistic turn in the humanities and may properly be called linguistic narrativism theory, a non-realist theory that is prominent within the philosophy of history, to which our work responds.

Under the linguistic turn, language is regarded as the condition of the possibility of both having and making sense of experience. Linguistic narrativism theory takes language, rather than any extra-linguistic referents, as the condition of the possibility of forming narratives. We hold that it thereby (1) strongly separates meaningfulness and judgment on the basis of a supposedly non-phenomenological, epistemic cleavage between discourse and the extra-linguistic, factually past and present, reality[6]; and (2) appeals to the moral philosophies of volitionism and intuitionism in order to deal morally with the epistemic cleavage it presupposes, and sets for itself.

[3] Hayden White, *The Practical Past* (Evanston, IL: Northwestern University Press, 2014), pp. 72–73.
[4] Hayden White, 'The Public Relevance of Historical Studies: A Reply to Dirk Moses', *History and Theory*, 44:3 (2005), p. 335; Michel Foucault, 'The History of Sexuality', *Power/Knowledge: Selected Interviews and Other Writings 1972-1977*, edited by Colin Gordon (Brighton: The Harvester Press, 1980), p. 193.
[5] Iris Murdoch, 'Metaphysics and Ethics', *Existentialists and Mystics: Writings on Philosophy and Literature*, edited by Peter Conradi (New York: Penguin Books, 1997), p. 73.
[6] Kalle Pihlainen, *The Work of History: Constructivism and a Politics of the Past* (London: Routledge. 2017), p. xiii; Eugen Zeleňák, 'Who Should Characterize the Nature of History? The Wrong Question?', *Storia della Storiografia* 59–60 (2011), pp. 173–174.

In order to avoid confusion, it is necessary to distinguish the two broad kinds of narrative theory in the philosophy of history. The first is phenomenological, associated with the work of Paul Ricoeur and David Carr, by which the meaningfulness of human life has a narrative structure.[7] On this account, the horizon of our lives is a coinciding of meaning and storytelling, within which explicit moral assessments are a part. The other is linguistic narrativism theory as described above. This second type of narrative theory—i.e. linguistic narrativism theory—is what we address in this study.

By highlighting the moral implications of linguistic narrativism theory, we argue that the "moral vision" expressed by this theory can entail, at worst, undesirable moral agnosticism if not related to a transcendental and supra-personal normativity in our moral life. We will show that the theory expresses a "moral vison" that involves, on the one hand, an implicit evaluation of what our relationships with other people look like, and on the other, an explicit understanding of the character of ethical assessments and moral judgments. Thus, even if these higher-order moral commitments are not philosophically argued for within linguistic narrativism theory as such, the kinds of normative ethics, as explained below, that remain for this and related non-realist theories may, by merciless subtraction, stunt the moral aims of their own proponents. In the analysis, we contend that the proponents of linguistic narrativism theory must strongly recognize and take responsibility for the "moral vison" their theory professes, in so far as they want to think of their theory as a morally responsible one.

2. Theoretical description and its moral implications

Amidst the boundaries that some philosophers hold language, or psychic drives, or socio-economic pressures, or the conservatism of logic to impose, how do we salvage reflection on moral life in order to parse out what is good? Murdoch's concept of "visions" are the changing observations, reflections, and conclusion, by which we can work our way towards deeper states of moral deliberation and understanding. How one sees the world affects how one thinks about it and, by the same logic, how one lives in it; acting and thinking connects in what we call moral life. The improvement of vision helps us to understand morality better, as well as allowing us to see people in a truer light, improving our judgments and choices in the world.[8] The process of vision is Murdoch's move against the Humean separation of facts from values in the various forms it took in the post-War Anglophone philosophy of her day. It points away from propositionalist concept analysis and empirical research and points toward accounts of the dynamic, temporal, and historical course that people's "moral visions" often takes.[9]

Before giving a more in-depth description of the "moral vision" of linguistic narrativism theory, it is necessary to provide a critique of the customary Humean bifurcation of "is" from "ought" and of facts from values that the description/evaluation distinction tends to rest on**.** The distinction seems in some lights to protect values from reduction to facts and in other lights fatally expose them to the same disintegrative peril.

[7] David Carr, *Experience and History: Phenomenological Perspectives on the Historical World* (Oxford: Oxford University Press, 2014); Paul Ricoeur, *Time and Narrative,* volumes 1 and 3, translated by Kathleen McLaughlin and David Pellauer (Chicago: University of Chicago Press, 1984, 1985).
[8] Murdoch, 'Vision and Choice in Morality', pp. 79–83.
[9] Talbot Brewer, *The Retrieval of Ethics* (Oxford: Oxford University Press, 2011).

This situation of opposing outcomes is the reason that G. E. Moore's "naturalistic fallacy" failed to solve the problem of the status of ethical values that it was intended to solve. If values are ontologically other than facts in a profound way, then they are unintelligible in terms of facts. In this case they (1) might be senseless as things other than facts and intelligible solely as facts about what people believe to be good and bad or (2) are ineffable and require discourse, practices, beliefs, or speculation of types wholly outside the domain of empirical verification.[10] Both results are possible from the is/ought distinction, although of course many people rigidly take one side or the other. Nevertheless, thinkers, sometimes without being unaware of it, breech, or even reject the fact/value binary, for instance, for the reason that when scientists attempt to find moral direction in immanent natural ends, their approach requires the surreptitious re-insertion of intuitive, teleological, political, religious, or metaphysical claims that subvert the original enterprise.[11]

Elizabeth Anscombe attempted to answer some of the difficulties in the ontological separation between fact and value. Her examples, derived from David Hume, tell us not only that evaluations of what we recognize as relevant to the factuality and truth of a situation concerning human states of affairs—such as the evaluative difference between "delivering the potatoes" and "supplying me with potatoes"—are intimately connected to how we come to frame or describe such states of affairs.[12] Her examples also emphasize that evaluation in terms of understanding or even judging human matters of fact supervenes on the way human states of affairs are described in the first place.[13] Descriptions of human matters—"supplying me with potatoes"—are, one could say, already *evaluations* that lean on what we ordinarily mean when describing such-and-such a deed, unless the circumstances are such that one must suspect one means something different. That, too, however, is an evaluation—namely a judgement that one should not understand the action the way we normally mean.

This is relevant to our understanding of theoretical perspectives not only because evaluative features do not belong exclusively to our descriptions of concrete human state of affairs, helping us to recognize what is relevant and not relevant when determining facticity and truth in a situation, but also because, as Murdoch argues, different degrees of evaluation belong to theoretical or hypothetical descriptions that are systematic expressions of a world view. For one could argue that in our life, even when we reiterate a theory: "…*various* values pervade and *colour* what we take to be the reality of our world; wherein we constantly evaluate our own values and those of others, and judge and determine forms of consciousness and modes of being."[14]

Theoretical descriptions often express, Murdoch argues, the values of what we take to be fundamental relationships or characteristics of our human lifeworld. Theories of

[10] George Edward Moore, *Principia Ethica* (Cambridge: Cambridge University Press, 1922), pp. 10–14; cf. G.E.M. Anscombe, 'On Brute Facts', *Ethics, Religion and Politics: Collected Philosophical Papers Vol. III* (Oxford: Basil Blackwell, 1981), p. 23.
[11] In philosophy, see Murdoch, *Metaphysics as a Guide to Morals*, pp. 25–26; in philosophy of science, see John Dupré, 'Fact and Value', *Value-Free Science? Ideals and Illusions*, edited by Harold Kincaid, John Dupré, and Alison Wylie (Oxford: Oxford University Press, 2007), pp. 27–41; and in discursivist theory of culture, see White, *The Practical Past*, pp. 72–73.
[12] David Hume, *A Treatise of Human Nature: Being and Attempt to Introduce the Experimental Method of Reasoning into Moral Subjects: Book I-III*, edited by L.A. Selby-Bigge (Oxford: Clarendon Press, 1881), p. 458.
[13] Anscombe, 'On Brute Facts', pp. 22–23.
[14] Murdoch, *Metaphysics as a Guide to Morals*, p. 26 (italics in original).

language, of mind, of ethics, and of culture hinge on predispositions that guide our ideas of consciousness or our "modes of being." Structuralism, one of the principal drives of linguistic narrativism theory[15], is an example of a world-view that leans on strong ethical presuppositions: "Structuralism, in so far as it offers itself as 'scientific', must profess to be morally neutral; but the large and various volume of structuralist writings contains innumerable value judgements."[16]

Structuralism, in which she includes most post-Saussurean theory of language and culture, makes strong value-judgements on the concept of "truth" in language, presumed as the morally neutral description of language from no particular point of view that posits that language in certain ways is estranged from the world it should denote.[17] Nevertheless, while we will not here inquire into the vision of life that Murdoch's presentation of structuralism entails or presupposes, the perspective we need to take for present purposes is that theory is explicated from a certain evaluative point of view in life. "Theory" itself is a certain vision *of* life. It expresses a relationship to the persons with whom we live.

From Anscombe's and Murdoch's claims that different evaluations enter any empirical and theoretical description of human states of affairs, we now turn to a prominent example of linguistic narrativism theory in order to understand its moral implications and the vision of life it professes. In *The Ethics of Theory* Robert Doran describes Hayden White's prominent theoretical vision of (historical) reality in the following way:

> As for relativism, White admits to being a pluralist with regard to the view that no single, overreaching perspective on reality or history can claim epistemological priority. But this does not thereby render moral judgment impossible or moot. On the contrary, moral judgement is returned to the realm of ethical responsibility (choice), from which it had become estranged by the supposedly "value-neutral" perspective of historical objectivism.[18]

Doran shows us that White's vision contains several *a priori* statements about the relationship between, on the one hand epistemology and reality, on the other epistemology and ethics. Analysis of this relationship leads us to three insights into linguistic narrativism theory:

(1). Different "perspectives on reality"—i.e., interpretations and narrations based on what is factually asserted as true or false by our world-view—cannot have epistemic status because they cannot be "subject to epistemological conditions of true and false."[19] White himself writes that "the plot-structures used to fashion the different stories are not in themselves in the nature of propositions that can be submitted to tests of verification or

[15] Kalle Pihlainen, 'The Work of Hayden White II: Defamiliarizing Narrative', *The SAGE Handbook of Historical Theory*, edited by Nancy Partner and Sarah Foot (Los Angeles: SAGE, 2013), pp. 121–123.

[16] Murdoch, *Metaphysics as a Guide to Morals*, pp. 46–47.

[17] Murdoch, *Metaphysics as a Guide to Morals*, pp. 192–195. One example of linguistic structuralism is Roland Barthes, 'Historical Discourse', *Structuralism: A Reader*, edited by Michael Lane, translated by Peter Wexler (London: Jonathan Cape, 1970), pp. 145–155.

[18] Robert Doran, *The Ethics of Theory: Philosophy, History, Literature* (London: Bloomsbury Publishing. 2017), p. 123.

[19] Doran, *The Ethics of Theory*, p. 122.

falsification in the way that 'singular existential statements' (i.e., facts) can be tested."[20] Thus, "perspectives on reality," conditioned by these plot-structures, or forms of story-telling, cannot be subject to epistemic evaluation as to "epistemological priority." Linguistic narrativism theory holds that only facts— "singular existential statements"— can be subject to such evaluation.

This distinction, however, rests on one deeper ontological assumption, or what Murdoch called a "mode of being."[21] Doran writes that "of course, on a deeper level, what *counts* as a fact is itself dependent on a priori world-disclosure...in our case that of the scientific-naturalist view of the world."[22] In other words, the judgement of what has "epistemological priority" (what is factually true and not) belongs not to narrative construction but to the deeper ontological stratum of one particular scientific and philosophical world-view:

> What counted as historical "fact" in medieval Europe was very different from what counts as fact in a contemporary context. We no longer permit miracles and witchcraft to serve as factual elements in the historical account; they are simply described with ironic distance. Thus, in the most general sense, all facts are dependent on a priori interpretation.[23]

Fact, then, refers to one particular vision of the world that conditions everything that we take as a state of affairs in our contemporary way of life. To use Bernard Williams' terms, one could say that the medieval fact of miracles and witchcraft is not factually "possible" in our world governed by the scientific-naturalist world-view.[24] This idea of the relationship between epistemology and reality, or in actual fact between epistemology and "perspectives on reality," has implications also for the relationship between epistemology and ethics. That consequence may be put as follows.

(2). The theory pursues the claim that what is epistemically true or false on the naturalist estimation is an epistemic judgment that is "subject to epistemological conditions of true and false."[25] Thus, fact is not a matter of ethical choice or interpretation; moral judgements typically belong to the practical construction of a story, which stipulates what is contextually true and false, aesthetically beautiful, or politically or ideologically effective. In other words, epistemology (with its epistemic judgments about states of affairs) is categorically separated from the realm of ethics (with its moral judgements about how one should *arrange* and *understand* these states of affairs).

There is, however, also a third result that is important for understanding the nature of linguistic narrativism theory.

(3). In this theory ethics is itself fundamentally envisioned as volitionism, that is to say, as a matter of making present choices on individual intuitions and preferences.[26] The

[20] Hayden White, 'Historical Pluralism and Pantextualism', *The Fiction of Narrative: Essays on History, Literature, and Theory 1957-2007*, edited by Robert Doran (Baltimore: The Johns Hopkins University Press, 2010), p. 232; cf. Doran, *The Ethics of Theory,* p. 122.
[21] Murdoch, *Metaphysics as a Guide to Morals*, p. 26.
[22] Doran, *The Ethics of Theory*, p. 122 (italics in original).
[23] ibid.
[24] Bernard Williams, *Ethics and the Limits of Philosophy* (Cambridge, Mass.: Harvard University Press, 2006), pp. 139–140, 142.
[25] Doran, *The Ethics of Theory*, p. 122.
[26] Jean-Paul Sartre, *Existentialism is a Humanism*, translated by Carol Macomber (New Haven: Yale University Press, 2007), pp. 23–28.

naturalist's assertion of fact is, in this respect, not subject to an ethical choice (a moral judgment) even if he describes facts (states of affairs). Facts are not evaluated on ethical decisions but are instead conditioned by what is epistemically true and false according to the naturalist world-view. The human scientist, by contrast, is inherently subject to an ethical choice, because any emplotment of story (even if it refers to states of affairs) is inevitably made on moral preferences. The story is written in the human scientist's ethical decisions.

The point of looking at linguistic narrativism theory in this way is to show what "vision of life," or "moral vison," it expresses.[27] On this description, Doran even argues that the theory involves strong categorical presuppositions that should be treated analogously to the categories in Immanuel Kant's *Critique of Pure Reason*. Thus, to say that linguistic narrativism theoretically legitimizes immoral histories, is "like asking if the Nazi also perceives the world according to the categories of the understanding and the forms of intuition as outlined in Kant's first *Critique*. From a metahistorical perspective, emplotment is simply what every historian does, Nazi or not."[28] This is where ethical tensions tend to arise. A generous reading of this idea would imply that theoretical legitimation of narrative work as simply the categorical forms that any writing of narrative as well as scholarly description of such work must take are quite different things from judging what that work morally means. This must mean that the author's epistemic responsibility is (exactly as a positivist or empiricist would claim) tied to being true to the asserted facts, whereas her moral responsibility is tied to the moral consequence of what her emplotted narrative may play out to mean in the present. The question is, however, against what normative background one is to judge whether an author is true to the things she claims, both epistemically and morally?

What distresses the "moral vision" to which linguistic narrativism theory can lead is the fact that descriptions of the mere nature of narrative can theoretically legitimize immoral narratives by virtue of describing *only* what anyone does when writing narrative. This entails that one will not be able to distinguish a theoretical description of the fact that such-and-such a narrative is undertaken from the agnostic or morally relativistic response to the fact that such-and-such a narrative is undertaken but not responded to as immoral. In that way the moral idleness of linguistic narrativism theory challenges Doran's own, or anyone's, claim that the existential and philosophical underpinnings of this kind of discursive cultural theory should be understood purely as "epistemological relativism" and not as an agnostic form of "moral relativism."[29] Nevertheless, on the assumption that this theoretical approach itself does not entail morally agnostic visions of life and that it should not do so, because one needs to distinguish theoretical description from moral response, moral agnosticism announces itself also in yet another sense. From the viewpoint of a writer of a narrative herself, in order not to provide justification for her own possible agnosticism or immorality, linguistic narrativism theory would need a normativity that is not bound to the intuitions and choices of the author herself. We will argue that it is not possible to establish this moral normativity by the means of Doran's and White's volitional

[27] Murdoch, 'Vision and Choice in Morality', p. 81; Murdoch, 'Metaphysics and Ethics', p. 73.
[28] Doran, *The Ethics of Theory*, p. 119; see also Elizabeth Deeds Ermarth, 'Ethics and Method', *History and Theory*, 43:4 (2004), pp. 75–76; and her *History in the Discursive Condition: Reconsidering the Tools of Thought* (New York: Routledge, 2011), p. 94.
[29] Doran, *The Ethics of Theory*, p. 119.

choice-ethics insofar as volitional ethics has only the resources of moral intuitionism by which to justify or deliberate moral claims.

3. Moral vision, volition, and intuition in linguistic narrativism theory

Even if linguistic narrativism theory is constructivist and non-realist in nature, its proponents take on a normative stance derived from moral intuitionism and volitionism regarding ethical matters—a stance that may arguably involve a sense of "realism." This may sound like a contradiction. As a meta-ethical theory, intuitionism, as in G. E. Moore's thought, is a form of "moral realism."[30] Our question, however, is how proponents of linguistic narrativism theory think of the "reality" that their moral judgements are supposed to be grounded in; furthermore, what "moral vision" does linguistic narrativism theory that appeals to the normativity of moral intuitionist and volitionist theses imply?

In this regard, although volitionism is sometimes used as a descriptive meta-ethical or "psychological" thesis whilst intuition is assigned to carry out the normative task, we use both terms chiefly to name prescriptive theses, although they both do some descriptive work as well.[31] On the one hand, this is because proponents of linguistic narrativism theory seldom themselves know whether they move on a meta-ethical or normative level. And on the other hand, as Murdoch argues, since all the objects of meta-ethical theory are normative ideas, it is difficult even in theory to empty the moral concepts from their evaluative content.[32]

Volitionism holds that it is justifiable to assign values and to make consequential choices on the basis of will, and according to one's interests or emotions, without strictly requiring evidentiary or philosophical reasons for the sake of moral responsibility.[33] Intuitionism (though the term sometimes also refers to psychological states or mechanisms) holds that assigning values and making consequential choices are sufficiently *justified* by intuitions because intuitions are direct cognition of right and wrong in so far as humans can possibly know what is right and what is wrong.[34] The important prescriptive purchase of the two approaches is that they provide what our moral judgements *should* be or, at the least, *must* be, since they are never anything else.

Regardless of whether we refer to the existentialist "doctrine"[35] or to the naturalist "psychology"[36] of volitionism and intuitionism, they both connect intimately to linguistic narrativism theory because it, in its canonical form, holds that the use of discourse constitutes our individual imaginative acts of enunciation and is not "epiphenomenal" to

[30] For a discussion of how Moore's "moral realism" in *Prinicipia Ethica* can be understood, see Camilla Kronqvist, 'Westermarck and Moore on the Sources of Morality', *Evolution, Human Behaviour and Morality: The Legacy of Westermarck*, edited by Olli Lagerspetz, Jan Antfolk, Ylva Gustafsson and Camilla Kronqvist (New York: Routledge, 2017), pp. 129–141.
[31] Pascal Engel, 'Volition and Voluntarism About Belief', *Belief, Cognition and the Will*, edited by A. Meijers (Tilburg: Tilburg University Press. 1999), pp. 9–25.
[32] Murdoch, 'Vision and Choice in Morality', pp. 80–82.
[33] Sartre, *Existentialism is a Humanism*, pp. 23–32; cf. Roland Barthes, *Empire of Signs*, translated by Richard Howard (New York: Hill and Wang, 1982), pp. 69–71.
[34] Walter Sinnott-Armstrong, *Moral Skepticisms* (Oxford: Oxford University Press, 2006), pp. 184–197.
[35] Sartre, *Existentialism is a Humanism*, pp. 18–20.
[36] Brian Leiter, 'Normativity for Naturalists', *Philosophical Issues: A Supplement to Noûs*, 25:1 (2015), pp. 64–66.

some other "reality" or some other human activity, including the classically humanistic subject or self.[37] It argues that analysis pursuant to this idea unsettles habit and undermines convention.[38] Kalle Pihlainen has emphasized the way this insures our taking moral responsibility away from the spurious authority of an empirically unavailable reality and onto one's self.[39]

In Pihlainen's terms, proponents of linguistic narrativism theory often want to actualize the emancipatory possibilities of theoretical reflection by thinking of language as performative acts and by not relying on the authority of "the reality of the past."[40] With this intention, although the scholar in one sense frees herself, she has also limited herself to her own sentiments and choices in the absence of "facts." She may well proceed beyond those choices through analysis, but what activates the impulse to do so? Although the theory itself, as Doran suggests, does not necessarily entail agnostic forms of "moral relativism,"[41] the scholar's volition nonetheless faces toward her own self and toward the self's intuitions if she has no other subject of moral inquiry to which she must be responsible. In that case, her preferred prejudices, or habits, or contemporary conventions take over—a result exactly the opposite of that which post-modernist approaches to understanding the world are supposed to yield.[42] The reason for this reversal lies in the ambiguous character of moral volition and intuition to which this kind of thinking resorts in default of any transcendental ground for judgment.

The fundamental tension as to what constitutes the normative ground for moral judgement in linguistic narrativism theory can be exemplified by a statement of Hayden White's. He writes: "…the ethical opens up a space in which 'something has to be done'. This is quite different from morality that, on the basis of some dogmatism, insists on telling us what we *must and must not do* in a given situation of choice."[43]

White tries to save the ethical responsiveness of moral agents by the distinction between a supposedly non-normative concept of "ethics" and a normative concept of "morality." Hence, if only the person herself can demand what she "must and must not do" with regard to her own action in a volitional situation, it strongly infers that normativity is not a supra-personal or transcendental "moral" demand but a matter of personal (what he calls "ethical") sentiment. On this view, demands on the moral agent are thought to be *intuitively instilled* on whatever principles or sentiments the moral agent herself happen to ethically appeal to (e.g., what she desires or chooses).

The difficulty of this vision is that recognizing "a space in which something has to be done" is not a non-normative claim. Instead, it is a normative one in two important senses: (1) it is normative because it proclaims that one *should* recognize a space as "ethical" (and not, say, as mechanical or natural)—a space in which one knows that one's actions will be judged morally by others in a human life-world; and (2) it is normative because it

[37] Elizabeth Deeds Ermarth, *Sequel to History: Postmodernism and the Crisis of Representational Time* (Princeton: Princeton University Press, 1992), pp. 116–119; cf. Barthes, *Empire of Signs*, p. 70.
[38] Ermarth *Sequel to History*, pp. 41–44.
[39] Pihlanen, *The Work of History*, pp. 48–51.
[40] Pihlainen, *The Work of History*, pp. 92–93.
[41] Doran, *The Ethics of Theory*, p. 119.
[42] Pihlainen, *The Work of History*, p. xvii; cf. Zygmunt Bauman, *Postmodern Ethics* (Oxford: Blackwell, 1995), pp. 31–34.
[43] White, 'The Public Relevance of Historical Studies', p. 338 (italics in original); cf. Michel Foucault, 'On the Genealogy of Ethics: An Overview of Work in Progress', *The Essential Works of Michel Foucault 1954-1984: Vol I*, edited by Paul Rabinow (New York: The New Press, 1997), p 263.

stipulates that one actually *must* do *something* in this "ethical" space of moral agency—thus, one must do not just *anything* but must do that which belongs to being a moral agent in a human lifeworld. Therefore, to critique the normativity of what is called "morality"—moral prescriptions about what we "must and must not do"—is not necessarily to undermine the normativity of what is called "ethics," to wit, that certain things (and not just *anything*) have to be done in certain situations given that we act with other people. In truth, even for White, to break with the first dogma already seems to presuppose the mandate of the second one.

The source of the paradox or tension in this vision is that "ethics" put normatively as "a space in which something has to be done" itself subverts the very possibility that normativity of moral intuition can be intuitively instilled.[44] The reason for this undermining is that, on this account, even if one thinks that the normativity of what exactly should be done in the "ethical" space is a personal, sensible, and volitional act (of enunciation) with no supra-personal or transcendental backgrounds, it is nowhere denied that "ethics" and "morality" are meaningful solely in being and acting for *another*. To recognize this, however, is already in important senses to relate to a supra-personal demand. In other words, if "ethics" is the "space in which something has to be done", it cannot possibly be instilled on my own sentiment, even if my sentiment may be my response. This exact thought reveals the tension in White's thinking.

Thus, from the tension within White's vison, we submit that linguistic narrativism theory invites two categorically different ways to take on what we, despite White's distinction, synonymously call ethics or morality. (1) It can either take the route that any talk about "ethics" and "morality" in linguistic narrativism theory presupposes a supra-personal human lifeworld against which the actions of the moral agent are judged on a normativity that is not instilled by the agent herself. Or (2) it can take the route of denial of any meaningful talk about normativity altogether (thus any meaningful talk about "ethics" and "morality"), falling into a morally agnostic description of "ethics" as textual code, or at worst solipsism.

Within the theoretical cluster of linguistic narrativism theory, Pihlainen is the one scholar who most strongly emphasizes the importance of the first of these two alternatives. He argues that any reasonable vision of morality cannot in theory overlook the fact that it constitutes a relationship to other real people. Regarding non-fiction, we are, he says, steadily reminded that:

> …historical narratives represent *particular real people*. Rather than always appropriate texts to our own personal concerns and particular points of view, we can thus at least aspire to another kind of understanding. As authors of historical accounts we are similarly reminded of our responsibilities in representing others, thus perhaps becoming better aware of the difficulties involved in understanding those who are different to ourselves.[45]

Despite the possible difficulties of understanding other people, the relationship to others prevents linguistic narrativism theory from falling into descriptive moral agnosticism or

[44] See this tension in Sartre, *Existentialism is a Humanism*, 23–26; and Keith Jenkins, 'The End of the Affair: On the Irretrievable Breakdown of History and Ethics', *Rethinking History*, 11:2 (2007), p. 283.
[45] Kalle Pihlainen, 'The Moral of the Historical Story: Textual Differences in Fact and Fiction', *New Literary History*, 33:1 (2002), pp. 56–57 (italics in original).

solipsism because of the normative demands that the existence of other "particular real people" force on us. Even in narrative construction, other people limit our actions and choices by being those for whom we are responsible. This is, indeed, the normativity that we contend proponents of linguistic narrativism theory must more carefully explicate and recognize in order to avoid the claim that the normativity of moral intuition is intuitively instilled on the agent's own performative acts.[46]

Nevertheless, for proponents of the second vision, such as Elizabeth Deeds Ermarth, "ethics" should be understood as a "multiplicity of…semantic systems," or "codes," whereas it seems to lack its normative relation to the other.[47] Therefore, she argues that in the discursive condition "…thought cannot linger long on an 'ethics' of 'good' or 'should' because adequacy involves qualities of enunciation: such as degrees of originality, proportion, flexibility (play), complexity, completeness."[48]

It may be true that moral assessment requires these heterogeneous qualities of the individual. Our argument, however, is that it is doubtful that any meaningful "conception of individual responsibility" survives if normativity is anchored in intuitive individual acts of "enunciation." If "ethics" is re-described as a set of textual "codes," and if "only enunciation adds value" to our preferring one "code" over another, while at the same time our moral compass cannot concern any transcendental "good" or "ought" responding to the demands that the existence of other persons put on us, one would have to conclude that normativity is instilled by the individual herself or ultimately that it is undone as senseless.[49] In the following section, we will consider more closely what moral implications that theoretical route may entail.

4. Difficulties of intuitively instilled normativity

One imperative principle of moral intuitionism is, as Walter Sinnott-Armstrong puts it, the "claim that some moral believers are justified in believing some moral claims independently of whether they are able to infer those moral beliefs from any other beliefs."[50] Moral intuitions are conceived as direct truths requiring no other moral substantiation. Advocates of moral intuitionism often suggest that as a practical matter intuition sufficiently supplies moral principles. In the case of the moral intuitionism attached to linguistic narrativism theory, proponents, such as Foucault (or Ermarth), confine themselves to problematizing the structure of moral justifications through discourse analysis of texts (and sometimes other media, such as paintings) of historical interest. The subjectively problematized history, or "genealogy," then, is reality as it is intuited or seen—and indeed it is very much a reality, a true understanding of the real things in human behavior, especially the cracks in our reasoning that we cover up—but it

[46] Émmanuel Lévinas, *Otherwise than Being, or Beyond Essence*, translated by Alphonso Lingis (Dordrecht: Springer Science & Business Media), p. 10; Anton Froeyman, 'Never the Twain Shall Meet? How Narrativism and Experience can be Reconciled by Dialogical Ethics', *History and Theory*, 54:2 (2015), pp. 162–177.
[47] Ermarth, *History in the Discursive Condition*, pp. 94–95; Ermarth, 'Ethics and Method', pp. 75–76; cf. Barthes, *Empire of Signs*, pp. 63–68.
[48] Ermarth, *History in the Discursive Condition,* p. 95; cf. Ermarth, 'Ethics and Method', p. 76.
[49] All quotes from Ermarth, *History in the Discursive Condition*, pp. 94-95; Ermarth, 'Ethics and Method', pp. 75–76; cf. Sartre, *Existentialism is a Humanism*, pp. 67–69.
[50] Sinnott-Armstrong, *Moral Skepticisms*, pp. 188–189; cf. Moore, *Prinicipia Ethica*, pp. 143–144.

is also generally the circumference of the real that is analyzed.[51] Thus, reality (be it "historical reality" or "morality") cannot on this account be grounded in moral reasoning, but can only be discursively appealed to by a subjectively enacted imagination related to an intuited politics of the present[52]; and so from a philosophical point of view, meta-ethical intuitionism or volitionism is tacitly accepted.

The chief advantage of this idea is that if it is correct it unburdens us of much of the effort to find axiomatic first principles for morality. This approach charms some proponents of linguistic narrativism theory because it helps in eluding the hard problems incident to piloting a normative system between notions of inward and outward, individual and social, rationalist and emotional, and empirical and evaluative.[53] Thus, when moral knowledge is granted validity or esteem because it is emergent from natural knowledge or continuous with it, rather than meaningless, or when one agrees with pure intuitionism on the plain good sense of not using old metaphysical and epistemological conundrums to dismiss the human activity of evaluation, what is allowed is a comfortable way to rely on a supposedly direct relation to the empirical or natural external world—or even an unreflective path to moral truth—for ethical guidance.

What, then, is regarded as instilling normativity in moral intuitions? First, it arises from the idea that intuitive judgments are conceived as quick ones. If we look at what we are told is a painting of a lemon, we can readily agree or deny that it is a lemon because we are familiar enough with lemons to tell at a glance. Were we less familiar with lemons, we would look more closely, maybe use a loop, and perhaps bring along a botanical field guide. Second, what enables this idea of judgements to be consonant with *moral judgements* is the supposition that "common sense" morality has survived because it is trustworthy.[54] Because common sense may be quick, it can provide sufficient directly intuited knowledge; its status as knowledge (justified true belief) is conferred by its longevity and its handy celerity. The "self-evidence" to which moral intuitionists restrict moral knowledge is as instantaneous as perception, unless other conditions, such as consistency or conformity to consensus, are added. However, when consistency or conformity are added to intuitions, the resulting moral claims are in fact no longer grounded on intuition.

Nevertheless, the view that our beliefs about right and wrong do not have any ground other than intuition commits theory that relies on intuition to the position that unless the justification of moral statements is itself intuitive, discussions of morality are undermined by the post-modern anti-foundationalist analysis of discourse.[55] Thus, it is necessary for the defenders to offer this disjunction: either moral knowledge is intuitive, or there is no normativity. However, on this logic, the only supposed justification for normativity would be the normativity with which moral psychology, or common sense, is credited by virtue of its intuitivity.[56] This leads to a syllogistic error: moral knowledge can be intuitive *and* there can be normativity; but these two premises do not add up to a valid conclusion about anything. Nothing, in fact, requires moral knowledge to be intuitive, or

[51] Foucault, 'On the Genealogy of Ethics', pp. 262–269; Ermarth, 'Ethics and Method', pp. 73–78.
[52] Foucault, 'The History of Sexuality', p. 193; cf. Ermarth, 'Ethics and Method', pp. 75–76.
[53] E.g., Ermarth, *History in the Discursive Condition*; Barthes, *Empire of Signs*.
[54] It needs to be noted that "common sense" is also sometimes critiqued for being a morality of prudence. Roland Barthes, *Mythologies*, translated by Annette Lavers (New York: The Noonday Press, 1972), pp. 32–34.
[55] Cf. Ermarth, 'Ethics and Method', pp. 61–83; Ermarth, *History in the Discursive Condition*; cf. Jenkins, 'The End of the Affair', pp. 275–285, Bauman, *Postmodern Ethics*, 28–31.
[56] Sinnott-Armstrong, *Moral Skepticisms*, pp. 187–192.

normativity to be instilled therein. Even if this were the case, it cannot be made a matter of necessity, especially not on the naturalistic syllogism that it must be true because certain pieces of natural (or moral-psychological) knowledge are true.[57]

Thus, from the gap between intuitive knowledge and natural properties arises the naturalist's skeptical position as to moral claims. It is for the consistent naturalist as it is for the linguistic narrativist who appeals to that vision: normativity ceases to "exist" because it is claimed to be unsubstantiated as a natural (factual) "property."[58] Furthermore, the presentation of moral claims is sometimes conceived to be advocacy or rejection of feelings,[59] not facts as to what does and does not suffice in the world when we respond to other people in our lives. The difficulty of this vision is that if proponents of intuitionism claim that "deontic or normative properties" are always "related to reasons" and also reject such reasons and their "properties," it suggests that normativity, even when described as a deontic matter, is framed as (1) a property or object and (2) a particular that should determine what justifies one's moral beliefs.[60] On the contrary, what is unique about the normative relationship we have to other people is not that it is best understood as a property, nor that it should justify one's own actions, nor that our relationship to others requires justification.[61] For one could claim that anyone's and everyone's moral agency, regardless of whether it is intuitive or reasoned, already constitutes a normative relationship with other persons that no skeptical regress can obliterate without then invoking moral nihilism as to other real people and beings.[62]

Prudence supports logic here in rejecting intuitionism: under moral intuitionism, what has already been learned, or is quickly conceived, would have to differ with regard to normativity from what is yet to be learned or more slowly conceived. This is a difficult way to approach moral reflection, because, as Sinnott-Armstrong claims, such intuitive particularism may justify the possibility of "moral nihilism" as a qualified take on morality.[63] The supposedly normative force of the quickness and particularity of moral judgment is thereby an assertion by which moral intuitionism enables an indifference to recognizing the world of laboring, struggling humankind, just because a person is comfortable with herself and her own intuitions, instead of asking herself how her own intuitions possibly disable her to responsibly care for others. Were we to follow this way of moral thinking, not only in accepting immediate understanding but also in hanging onto our received notions, we must be like Prince Oblonsky in *Anna Karenina*, whose:

> ...tendencies and opinions were not his by deliberate choice: they came of themselves, just as he did not choose the fashion of his hats or coats but wore those of the current style. Living in a certain social set, and having a desire, such as generally develops with maturity, for some kind of mental activity, he was obliged to hold views, just as he was obliged to have a hat.... Thus [his views] became habitual to Oblonsky, and he

[57] Cf. Moore, *Principia Ethica*, pp. 10–14; Immanuel Kant, *Groundwork of The Metaphysics of Morals,* in *The Cambridge Edition of the Works of Immanuel Kant: Practical Philosophy*, edited Mary J. Gregor (Cambridge: Cambridge University Press, 1999), pp. 90–91.
[58] In line with naturalism, linguistic narrativism theory argues that fact is determined by a naturalist-scientific worldview. Doran, *The Ethics of Theory*, p. 122.
[59] Leiter, 'Normativity for Naturalists', p. 74.
[60] Leiter, 'Normativity for Naturalists', p. 65; Sinnott-Armstrong, *Moral Skepticisms*, pp. 189–192.
[61] Cf. Kant, *Groundwork of The Metaphysics of Morals,* pp. 86, 90–91.
[62] Cf. Sinnott-Armstrong, *Moral Skepticisms*, p. 167.
[63] Sinnott-Armstrong, *Moral Skepticisms*, p. 191.

> loved his paper as he loved his after-dinner cigar, for the light mistiness it produced in his brain.[64]

The distinction between the most upright and impartialist intuition and the smugness of a rich prince or plutocrat is thinner than one thinks at a quick glance. And it is a fair bet that intuitionism underestimates the scrutiny to which people put their moral opinions or the percentage of people who put their morality under high scrutiny. Whether verbal and educated or not, people feel very deeply, such that they look for the moral good in what they do. Everything will come into question in life as it is actually lived with others.

If the force of our ratiocinative, affective, and even unconscious ways of living morally responsible lives is so evacuated by the scepticist critique that moral deliberation and normativity are deflated to intuition, then we are mistakenly purging some of the visions, ideas, and stories that enable us to be moral agents and to renew that agency in the first place. While critique valuably leads us to de-stabilize and complicate the stories and discourses comprising our "moral visions", we must necessarily rethink those human relationships and discursive reflections on life that enable us to have a "*moral vision*" at all. Even if intellect complicates and might erase the meaningfulness of narrative, even if this erasure and uncertainty of meaningfulness is our intuition, we must ask whether that intuition is true to the other beings we answer to. That moral understanding requires reflective thought about our responsibilities through experiential and deliberative temporal depth of our "moral vision" where our relationship with others is a presupposition and not itself a choice.

The proponent of linguistic narrativism theory may well readily admit this. However, if she nonetheless relies on intuition and nonetheless does not return to that meaningfulness of actions that grows quite separately from personal will and from the automatism of materially willed or desired actions, then she has not understood the moral or immoral meaning of her own "moral vision" implicated in the theory she pursues. In order to retake responsibility, as linguistic narrativism theory claims to do[65], she must consider the moral responsiveness of moving beyond volition and/or intuition in the light of what it normatively means to be with and to care for other persons.[66] Without this self-reflection, the proponent of linguistic narrativism theory will, again, end up in the morally agnostic loop of re-confirming the correctness and goodness of her own intuitions on her own intuitions.

5. Conclusion

Whether or not one regards moral volitionism and moral intuitionism as two different roads that linguistic narrativism theory can take, its proponents must still engage the really hard problem of aligning a forceful and humane way to understand how the theory affects people and their relationships to one another. In other words, even if it is claimed that moral volition and intuition should be able to lead proponents of linguistic narrativism theory back to reference to "reality,"[67] the position does not escape ethical evaluation and

[64] Leo Tolstoy, *Anna Karenina*, translated by L. and A. Maude (New York: Knopf, 1999), pp. 7–8.
[65] Ermarth, 'Ethics and Method', pp. 75–76.
[66] Cf. Murdoch, 'Vision and Choice in Morality', pp. 87–89.
[67] Ermarth, 'Ethics and Method', pp. 75–76; Barthes, 'Historical Discourse', pp. 153–155.

reproach. Our analysis shows that the position clearly consolidates, rather than liquidates, the gap between facts and values and invites moral skepticism. As a "moral vision,"[68] it leaves not only the agent's personal sentiments and choices, but also her normative commitments, up to individual will and intuition, on the epistemic presumption of an empirically unavailable "reality" and on the ethical presumption of the unavoidable subjectivity and freedom of judgment.[69]

However, if what makes theoretical description different from an index of facts or causes or from a fairy-tale is that it helps us step into intimate, ethical connections with other persons, then these people themselves cannot be the product of subjective will or intuition. Pihlainen suggests that if theory-based accounts are not to become "entertainment" or to stay trapped in a "dead in the water fact-fiction debate," one must turn to "experientiality and emotional impact" in order to envision right action.[70]

In other words, one may reasonably hold that non-referentiality and non-realism makes exploration of experience and emotion in moral life possible, but only with a certain normative preunderstanding. As Georges Didi-Huberman has argued (in a line of thought descending chiefly from Aby Warburg and Walter Benjamin), not only do personal accounts and the testimony of memory not need "a clearly visible...referent," but also the very absence of the referent in a verbal or even visual account of past experience can enhance its affective power.[71] Passing-away and loss outline the notion of the past, and what is represented must be at the least what is no longer unequivocally intelligible through reference.[72] The moral philosophical insight, however, is that even if the referent is fragmentary, there is no theoretical reason for us to deviate from the normative preunderstanding that even in theoretical reflection we relate to "particular real people."[73]

In this investigation, we have contended that linguistic narrativism theory needs to rethink the moral implications of its presuppositions and claims in order not to justify morally agnostic "moral visions." Relying on Murdoch's philosophical insight that theoretical descriptions involve degrees of ethical evaluation and express a "moral vision," we have contended that scholars who forward theoretical perspectives, such as linguistic narrativism theory, in the end cannot elide relationships to other real people. Our normative contention is that in so far as the proponents of linguistic narrativism theory want to think of their theory as a morally responsible one, they must clearly recognize the moral implications of the theory and take responsibility for those implications, accordingly to the kind of "moral vision" their theory professes. Exactly as there is never a result in consciousness that floats entirely away from reflection, like a helium balloon flying up from earth when its ropes are cut, there is no theoretical inquiry about ethics and responsibility without the face of the other.

[68] Murdoch, 'Vision and Choice in Morality', pp. 82–87.
[69] Doran, *The Ethics of Theory*, p. 119; Ermarth, 'Ethics and Method', p. 79; White, 'Historical Pluralism and Pantextualism', pp. 230–236; see also White, 'The Public Relevance of Historical Studies', p. 338.
[70] Pihlainen, *The Work of History*, pp. 29–30.
[71] Georges Didi-Huberman, *Bark*, translated by Samuel E. Martin (Cambridge, MA: MIT Press. 2017), p. 81.
[72] Edith Wyschogrod, *The Ethics of Remembering* (Chicago: University of Chicago Press, 1998), pp. 20–22.
[73] Pihlainen, 'The Moral of the Historical Story', p. 56; Bennett Gilbert, *A Personalist Philosophy of History* (New York: Routledge, 2019).

Natan Elgabsi and Bennett Gilbert
natan.elgabsi@abo.fi, bbg2@pdx.edu

Bibliography

Anscombe, G.E.M. 'On Brute Facts', in *Ethics, Religion and Politics: Collected Philosophical Papers Vol. III*. Oxford: Basil Blackwell, 1981, pp. 22–25.

Barthes, Roland. *Empire of Signs*. Translated by Richard Howard. New York: Hill and Wang, 1982

Barthes, Roland. 'Historical Discourse', in *Structuralism: A Reader*, edited by Michael Lane, translated by Peter Wexler. London: Jonathan Cape, 1970, pp. 145–155.

Barthes, Roland. *Mythologies*. Translated by Annette Lavers. New York: The Noonday Press, 1972.

Bauman, Zygmunt. *Postmodern Ethics*. Oxford: Blackwell, 1995.

Brewer, Talbot. *The Retrieval of Ethics*. Oxford: Oxford University Press, 2011.

Carr, David. *Experience and History: Phenomenological Perspectives on the Historical World.* Oxford: Oxford University Press, 2014.

Didi-Huberman, Georges. *Bark.* Translated by Samuel E. Martin. Cambridge, Mass.: MIT Press, 2017.

Doran, Robert. *The Ethics of Theory: Philosophy, History, Literature*. London: Bloomsbury Publishing, 2017.

Dupré, John. 'Fact and Value', in *Value-Free Science? Ideals and Illusions*, edited by Harold Kincaid, John Dupré and Alison Wylie. Oxford: Oxford University Press, 2007, pp. 27–41.

Engel, Pascal. 'Volition and Voluntarism About Belief', in *Belief, Cognition and the Will*, edited by A. Meijers. Tilburg: Tilburg University Press, 1999, pp. 9–25.

Ermarth, Elizabeth Deeds. 'Ethics and Method', *History and Theory* 43:4 (2004), pp. 61-83.

Ermarth, Elizabeth Deeds. *History in the Discursive Condition: Reconsidering the Tools of Thought*. New York: Routledge, 2011.

Ermarth, Elizabeth Deeds. *Sequel to History: Postmodernism and the Crisis of Representational Time.* Princeton: Princeton University Press, 1992.

Foucault, Michel. 'On the Genealogy of Ethics: An Overview of Work in Progress', in *The Essential Works of Michel Foucault 1954-1984: Vol I*, edited by Paul Rabinow. New York: The New Press, 1997, pp. 253–280.

Foucault, Michel. 'The History of Sexuality', in *Power/Knowledge: Selected Interviews and Other Writings 1972-1977*, edited by Colin Gordon. Brighton: The Harvester Press, 1980, pp. 183–193.

Froeyman, Anton. 'Never the Twain Shall Meet? How Narrativism and Experience can be Reconciled by Dialogical Ethics', *History and Theory* 54:2 (2015), pp. 162–177.

Gilbert, Bennett. *A Personalist Philosophy of History*. New York: Routledge, 2019.

Hume, David. *A Treatise of Human Nature: Being and Attempt to Introduce the Experimental Method of Reasoning into Moral Subjects: Book I-III*, edited by L.A. Selby-Bigge. Oxford: Clarendon Press, 1881.

Jenkins, Keith. 'The End of the Affair: On the Irretrievable Breakdown of History and Ethics', *Rethinking History* 11:2 (2007), pp. 275–285.

Kant, Immanuel. *Groundwork of The Metaphysics of Morals*, in: *The Cambridge Edition of the Works of Immanuel Kant: Practical Philosophy*, edited by Mary J. Gregor. Cambridge: Cambridge University Press, 1999, pp. 37–108.

Kronqvist, Camilla. 'Westermarck and Moore on the Sources of Morality', in *Evolution, Human Behaviour and Morality: The Legacy of Westermarck*, edited by Olli Lagerspetz, Jan Antfolk, Ylva Gustafsson and Camilla Kronqvist. New York: Routledge, 2017, pp. 129–141.

Leiter, Brian. 'Normativity for Naturalists', *Philosophical Issues: A Supplement to Noûs* 25:1 (2015), pp. 47–79.

Lévinas, Émmanuel. *Otherwise than Being, or Beyond Essence*. Translated by Alphonso Lingis. Dordrecht: Springer Science & Business Media, 1991.

Moore, George Edward. *Principia Ethica*. Cambridge: Cambridge University Press, 1922.

Murdoch, Iris. 'Metaphysics and Ethics', in *Existentialists and Mystics: Writings on Philosophy and Literature*, edited by Peter Conradi. New York: Penguin Books, 1997, pp. 59–75.

Murdoch, Iris. 'Vision and Choice in Morality', in *Existentialists and Mystics: Writings on Philosophy and Literature*, edited by Peter Conradi. New York: Penguin Books, pp. 76–98.

Murdoch, Iris. *Metaphysics as a Guide to Morals*. London: Vintage, 2003.

Pihlainen, Kalle. 'The Moral of the Historical Story: Textual Differences in Fact and Fiction', *New Literary History* 33:1 (2002), pp. 39–60.

Pihlainen, Kalle. 'The Work of Hayden White II: Defamiliarizing Narrative', in *The SAGE Handbook of Historical Theory*, edited by Nancy Partner and Sarah Foot. Los Angeles: SAGE, 2013, pp. 117–135.

Pihlainen, Kalle. *The Work of History: Constructivism and a Politics of the Past.* London: Routledge, 2017.

Ricoeur, Paul. *Time and Narrative.* Volume 1. Translated by Kathleen McLaughlin and David Pellauer. Chicago: University of Chicago Press, 1984.

Ricoeur, Paul. *Time and Narrative.* Volume 3. Translated by Kathleen Blamey and David Pellauer. Chicago: University of Chicago Press, 1985.

Sartre, Jean-Paul. *Existentialism is a Humanism*. Translated by Carol Macomber. New Haven: Yale University Press, 2007.

Sinnott-Armstrong, Walter. *Moral Skepticisms*. Oxford: Oxford University Press, 2006.

Tolstoy, Leo. *Anna Karenina*. Translated by L. and A. Maude. New York: Knopf, 1992.

White, Hayden. 'Historical Pluralism and Pantextualism', in *The Fiction of Narrative: Essays on History, Literature, and Theory 1957-2007*, edited by Robert Doran. Baltimore: The Johns Hopkins University Press, 2010, pp. 223–236.

White, Hayden. *The Practical Past*. Evanston, IL: Northwestern University Press, 2014.

White, Hayden. 'The Public Relevance of Historical Studies: A Reply to Dirk Moses', *History and Theory* 44:3 (2005), pp. 333–338.

Williams, Bernard. *Ethics and the Limits of Philosophy*. Cambridge, Mass.: Harvard University Press, 2006.

Wyschogrod, Edith. *The Ethics of Remembering.* Chicago: University of Chicago Press, 1998.

Zeleňák, Eugen. 'Who Should Characterize the Nature of History? The Wrong Question?', *Storia della Storiografia* 59–60 (2011), pp. 171–182.