Conceptualizing Policy in Value Sensitive Design:

A Machine Ethics Approach

Steven Umbrello

*Institute for Ethics and Emerging Technologies; University of Turin, Italy*

**Abstract**

*The value sensitive design (VSD) approach to designing emerging technologies for human values is taken as the object of study in this chapter. VSD has traditionally been conceptualized as another type of technology or instrumentally as a tool. The various parts of VSD’s principled approach would then aim to discern the various policy requirements that any given technological artifact under consideration would implicate. Yet, little to no consideration has been given to how laws, policies and social norms engage within VSD practices. Similarly, how the interactive nature of the VSD approach can, in turn, influence those directives. This is exacerbated when considering machine ethics policy that has global consequences outside their development spheres. This chapter begins with the VSD approach and aims to determine how policies come to influence how values can be managed within VSD practices. It shows that the interactional nature of VSD permits and encourages existing policies to be integrated early on and throughout the design process.*

Keywords: value sensitive design, VSD, value tradeoffs, governance, policy, policy innovation, design psychology, applied ethics

**AKNOWLEDGMENTS**

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors. All remaining errors are the author’s alone. The views expressed are not necessarily those of the Institute for Ethics and Emerging Technologies. The content of this publication has not been approved by the United Nations and does not reflect the views of the United Nations or its officials or Member States.

**REFERENCES**

Barrett, A. M., & Baum, S. D. (2017). Risk Analysis and Risk Management for the Artificial Superintelligence Research and Development Process. In V. Callaghan, J. Miller, R. Yampolskiy, & S. Armstrong (Eds.), *The Technological Singularity: Managing the Journey* (pp. 127–140). Berlin, Heidelberg: Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-662-54033-6\_6

Baum, S. D. (2014). The great downside dilemma for risky emerging technologies. *Physica Scripta*, *89*(12), 10. https://doi.org/10.1088/0031-8949/89/12/128004

Bay, C. (1961). The Structure of Freedom. *Science and Society*, *25*(1).

Ben-Haim, Y., Osteen, C. D., & Moffitt, L. J. (2013). Policy dilemma of innovation: An info-gap approach. *Ecological Economics*, *85*, 130–138. https://doi.org/10.1016/j.ecolecon.2012.08.011

Bødker, K., Kensing, F., & Simonsen, J. (2009). *Participatory IT design: Designing for business and workplace realities*. MIT press.

Borning, A., & Muller, M. (2012). Next steps for value sensitive design. *Proceedings of the 2012 ACM Annual Conference on Human Factors in Computing Systems - CHI ’12*, 1125. https://doi.org/10.1145/2207676.2208560

Bos, C., Walhout, B., Peine, A., & van Lente, H. (2014). Steering with big words: articulating ideographs in research programs. *Journal of Responsible Innovation*, *1*(2), 151–170. https://doi.org/10.1080/23299460.2014.922732

Busch, D. (2016). MiFID II: regulating high frequency trading, other forms of algorithmic trading and direct electronic market access. *Law and Financial Markets Review*, *10*(2), 72–82.

Corak, M. (2016). ‘Inequality is the root of social evil,’or Maybe Not? Two Stories about Inequality and Public Policy. *Canadian Public Policy*, *42*(4), 367–414.

Correljé, A., Cuppen, E., Dignum, M., Pesch, U., & Taebi, B. (2015). Responsible Innovation in Energy Projects: Values in the Design of Technologies, Institutions and Stakeholder Interactions. In B.-J. Koops, I. Oosterlaken, H. Romijn, T. Swierstra, & J. van den Hoven (Eds.), *Responsible Innovation 2* (pp. 183–200). Springer International Publishing. https://link.springer.com/chapter/10.1007%2F978-3-319-17308-5\_10

Davis, J., & Nathan, L. P. (2014). Value Sensitive Design: Applications, Adaptations, and Critiques. In J. van den Hoven, P. E. Vermaas, & I. van de Poel (Eds.), *Handbook of Ethics, Values, and Technological Design: Sources, Theory, Values and Application Domains* (pp. 1–26). Dordrecht: Springer Netherlands. https://doi.org/10.1007/978-94-007-6994-6\_3-1

Dignum, V. (2019). *Responsible artificial intelligence: How to develop and use AI in a responsible way*. Springer International Publishing. https://doi.org/10.1007/978-3-030-30371-6

Ehn, P. (2016). Design, Democracy and Work: Exploring the Scandinavian Participatory Design Tradition.

Fallan, K. (2015). Our Common Future. Joining Forces for Histories of Sustainable Design. *TECNOSCIENZA: Italian Journal of Science & Technology Studies*, *5*(2), 15–32. http://www.tecnoscienza.net/index.php/tsj/article/view/201%5Cnhttp://www.tecnoscienza.net/index.php/tsj/article/download/201/136%5Cnhttp://www.tecnoscienza.net/index.php/tsj/article/view/201/136?utm\_source=non+iscritti+a+TS&utm\_campaign=e868895d7d-newissu

Floridi, L. (2016). On human dignity as a foundation for the right to privacy. *Philosophy & Technology*, *29*(4), 307–312.

Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., … Vayena, E. (2018). AI4People - An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations. *Minds and Machines*, *28*(December), 1–24. https://doi.org/10.12932/AP0443.32.4.2014

Floridi, L., Cowls, J., King, T. C., & Taddeo, M. (2020). Designing AI for Social Good: Seven Essential Factors. *Science and Engineering Ethics*, 1–26. https://doi.org/10.1007/s11948-020-00213-5

Friedman, B., & Grudin, J. (1998). Trust and accountability: preserving human values in interactional experience. In *CHI 98 conference summary on Human factors in computing systems* (p. 213). ACM.

Friedman, B., & Hendry, D. G. (2019). *Value sensitive design: Shaping technology with moral imagination*. Cambridge, MA: MIT Press.

Friedman, B., Hendry, D. G., & Borning, A. (2017). A Survey of Value Sensitive Design Methods. *Foundations and Trends® in Human–Computer Interaction*, *11*(2), 63–125. https://doi.org/10.1561/1100000015

Friedman, B., & Kahn Jr., P. H. (2002). Value sensitive design: Theory and methods. *University of Washington Technical*, (December), 1–8. https://doi.org/10.1016/j.neuropharm.2007.08.009

Friedman, B., Kahn Jr., P. H., & Borning, A. (2008). Value Sensitive Design and Information Systems. *Human-Computer Interaction and Management Information Systems: Foundations*, 69–101. https://doi.org/10.1145/242485.242493

Friedman, B., Kahn, P. H., Borning, A., & Huldtgren, A. (2013). Value Sensitive Design and Information Systems. In N. Doorn, D. Schuurbiers, I. van de Poel, & M. E. Gorman (Eds.), *Early engagement and new technologies: Opening up the laboratory* (pp. 55–95). Dordrecht: Springer Netherlands. https://doi.org/10.1007/978-94-007-7844-3\_4

Gazzaneo, L., Padovano, A., & Umbrello, S. (2020). Designing Smart Operator 4.0 for Human Values: A Value Sensitive Design Approach. In *International Conference on Industry 4.0 and Smart Manufacturing (ISM 2019) in Procedia Manufacturing* (pp. 219–226). Rende, CS: Elsevier. https://doi.org/10.1016/j.promfg.2020.02.073

Gregor, P., Sloan, D., & Newell, A. F. (2005). Disability and Technology: Building Barriers or Creating Opportunities? *Advances in Computers*. https://doi.org/10.1016/S0065-2458(04)64007-1

Hyppönen, H., Kemppainen, E., Gill, J., Slater, J., & Poulson, D. (2000). *Handbook on Inclusive Design of Telematics Applications.* (H. Hyppönen, E. Kemppainen, J. Gill, J. Slater, & D. Poulson, Eds.). Themes/STAKES 2.

Ingram, H., deLeon, P., & Schneider, A. (2016). Conclusion: Public policy theory and democracy: The elephant in the corner. In *Contemporary Approaches to Public Policy* (pp. 175–200). Springer.

Lockton, D., Harrison, D., & Stanton, N. A. (2016). Design for Sustainable Behaviour: investigating design methods for influencing user behaviour. *Annual Review of Policy Design*, *4*(1), 1–10.

McConnell, A. (2010). *Understanding policy success: Rethinking public policy*. Macmillan International Higher Education.

Mintrom, M., & Luetjens, J. (2017). Creating public value: Tightening connections between policy design and public management. *Policy Studies Journal*, *45*(1), 170–190.

Muehlhauser, L., & Bostrom, N. (2014). Why We Need Friendly AI. *Think*, *13*(36), 41–47. https://doi.org/10.1017/S1477175613000316

Müller, V. C., & Bostrom, N. (2016). Future Progress in Artificial Intelligence: A Survey of Expert Opinion BT - Fundamental Issues of Artificial Intelligence. In V. C. Müller (Ed.) (pp. 555–572). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-26485-1\_33

Nissenbaum, H. (2009). *Privacy in context: Technology, policy, and the integrity of social life*. Stanford University Press.

Peters, B. G. (2018). *Policy problems and policy design*. Edward Elgar Publishing.

Quah, E., & Mishan, E. J. (2007). *Cost-benefit analysis*. Routledge.

Rawls, J. (2001). *Justice as fairness: A restatement*. Harvard University Press.

Ruzic, L., & Sanfod, J. A. (2017). Universal Design Mobile Interface Guidelines (UDMIG) for an Aging Population. In *Mobile e-Health* (pp. 17–37). Springer.

Santoni de Sio, F., & van den Hoven, J. (2018). Meaningful Human Control over Autonomous Systems: A Philosophical Account. *Frontiers in Robotics and AI*. https://www.frontiersin.org/article/10.3389/frobt.2018.00015

Scheve, K., & Stasavage, D. (2017). Wealth inequality and democracy. *Annual Review of Political Science*, *20*, 451–468.

Sieber, S. (2013). *Fatal remedies: The ironies of social intervention*. Springer Science & Business Media.

Simon, H. A. (1988). The science of design: Creating the artificial. *Design Issues*, 67–82.

Soares, N., & Fallenstein, B. (2014). *Agent Foundations for Aligning Machine Intelligence with Human Interests : A Technical Research Agenda*. https://doi.org/10.1007/978-3-662-54033-6\_5

Solove, D. J. (2008). *Understanding privacy* (Vol. 173). Harvard university press Cambridge, MA.

Sunstein, C. R. (2013). The value of a statistical life: some clarifications and puzzles. *Journal of Benefit-Cost Analysis*, *4*(2), 237–261.

Umbrello, S. (2018a). *Safe-(for whom?)-by-Design: Adopting a Posthumanist Ethics for Technology Design*. York University. https://doi.org/10.13140/RG.2.2.29726.38720

Umbrello, S. (2018b). The moral psychology of value sensitive design: the methodological issues of moral intuitions for responsible innovation. *Journal of Responsible Innovation*, *5*(2), 186–200. https://doi.org/10.1080/23299460.2018.1457401

Umbrello, S. (2019a). Atomically Precise Manufacturing and Responsible Innovation: A Value Sensitive Design Approach to Explorative Nanophilosophy. *International Journal of Technoethics*, *10*(2), 1–21. https://doi.org/10.4018/IJT.2019070101

Umbrello, S. (2019b). Beneficial Artificial Intelligence Coordination by Means of a Value Sensitive Design Approach. *Big Data and Cognitive Computing*, *3*(1), 5. https://doi.org/10.3390/bdcc3010005

Umbrello, S. (2020a). Imaginative Value Sensitive Design: Using Moral Imagination Theory to Inform Responsible Technology Design. *Science and Engineering Ethics*, *26*(2), 575–595. https://doi.org/10.1007/s11948-019-00104-4

Umbrello, S. (2020b). Meaningful Human Control Over Smart Home Systems: A Value Sensitive Design Approach. *HUMANA.MENTE Journal of Philosophical Studies*, *13*(37), 40-65. https://www.humanamente.eu/index.php/HM/article/view/315

Umbrello, S., & Baum, S. D. (2018). Evaluating future nanotechnology: The net societal impacts of atomically precise manufacturing. *Futures*, *100*(June), 63–73. https://doi.org/10.1016/j.futures.2018.04.007

Umbrello, S., & De Bellis, A. F. (2018). A Value-Sensitive Design Approach to Intelligent Agents. In R. V. Yampolskiy (Ed.), *Artificial Intelligence Safety and Security* (pp. 395–410). CRC Press. https://doi.org/10.13140/RG.2.2.17162.77762

Umbrello, S., & van de Poel, I. (2020). *Mapping Value Sensitive Design onto AI for Social Good Principles*. https://www.academia.edu/43347384/Mapping\_Value\_Sensitive\_Design\_onto\_AI\_for\_Social\_Good\_Principles

United Nations. (2017). *Data Privacy, Ethics and Protection Guidance Note on Big Data For Achievement of the 2030 Agenda.* https://unsdg.un.org/resources/data-privacy-ethics-and-protection-guidance-note-big-data-achievement-2030-agenda

United Nations. (2019a). Democracy & the SDGs. https://sustainabledevelopment.un.org/index.php?page=view&type=20000&nr=5780&menu=2993

United Nations. (2019b). Sustainable development goals. *GAIA*. https://doi.org/10.14512/gaia.28.2.1

van de Kaa, G., Rezaei, J., Taebi, B., van de Poel, I., & Kizhakenath, A. (2019). How to Weigh Values in Value Sensitive Design: A Best Worst Method Approach for the Case of Smart Metering. *Science and Engineering Ethics*. https://doi.org/10.1007/s11948-019-00105-3

van de Poel, I. (2014). Conflicting Values in Design. In J. van den Hoven, P. E. Vermaas, & I. van de Poel (Eds.), *Handbook of Ethics, Values, and Technological Design: Sources, Theory, Values and Application Domains* (pp. 1–23). Dordrecht: Springer Netherlands. https://doi.org/10.1007/978-94-007-6994-6\_5-1

van den Hoven, J. (2013). Architecture and Value-Sensitive Design. In C. Basta & S. Moroni (Eds.), *Ethics, design and planning of the built environment* (p. 224). Springer Science & Business Media. https://books.google.ca/books?id=VVM\_AAAAQBAJ&dq=moral+value+such+as+freedom,+equality,+trust,+autonomy+or+privacy+justice+%5Bthat%5D+is+facilitated+or+constrained+by+technology&source=gbs\_navlinks\_s

van den Hoven, J. (2017). The Design Turn in Applied Ethics. In J. van den Hoven, S. Miller, & T. Pogge (Eds.), *Designing in Ethics* (pp. 11–31). Cambridge, UK: Cambridge University Press. https://doi.org/10.1017/9780511844317

van den Hoven, J., Lokhorst, G. J., & van de Poel, I. (2012). Engineering and the Problem of Moral Overload. *Science and Engineering Ethics*, *18*(1), 143–155. https://doi.org/10.1007/s11948-011-9277-z

van den Hoven, J., & Manders-Huits, N. (2009). Value-Sensitive Design. In *A Companion to the Philosophy of Technology* (pp. 477–480). Wiley-Blackwell. https://doi.org/10.1002/9781444310795.ch86

van Wynsberghe, A. (2013). Designing Robots for Care: Care Centered Value-Sensitive Design. *Science and Engineering Ethics*, *19*(2), 407–433. https://doi.org/10.1007/s11948-011-9343-6

van Wynsberghe, A., & Robbins, S. (2014). Ethicist as Designer: A Pragmatic Approach to Ethics in the Lab. *Science and Engineering Ethics*, *20*(4), 947–961. https://doi.org/10.1007/s11948-013-9498-4

Warnier, M., Dechesne, F., & Brazier, F. (2014). Design for the Value of Privacy. In J. van den Hoven, P. E. Vermaas, & I. van de Poel (Eds.), *Handbook of Ethics, Values, and Technological Design: Sources, Theory, Values and Application Domains* (pp. 1–14). Dordrecht: Springer Netherlands. https://doi.org/10.1007/978-94-007-6994-6\_17-1

Winkler, T., & Spiekermann, S. (2019). Human Values as the Basis for Sustainable Information System Design. *IEEE Technology and Society Magazine*, *38*(3), 34–43. https://doi.org/10.1109/MTS.2019.2930268

Woelfer, J. P., Iverson, A., Hendry, D. G., Friedman, B., & Gill, B. T. (2011). Improving the Safety of Homeless Young People with Mobile Phones: Values, Form and Function. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 1707–1716). New York, NY, USA: ACM. https://doi.org/10.1145/1978942.1979191

Woolgar, S. (1991). The Turn to Technology in Social Studies of Science. *Science, Technology, & Human Values*, *16*(1), 20–50. https://doi.org/10.1177/016224399101600102

Yoo, D., Derthick, K., Ghassemian, S., Hakizimana, J., Gill, B., & Friedman, B. (2016). Multi-lifespan design thinking: Two methods and a case study with the Rwandan diaspora. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (pp. 4423–4434). ACM.