

Value sensitive architecture

Authors

Marlies van Steenbergen, marlies.van.steenbergen@sogeti.com,
Ton Eusterbrock, ton.eusterbrock@sogeti.com,
Hans Nouwens, hans.nouwens@sogeti.com,
Edzo A. Botjes, edzo.botjes@sogeti.com

Enterprise Architecture, Sogeti Netherlands

Editors

Lisette Atsma, Mariëtte Draaisma



Value sensitive architecture

– a DYA white paper by Sogeti –

Marlies van Steenberg, marlies.van.steenbergen@sogeti.com
Ton Eusterbrock, ton.eusterbrock@sogeti.com,
Hans Nouwens, hans.nouwens@sogeti.com,
Edzo A. Botjes, ebotjes@xebia.com
from the departments of
Enterprise Architecture, Sogeti Netherlands & Xebia

I. Value sensitive architecture

Abstract—Every action is a moral choice. Every design constitutes a subset of all theoretically available possibilities. This also holds for the design of a digital solution. A design always defines the behavioural choices of its users in some way: discouraging some behaviour and stimulating other behaviour. Technology cannot be neutral.

Value sensitive architecture is not only aware of the economical aspects of enterprise architecture design choices, but also of their impact on personal and public values.

In this whitepaper we introduce the main lenses to look at ethics and provide examples of the moral impact of design choices. We discuss how, by applying value sensitive design and an ethical matrix, you can take human values into account during the design process. At the end we place ethical thinking, sensemaking and enterprise architecture into relation to each other.

II. Everything is a moral choice

A. The Cookie dilemma

Recently, I spoke with a new colleague who is preparing her Master thesis research. Her topic is how websites deal with asking permission from their visitors to place Cookies. She is particularly interested in the question how informed a given consent from a website visitor really is, or can be. She told me that studies have been done that reveal that the colour of buttons is decisive in nudging persons into giving consent. I could immediately relate to this topic, from personal experience, and I am sure I am not the only one.

When surfing on the Internet I am torn between meticulously reading the privacy statements and trying to adjust my settings concerning the permissions for cookies, or just clicking on the OK button and get on with it. Most of the time it feels I have no real choice. The diversity in the ways in which permission for placing Cookies is asked, is huge and varies from extremely annoying to moderately user-friendly. To me, the way in which websites handle the required request for permission regarding Cookies, is indicative of the way in which organisations apparently value their customers. The organisations that unobtrusively nudge the visitors of their sites into accepting all possible Cookies, to me, seem to value their customers differently from the ones that make it very

easy for visitors to make a fast but considerate choice that suits the visitor's own preferences and values.

B. Moral choices

Everything is a moral choice. We all recognise the obvious moral choices, where the issues of right or wrong stare us in the face: whether we help someone in need, whether we harm someone for our personal gain, whether we break our promises. Of course, not all moral choices are easy. Do we need to rescue someone if it endangers our own life? Does it make a difference whether the person in need is harmful to others or not, and whether I am twenty years of age or eighty? Complicating things further, we must realise that not all moral choices stare us in the face. Sometimes, a keen awareness is required to realise that one is dealing with a moral issue. For instance, because the consequences of a choice are distant in place or time. Sometimes people are genuinely not aware that they are dealing with a moral issue. And sometimes the system may induce people to consciously or unconsciously close their eyes to moral consequences, as happened in the financial crisis of 2008.

Moral choices are of all times. And technological innovations have always evoked moral discussion. For instance about their potential impact on jobs. With the rise of data-driven technology new moral discussions appear. About bias in Machine Learning, about a digital divide leading to exclusion, about threats to autonomy, about breaches of privacy and lack of transparency, even about threats to democracy.

Moral consequences are not always straightforwardly visible. Knowing this, obliges us to take the time to deeply reflect on the choices we make. A well-known example are the overpasses to Long Island in New York. The design of the overpasses was so low that busses could not pass under them. This hindered poor people from travelling to Long Island. Was this intentionally? Some people think it was, others take a more nuanced view (Joerges, 1999). Either way, the effect was one of exclusion of certain people from travelling to a desirable part of the city.

III. Ethics: thinking about what is right or wrong

It is not always easy to do the right thing, to know what the right thing to do would be, or even to recognise that

Choice architecture, nudging and affordances

An important concept when talking about the impact of our design on individuals or society, is the concept of choice architecture, and related to that, the concepts of nudging, seducing people to act in a certain manner and affordances^a, the perceivable action possibilities of an object or situation (van den Hoven, 2017).

Every design constitutes in a subset of all theoretically available possibilities. This also holds for the design of a digital solution. A design always limits the behavioural choices of its users in some way. Thaler and Sunstein (2009) call this a ‘choice architecture’. A choice architecture structures the dimensions of freedom for individuals. It determines the affordances and restrictions of the environment. Also, it can stimulate behaviour by intentionally making certain choices easier than others, ‘nudging’ one towards a particular choice.

A nudge stimulates people to voluntarily exhibit certain behaviour without forcing them. People can ignore a nudge without repercussions of any kind. However, it is made easy for them to, consciously or unconsciously, follow the nudge. Making use of nudges has a paternalistic flavour. Indeed, Thaler and Sunstein argue that this is not a bad thing, but is acceptable for the common good.

^a<https://www.merriam-webster.com/dictionary/affordance>

a question of right or wrong is at stake. For centuries philosophers have thought and debated about what is morally just. Various ethical schools emerged that differ in their perspective on what constitutes “doing the right thing”. Three main streams of ethical thinking are consequentialism, deontology and virtue ethics. These streams differ in the perspective from which they consider the question of what is right.

A. Consequentialism

The consequentialists look upon what is right from the perspective of the consequences of an act. Whether an act is right or wrong depends solely on its consequences. The best-known school of consequentialism is utilitarianism, proposed by Jeremy Bentham¹ (1748-1832) and John Stuart Mill² (1806-1873). Utilitarianism claims that the right act is the act that brings most happiness in the world, i.e. the act where the total amount of happiness generated minus the total amount of pain generated is higher than with any other possible act.

Let’s use the infamous trolley problem, introduced by Philippa Foot³ in the sixties, to illustrate this vision (Foot, 1967; Thomson, 1985). Imagine the situation where a trolley has gone haywire and threatens to kill five persons

¹<https://plato.stanford.edu/entries/bentham>

²<https://plato.stanford.edu/entries/mill>

³<https://tudelft.openresearch.net/page/15836/introduction-to-the-trolley-problem>

Examples of technological choices that require moral deliberation are many.

Dynamic pricing

Big data allows for dynamic pricing. Prices may go up the more interest you show in a particular purchase. Prices may go up in times of scarcity. Prices may go down following the prices of competitors. Prices may go down if you have a healthy lifestyle, are a careful driver or live in a secure neighbourhood.

Are there limits to what is acceptable in the application of dynamic pricing?

The responsible mortgage lender

A mortgage lender wants to develop a Machine Learning algorithm that calculates when customers run a high risk of not being able to pay their mortgage fees in a few months’ time. Customers selected by the algorithm can be approached proactively to try and prevent arrears in payments.

What is acceptable concerning the types of data to be used for this purpose? What types of actions are acceptable to take, based on the data?

Online proctoring

Lately, because of social distancing and travel restrictions, higher education students have been required to take their exams from home, via Internet, using their own laptops. This is the only way to prevent serious study delays. To ensure students do not commit fraud, online proctoring software is used.

This type of software records everything the student does during the exam, through the webcam and through capturing keystrokes. The recordings, including images of the webcam, are analysed by an AI algorithm and when the algorithm detects any anomalies in behaviour a report is made and sent to the examiner, together with the recordings, for further inspection.

Is it acceptable to force students to allow themselves to be filmed, to ask them to arrange the correct circumstances to take the exam, and to subject them to machine scrutiny for potential fraud? How might this affect well-meaning students?

who are working on the track. The only way to prevent this from happening and save the five, is to flip the switch on the track so the trolley takes a side track, on which only one person is at work, offering the life of this one worker for the lives of the other five. What would you do? Utilitarianists would state that the right thing to do is to flip the switch, reasoning that the death of one person brings less pain than the death of five. But how about if the one person is young and healthy and has a loving wife and three children, while the other five are old and without any family? How do we calculate the amount of happiness versus pain? This is not an easy question to answer.

The political philosopher John Rawls⁴ (1921 – 2002) argues that utilitarianism may lead to sacrificing the interests of a minority to the happiness of the majority and introduces the concept of ‘justice as fairness’ to prevent this. Rawls introduces the ‘veil of ignorance’: justice comes when we adhere to those principles of government that rational individuals would agree to when operating from a ‘veil of ignorance’. We operate from a ‘veil of ignorance’ when we have no knowledge about the social group we would belong to, about our own circumstances and capacities, and about our own basic values and goals.

In such a situation rational individuals would adhere to the following two principles: (1) that each person is to have an equal right to the most extensive basic liberty that is compatible with a similar liberty for others; (2) that social and economic inequalities are to be arranged in such a way that they are to the greatest benefit of the least advantaged and are attached to offices and positions open to all under conditions of fair equality of opportunity.

Consequentialists aim to produce the most good. Morally right is the action that will achieve the best consequences. The question they ask themselves is: what kind of outcomes should I try to produce? Utilitarians only take the amount of happiness into account. Other consequentialists may also take other types of consequences into account, such as sustainability, freedom or knowledge.

B. Deontology

The deontological school claims that humans should act according to certain rules. People have duties according to which they should behave. For instance, the duty not to harm another person. Or the duty to adhere to our promises. The most famous deontologist was Immanuel Kant⁵ (1724-1804). Where utilitarians might argue that the end justifies the means, deontologists will not agree with this statement. In their opinion the end never justifies the means. Returning to the trolley problem, a deontologist may argue that one should never actively harm another being and therefore one should not flip the switch, whatever the consequences.

As duties may conflict with each other, though, a deontologist might also argue that in this case flipping is allowed as it is more important to adhere to the duty of saving lives if one possibly can. This same deontologist may draw the line, however, when asked if it is right to push a fat man on the track, if that would be the only way to stop the train and save the group of five. He may sense a difference between flipping a switch and actively murdering an innocent bystander. This immediately begs the question: what are these duties or rules that humans should adhere to? Who defines them? Some central authority such as government or the church? Not according to Kant.

Kant argues that reasonable people can determine by themselves what is right and what is wrong by considering whether they could reasonably wish for their

act to become a general rule for all people. Thus, a reasonable person will not think that it is o.k. to make a promise to someone, for instance to repay a loan, while knowing for sure that they will never keep this promise. For, if making promises while knowing you will break them, becomes the general rule, the entire meaning of the concept of a promise becomes useless. And nobody in their right mind will find that desirable. In other words, what is right or wrong is not determined by government or religion, but derives from reason.

Kant greatly values ‘autonomous agency’: because the rules are dictated by reason you follow your own conscience and you show respect for other such agents. A human should never be regarded as solely a means to an end. Humans are not just ‘resources’, they are an end in themselves.

Jürgen Habermas⁶ (1929) agrees with Kant in that individuals should act according to moral rules derived from reason. However, unlike Kant he does not think that any rational individual will always arrive at these rules by themselves. Habermas argues that these rules should be agreed upon among individuals through discourse. In his so-called ‘discourse ethics’ he formulates the discourse principle that a rule or action is justified only if all affected by the rule or action could accept it in a reasonable discourse. Habermas derives the following dialogical principle of universalization: a moral norm is valid, just in case all the foreseeable consequences and side-effects of its general observance for the interests and value-orientations of each individual, could be jointly accepted by all concerned, of their own free will. Thus, Habermas stresses the importance of persons engaging in moral dialogue, and having the capacity to do so.

Deontologists aim to perform the right action. Ethical conduct always involves doing the right thing: never failing to do one’s duty. The question they ask themselves is: what are my obligations in this situation, and what are the things I should never do?

C. Virtue ethics

The third ethical school to discuss here is virtue ethics. This school looks at the person doing the act, rather than the act itself. A good act is an act that a virtuous, or good, person would do. The name primarily associated with virtue ethics is Aristotle⁷ (384 BC – 322 BC). Like Socrates⁸ and Plato⁹, Aristotle regards virtue central to a well-lived life. Virtues, such as wisdom, justice, courage and temperance (the cardinal virtues), are to be seen as skills that must be trained by repeatedly applying them in various circumstances.

An important concept with Aristotle, is the ‘virtuous mean’. A virtue is a mean state between vices of excess and deficiency. Too much courage becomes the vice of recklessness, too little courage becomes the vice of cowardice. In a similar way, each virtue constitutes a mean state between a vice of excess and a vice of deficiency. Training the virtue of courage does not imply

⁴<https://plato.stanford.edu/entries/rawls>

⁵<https://plato.stanford.edu/entries/kant>

⁶<https://plato.stanford.edu/entries/habermas>

⁷<https://plato.stanford.edu/entries/aristotle>

⁸<https://plato.stanford.edu/entries/socrates>

⁹<https://plato.stanford.edu/entries/plato>

that one tries to gradually move towards recklessness, but that one exercises good judgement in the face of ethical issues and tries to strengthen and deepen one's capacity to display the right amount of courage in each circumstance. This can only be achieved by practice. Only in practice can we gain an understanding of what, in a particular situation, is the right action displaying the right amount of courage for the case.

Becoming virtuous, thus, is not a purely intellectual exercise. Virtues are developed over time through regular practice and through repetition of doing the right actions. In the eyes of Aristotle, ethics cannot be reduced to a mere decision procedure. Crossan et al. (2013) develop this idea into what they call a virtue-based orientation in ethical decision-making, which they define as the capacity to deepen through repeated self-reflection your character strengths along the virtuous mean, while avoiding the vices of excess or deficiency. To develop a virtue-based orientation an explicit step of reflection must be inserted in the cycle of decision making.

Virtue ethics aims to develop character. Ethical conduct is whatever a fully virtuous person would do in the circumstances. The question that they ask themselves is: what kind of person should I be and what will my actions show about my character?

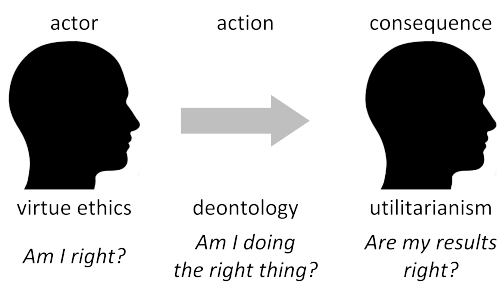


Figure 1: Three schools of ethics.

D. Which ethical school to prefer?

We might ask the question which ethical school provides the best answer to ethical questions. However, this is not a question that can be answered. Each school has its merits. A more useful question is how these ethical perspectives together can help us in addressing ethical issues. They do not provide a decisive ethical decision procedure. They can, however, broaden our repertoire to deal with hard ethical questions (see figure ??).

The consequentialist perspective urges us to think deeply about the potential impact of our choices and actions on others. The deontology perspective urges us to think deeply about the kind of behaviour we are exhibiting ourselves as well as stimulating in others with our choices and actions. The virtue perspective urges us to work hard on the skill of ethical acting in both ourselves and others.

Essential in all of this, is to realise that true ethical dilemmas can only be brought to a satisfying conclusion through dialogue and engaging others. And that sometimes there is no best solution to an ethical dilemma, but that the best we can do is make a choice, after considered involvement of all relevant parties, thorough deliberation and deep contemplation, that we

feel we can explain in good conscience.

Eastern philosophy

The three ethical schools discussed in this paper are exemplars of Western philosophy. Western philosophy is much concerned with abstracting from experience into theorising about how to deal with moral dilemmas. Eastern philosophy (Indian^a, Chinese^b, Japanese^c) is more oriented towards everyday practice. Chinese philosophy texts, for instance, are primarily concerned with 'pretheoretical' experience. Confucius talks about how to act in particular circumstances.

The same ethical question may be answered differently for different persons in different circumstances. From such practical examples various, sometimes mutually exclusive, theories might be developed. The dilemmas arising from this ambiguity are often not conclusively solved. Knowledge can be achieved from learning from exemplary figures. The focus is on relations with others and how people treat each other in the daily life.^d

^a<https://plato.stanford.edu/entries/ethics-indian-buddhism>

^b<https://plato.stanford.edu/entries/ethics-chinese>

^c<https://plato.stanford.edu/entries/japanese-philosophy>

^d<https://plato.stanford.edu/entries/ethics-chinese>

IV. Human values

A. What we find important in our lives

A fundamental concept in ethics is the concept of human value. Human values are what people find important in their lives. They are desirable, trans-situational goals that serve as guiding principles (Schwartz, 2012). Values motivate people to act in a certain manner.

It is important to distinguish human value from 'economic' value. Economic value is attached to a product or service and is ultimately expressed in monetary terms. Human value is not attached to a specific artefact, but is related to how we feel about matters in general, to what in our opinion would be a desirable state of the world. Human values are guiding principles that transcend specific situations. The economic value a person attaches to an object will be partly determined by her human values.

Human values are held by individuals, but values can also be shared within groups of people. In the latter case we might speak for instance of organisational values or public values. Researchers have been compiling value inventories from perspectives such as work-related values (business managers, employees), values related to consumer attitudes, psychology-related values and values related to technology design (Cheng and Fleischmann, 2010).

B. Ethical pluralism

People differ in the values that they find most important. Utilitarians consider the value of happiness or well-being very important. They strive for the highest amount of happiness in the world. Deontologists highly rate the value of autonomy. Humans should never be regarded only as a means, they are an end in themselves, they are autonomous beings.

According to [Crossan et al. \(2013\)](#), the differences in importance that individuals attach to values are reflected in the way they display virtues. Thus, the virtue of wisdom can be displayed in the form of curiosity and independent thought by people who value autonomy, in the form of the pursuit of intellect by people who value self-enhancement, and in the form of understanding and tolerance by people who value universalism and benevolence ([Crossan et al., 2013](#)).

Many philosophers are of the opinion that there are universal values such as well-being or justice that are shared by all humanity. Others, the ethical relativists, oppose this and say that values are, among others, culturally determined. A useful middle way seems to be the stance of ethical pluralism. It assumes that there are indeed values that everyone recognises as such, but that individuals differ in the relative importance they attach to these values, and in their norms of what is acceptable or not regarding violation of these values ([Ess, 2006](#)).

C. Intrinsic versus instrumental values

When considering personal values, a further distinction can be made between intrinsic values and instrumental values ([Rokeach, 1973](#)). Intrinsic values are values that are important to humans in their own right. They are an end in themselves. Instrumental values are values that are important because they contribute to intrinsic values. Well-being or happiness is considered an intrinsic value by many. It serves no other higher purpose, but is something to strive for in itself. If you ask the question 'what is well-being good for?', the answer is that it is simply there as an ultimate goal for mankind ([Spiekermann, 2015](#)).

Examples of instrumental values that may contribute to well-being are convenience or independence. However, we have to take into account that what is considered an intrinsic value might also be culturally determined.

Values are complex concepts in the sense that they usually have many facets. Taking for instance the value of transparency and applying that to the context of knowledge creation and communication, [Spiekermann \(2015\)](#), inspired by [Turilli and Floridi \(2009\)](#), conceptually translates transparency into five information quality criteria: meaningful, veridical, comprehensive, accessible and appropriate. Meaningful means 'conveying a message that has significance for a recipient in a particular context'. If information cannot be interpreted, it is not meaningful.

Comprehensiveness means that 'the information is easy to read and understand'. Data without metadata are not comprehensible. Hundreds of pages with fine print are not comprehensible. Also, it may be necessary to know about how information was created to be able to

Diversity in value frameworks

Many overviews of values have been compiled by researchers over the decades. Below are four examples.

Milton Rokeach identifies 18 intrinsic values: true friendship, mature love, self-respect, happiness, inner harmony, equality, freedom, pleasure, social recognition, wisdom, salvation, family security, national security, a sense of accomplishment, a world of beauty, a world of peace, a comfortable live, an exciting live and 18 instrumental values: cheerfulness, ambition, love, cleanliness, self-control, capability, courage, politeness, honesty, imagination, independence, intellect, broad mindedness, logic, obedience, helpfulness, responsibility, forgiveness (Rokeach, 1973). These values are incorporated in a measurement instrument, the so-called Rokeach Value Survey (Rokeach, 1973).

Shalom Schwartz argued that the values of Rokeach were culturally biased. Schwartz defined a list of 10 motivational value types that are recognised in different cultures (Schwartz, 1992, 2012). Schwartz clusters these ten basic value types into four dimensions: openness to change (stimulation, self-direction), conservation (tradition, conformity, security), self-enhancement (power, achievement, hedonism) and self-transcendence (universalism, benevolence). The ten value types are further elaborated into 56 basic human values.

Friedman and Hendry (2019) list the following values as frequently occurring in system design: human welfare, ownership and property, privacy, freedom from bias, universal usability, trust, autonomy, informed consent, accountability, courtesy, identity, calmness and environmental sustainability.

Pereira and Baranauskas (2014) present an interesting framework for understanding values in software design taking a cultural perspective. Their Value Pie places values in two dimensions. The first dimension is formed by ten so-called 'building blocks of culture'. The second dimension is formed by levels on which humans operate, i.e. informal, formal and technical.

comprehend it.

Accessibility to information is a legal right in Europe, but in practice often requires a lot of effort and stubbornness.

Veridical means truthful, not telling lies. But not telling lies is not enough, because reality can be disguised without telling lies, transparency also needs appropriateness.

Appropriateness implies selecting 'those essential pieces of information that best reflect reality'. This may mean not showing outdated information or information that deflects from what is really going on.

Digitalisation and the capacity or shame

A striking example of the impact of digitalisation on human values, is the financial crisis around 2008. When financial institutions started to fall down, bankers were accused of having enriched themselves shamelessly without regard for the negative consequences to many people and society at large.

Joris Krijger makes the connection between this apparent lack of shame of bankers and the virtual environment they operated in, created by far-reaching digitisation (Krijger, 2016). Krijger derives from philosophy, three necessary conditions for the capacity to experience shame: integrity (Kierkegaard^a), physicality (Merleau-Ponty^b) and responsibility (Sartre^c). Next, he argues that each of these three conditions is diminished by virtualisation. Integrity is concerned with wholeness, being one with your actions, always acting from your inner self.

People that possess integrity experience a personal connectedness with their acts. They cannot hide behind others. Virtualisation, however, enables people to become an anonymous spectator, to live a simulated life without risks or commitments.

Physicality is about being physically present and visible to other people. Shame is connected to this visibility. Virtualisation, however, has removed physicality. We can observe without being observed ourselves.

Personal responsibility means being aware of the fact that one is an autonomous actor and is seen as such by others. Taking personal responsibility means realising that you are a free agent, and thus responsible for your actions. That you cannot hide behind systems or roles. Virtualisation, however, has made it much easier for people to shed the 'burden of freedom' and to blame the complexity of modern systems, with their complicated digital algorithms and machine decisions, for any untoward actions.

The banking sector in the first decade of this century can be characterised by maximum virtualisation. This probably contributed strongly to its actors' inability to feel shame and thus take moral responsibility for the huge negative consequences of their acts. It is another example of the need to consciously and deliberately incorporate moral reflection in our daily work.

^a<https://plato.stanford.edu/entries/kierkegaard>

^b<https://plato.stanford.edu/entries/merleau-ponty>

^c<https://plato.stanford.edu/entries/sartre>

D. Responsible innovation

Ethics has been a topic of contemplation by philosophers for over millennia (Grayling, 2019). Over the last decade, however, we have seen a new surge in discussions in society at large about ethics. Mainly caused by people starting to realise the huge, often unforeseen, impact of digitalisation in all its forms on society. Part of these discussions is the growing call for an explicit ethical approach to digital innovation, or responsible innovation.

Responsible innovation does not only aim at contributing to the general good of mankind, for instance by contributing to the United Nations Sustainable Development Goals¹⁰, but also at taking the values of individuals into account, while doing so (van den Hoven, 2017). This latter aspect, taking into account from the start the values of individuals, not only seems to be the morally right thing to do, but it also has economic value, as people simply do not accept unmotivated or disproportional violations of what they value. Examples of this in the Netherlands are the failure of government to enforce initiatives such as smart meters, electronic health records and a corona tracking-and-tracing app, because government could not convince citizens that their privacy was guaranteed by the design and implementations of these initiatives (van den Hoven, 2017).

It is not surprising, therefore, that over the last years we have seen an increase in interest in several variations on so-called value-sensitive design. Indeed, we might say that after the realisation decades ago that besides the required function of a system (the so-called functional requirements), we also need to take into account the quality of how well it performs that function (so-called non-functional requirements), such as performance or ease of use, we now are becoming aware that there is a third type of requirements that are important to success: the moral requirements. Moral requirements are about how the design affects individuals and society (moral impact).

E. Value-sensitive design

In the nineties Batya Friedman and colleagues introduced Value-sensitive design (VSD) as an approach to designing technological solutions that takes human values into account, throughout the entire design process (Friedman, 1996; Friedman and Hendry, 2019). Since then similar approaches have come into being, going under names such as 'values in design' and 'values for design'.

Friedman and colleagues stress the importance of not only considering the values of direct stakeholders, i.e. the persons who are going to use the solution, but also the indirect stakeholders, i.e. individuals who do not use the solution, but may be affected by it. For instance, what happens to people who do not want or are not able to use a particular app. Do we really have a choice whether we install Whatsapp on our smart phones or are we excluded from our social circles if we do?

The first step in VSD is to identify the most relevant direct and indirect stakeholders as well as the values that

¹⁰<https://www.un.org/sustainabledevelopment/sustainable-development-goals>

might be affected by the technological solution under consideration. This is called the conceptual investigation. Besides the conceptual investigation, VSD distinguishes the empirical investigation and the technical investigation. The empirical investigation is about determining how stakeholders actually experience and perceive the identified values in the specific context of the solution, and what norms they apply. This is done by interviewing individuals, organizing focus groups, making observations or conducting experiments. The technical investigation entails translating everything that has been learned into concrete design requirements that can be implemented. The three types of investigation are executed iteratively and in an integrated fashion. They can be regarded as three perspectives from which, simultaneously, the question of how to translate human values into design, is approached. VSD has been applied to a diversity of technological designs, as diverse as informed consent (Friedman et al., 2002) and windmill parks (Oosterlaken, 2015).

F. Ethical matrix

VSD is a vision and general approach rather than a step by step recipe. It provides some practical suggestions on how to approach the design process, but does not prescribe specific techniques or sequences of actions. This leaves room to apply one's own methods and instruments. One such instrument, that is very useful in building an overview of stakeholders, values and the specific impact of a technological innovation, is the so-called ethical matrix.

The ethical matrix is first introduced in Mepham (2000) as an instrument to lower the threshold for non-ethicists to engage in rational ethical evaluation of biotechnological innovations in agriculture and food production. Since then the matrix has been applied, often in an adapted version, in areas such as fishery (Kaiser et al., 2007), medicine (Chatfield, 2018) and waste management (Kermisch and Depaus, 2018). The original ethical matrix is a 3x4 matrix with stakeholder groups on one dimension and ethical principles on the other (Mepham et al., 2006). The default stakeholder groups are producers, consumers, treated organisms and biota. The ethical principles are well-being (based on the ideas of utilitarianism), autonomy (inspired by deontology) and fairness (based on Rawls' theory of justice of fairness). The cells of the matrix contain the impact, negative or positive, of the technological innovation under consideration on each of the stakeholder groups with regard to the ethical principles.

Over the years, adaptations have been suggested. Vinnari et al. (2017) propose giving more voice to non-human stakeholders. Schroeder and Palmer (2003) suggest adding future generations to the list of stakeholders and replacing the principle of fairness with the principle of solidarity.

van der Stappen and van Steenbergen (2020) adapted the ethical matrix for use in a VSD approach by replacing the default stakeholders and ethical principles by respectively the stakeholders and values identified in the conceptual investigation. Furthermore, they extend its use from evaluation of a proposed technological innovation to structuring and recording discussions about values

The corona tracking-and-tracing Appathon

To illustrate the three VSD perspectives we apply them, after the fact, to the discussion in Europe about a corona tracking-and-tracing app.

In the weekend of 18 and 19 April 2020 the Dutch government organised an Appathon. During the Appathon experts evaluated seven designs of a corona tracking-and-tracing app proposed by seven consortia, with the purpose of selecting one of them to be developed and implemented in the Netherlands. The entire Dutch population could follow the Appathon live on YouTube and could submit questions about the designs^a. The group of experts included privacy officers and ethicists, as well as epidemiologists, IT professionals and behaviourists.

This set-up can be viewed as a form of conceptual investigation: identifying the stakeholders and values at stake by having a public dialogue among experts. However, the Appathon led to much commotion because lots of people distrusted the app and the haste with which it was being introduced. The main fear was the threat to privacy and autonomy of citizens. Had the government contemplated the Appathon more carefully, this might have been a very nice example of a conceptual investigation into the relevant stakeholders and values.

Stakeholders that might have been identified in this case are not only citizens and government, but also indirect stakeholders such as health professionals, citizens that do not want to use the app, entrepreneurs and employees. Values that undoubtedly would have been identified would have included privacy, well-being and autonomy.

After the Appathon, the government decided to develop the app themselves, involving many organisations. A pilot was conducted in two regions. This pilot can be seen as part of the empirical investigation. To serve this aim the pilot must not only be concerned with the functioning of the app, but also with how both participants and non-participants experience the use or non-use of the app.

During the Appathon, the technical perspective was also applied, as the consortia presented and discussed their proposed designs. From these discussions a number of design principles arose. Thus, it was established that decentralised storage of data on the smart phone was preferable to central storage in terms of privacy impact. Autonomy of citizens can be respected by leaving it to the user of the app to take appropriate action when warned about possible infection and by safeguarding that citizens who do not use the app do not experience any negative consequences for their voluntary or involuntary refusal.

^a<https://www.youtube.com/watch?v=kV0y36UDx88>

throughout the entire design, implementation and use phases of digital innovations, including the comparison of design alternatives.

Table 1 shows, for illustrative purposes, part of an ethical matrix for an app that is to support students performing preventive health checks for citizens in community centres.

Table 1: Part of the Ethical Matrix of one of four design alternatives (van der Stappen and van Steenberg, 2020).

	Transparency	Responsibility
Citizen	Origin of advice is not clear	
Student		High responsibility for student: they may not yet be ready for that
Lecturer	Process towards advice is not clear	Requires close monitoring of student

	Security	Autonomy
Citizen	May cause sense of insecurity when student hesitates a lot about advice	
Student	May cause sense of insecurity about the soundness of their advice	Much autonomy for students, who formulate advice entirely by themselves
Lecturer	May cause sense of insecurity about whether all advices will be sound.	

Once the stakeholders and their values are identified, the cells can be filled for both the design as a whole and for specific parts of the design. By construing a series of ethical matrices during the design and implementation process, a traceable record is created of all value-related design considerations and final decisions.

Respecting human values in a design is not always straightforward. It may require creative thinking to find solutions that reduce potential negative impacts on stakeholders. The situation becomes even more tricky when two values cannot be fully satisfied at the same time. Value tensions can occur between values of different stakeholders or between personal values and public values. Examples of the latter are the offer that may be asked by governments of citizens in trading their personal privacy or autonomy for the sake of the safety or well-being of society as a whole (as in the case of a corona tracking-and-tracing app or surveillance cameras in public places). These so-called value tensions cause ethical dilemmas. Which takes us back to the ethical schools that can support us in approaching such ethical dilemmas. But not only the solution of ethical dilemmas may be challenging, sometimes we don't even realise we are dealing with an ethical dilemma, suffering from ethical blindness.

Ethical Blindness

Ethical behaviour starts with awareness. After awareness comes knowledge: being able to make ethical judgements. And after knowledge comes capacity: being able to act ethically, even in difficult circumstances. But even ethically aware and skilled persons can make unethical decisions, because of ethical blindness. Ethical blindness is a temporary state in which ethical people unconsciously take decisions that are contrary to their own values (Palazzo et al., 2012). Palazzo et al. identify 'rigid framing' as an important cause of ethical blindness.

People make sense of the world by interpreting what is happening from an, often implicit, frame, as it is called by Palazzo et al. (others will call it a mental model or lens). This frame provides the concepts to understand what we see and relate these observations to each other and to previous experiences. We construct our frames from experience, and we use them to create our world. We build frames from the moment we are born, and they are strengthened or adapted through upbringing, education, peers and experience. We may have multiple frames, but we can only use one frame at a time.

Rigid framing occurs when we are unable to switch to a different frame. Rigid framing limits our capability for moral imagination, the "ability to imaginatively discern various possibilities for acting within a given situation and to envision the potential help and harm that are likely to result from a given action" (Palazzo et al., 2012, p. 28).

The lack of moral imagination may easily lead to ethical blind spots. To prevent this, it is important that we remain 'frame vigilant', i.e. that we remain aware that we are unable to see certain aspects because of the frames we use. One way to prevent ethical blindness caused by rigid framing is to introduce diversity in a team, generating the interplay between a variety of frames.

V. The causal loop: Ethical thinking, sense making, decision making and ethical acting

So, how does all of this relate to architecture? The key question is whether in enterprise architecture we make or stimulate any moral choices. According to van den Hoven (2017) we do: "The first thing we need to realise is that the technologies we end up using are consolidated sets of choices that were made in their design, development and implementation. These choices are about e.g. interfaces, infrastructures, algorithms, ontologies, code, protocols, integrity constraints, architectures, governance arrangements, identity management systems, authorisation matrices, procedures, regulations, incentive structures, monitoring and inspection and quality control regimes ...The technology that we are using is not

neutral, since its design is informed by the world views and values of its makers. Once their ideas, values and assumptions have been embedded or expressed in digital artefacts, they start to influence the options, behaviour and thinking of users.” (van den Hoven, 2017, p. 66). In our previous paper¹¹ on Sensemaking Architecture we argue that one of the aspects of Sensemaking Architecture is that it requires architects to develop ethical thinking to be able to make sense of the world. This has to be done in such a manner that it informs the kind of decision making that is needed in today’s world.

Ethical thinking basically means applying all that has been said so far in this chapter. The core of ethical thinking is being aware of the fact that any architectural choice being made and advice being given may have a moral impact on individuals or society and consequently making the contemplation of moral requirements a standard part of all architectural considerations. This does not only concern the direct effects of the choices made in the architecture principles and models, but also indirect effects of the associated architecture governance.

Sensemaking architecture is concerned with translating relevant trends and events, within and outside the organisation, to advice that informs decision making. This translation involves noticing and recognising what events and trends are relevant, understanding and interpreting these events and trends and drawing conclusions about what is sensible for the organisation to do, given these events and trends. Sensemaking architecture based on ethical thinking implies applying an ethical lens in the interpretation and translation of events and trends. In the interpretation, by contextualising events in a moral frame, interpreting them not only in economic terms but also in moral terms. In the translation, by taking human values into account when making architectural choices, not only economic ones.

The past years have seen a surge in ethical codes and guidelines. This seems to have started with technology-related guidelines. Examples are the Machine Intelligence Garage Ethics Framework¹², the Blockchain ethical design framework¹³ and the Ethics Guidelines for Trustworthy AI¹⁴. The UN Sustainable Development Goals¹⁵ may be regarded another display of a new moral awareness. This trend continues with organisations formulating their own ethical codes. Such ethical codes can be sources for architecture principles that translate them into architectural norms. Examples of morally relevant architectural choices are the way the architecture handles customer data, the way it employs different communication channels, the extent to which it assumes self-reliance from customers, the purposes data analytics are put to, the extent to which it contributes to the Sustainable Development Goals.

¹¹<https://labs.sogeti.com/architecture-in-this-new.world-we-live-in-a-dya-whitepaper-by-sogeti>

¹²<https://www.migarage.ai/ethics-framework>

¹³<https://www.researchgate.net/publication/330069634>

¹⁴<https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai>

¹⁵<https://www.un.org/sustainabledevelopment/sustainable-development-goals>

Morally inspired architecture principles: the Madaster case

Every modern organisation has formulated some principles about how to deal with data. They are about concepts such as a single source of truth, accessibility to data for all relevant employees, complying with the GDPR and ownership of data within the organisation. These are sensible principles. They do not, however, safeguard moral handling of customers’ data.

A principle that does indeed represent a moral requirement, is a principle formulated by Madaster, a Dutch start-up company that provides identities for building materials, making them traceable and reusable and in this way stimulating the circular economy. Madaster works from the vision that the data about materials belong to the owners of these materials. And thus, they embrace the principle that no other party can take these data and use them for their own purposes or sell them to third parties.

This has important implications, such as the exclusion of some well-known platform suppliers who base their business model on earning money with valorizing the data that passes through their platform.

Whereas the architectural choices often have a direct impact on the environment of the organisation, the way architectural governance is implemented, has a direct impact on the organisation itself and its fitness for its environment. Every form of architectural governance favours particular values and stimulates particular behaviour from employees. A rule-based architecture with centrally formulated specific rules does not favour autonomy of its development teams, nor does it display trust. The same goes for the use of detailed SLAs within the organisation. But in some occasions strict top-down rules are needed, for instance to comply with regulations or interoperability standards. This is just one example. The challenge for the architect is to be able to distinguish which type of governance is required in which situation and to apply a discretionary view on architectural governance. This topic of discretionary architecture will be addressed in a forth coming paper.

Architects are already aware that they must take their stakeholders into account to be effective: design teams must be able to translate the architecture into system designs, management is more or less well-supported through the architecture in their decision making, knowledge workers are more or less well-supported through the architecture in their work, budget owners have to spend more or less money on IT because of the architecture. Most of the time, however, considerations are restricted to classical functional and non-functional characteristics of the architecture. Unless an explicit culture change is at stake, as with the rise of agile development, which generated attention to values such as autonomy. Still, this is often translated into restricting the authority of the architects instead of into designing an architecture that respects the autonomy of design teams. And the value of autonomy for design teams is

not often explicitly balanced against for instance the value of certainty for customers. We need to add moral requirements as a third type of requirements to the functional and non-functional requirements.

With architecture, as with all design processes, it is important to realise that architecture cannot be neutral. Architectures have implicit or explicit built-in values and affect the values of others by favouring particular behaviour. Enterprise architecture inherently constitutes a choice architecture and provides a subset of all potential affordances, nudging developers, decision-makers, colleagues, partners and customers to certain behaviour and designs. It is up to the architect to reflect, with every architectural decision being made, upon the impact their advises may have, intentionally or unintentionally, on direct and indirect stakeholders, using the various ethical perspectives and instruments available, but above all, by engaging in a dialogue with as diverse a population as possible.

References

- Chatfield, K. (2018). *An Ethical Matrix for Traditional and Complementary Medicine*, pages 29–40. Springer International Publishing, Cham.
https://dx.doi.org/10.1007/978-3-030-05300-0_3.
- Cheng, A.-S. and Fleischmann, K. R. (2010). Developing a meta-inventory of human values. *Proceedings of the American Society for Information Science and Technology*, 47(1):1–10.
<https://dx.doi.org/10.1002/meet.14504701232>.
- Crossan, M., Mazutis, D., and Seijts, G. (2013). In search of virtue: The role of virtues, values and character strengths in ethical decision making. *Journal of Business Ethics*, 113(4):567–581.
<https://dx.doi.org/10.1007/s10551-013-1680-8>.
- Ess, C. (2006). Ethical pluralism and global information ethics. *Ethics and Information Technology*, 8(4):215–226.
<https://dx.doi.org/10.1007/s10676-006-9113-3>.
- Foot, P. (1967). The problem of abortion and the doctrine of double effect. *Oxford Review*, 5.
<https://dx.doi.org/10.1093/0199252866.003.0002>.
- Friedman, B. (1996). Value-sensitive design. *interactions*, 3(6):16–23. <https://dx.doi.org/10.1145/242485.242493>.
- Friedman, B. and Hendry, D. G. (2019). *Value sensitive design: Shaping technology with moral imagination*. MIT Press, Cambridge, Massachusetts, USA.
<https://dx.doi.org/10.7551/mitpress/7585.001.0001>.
- Friedman, B., Howe, D. C., and Felten, E. (2002). Informed consent in the mozilla browser: Implementing value-sensitive design. In *Proceedings of the 35th annual hawaii international conference on system sciences*, volume 8, pages 96–99. IEEE.
<https://dx.doi.org/10.1109/HICSS.2002.994366>.
- Grayling, A. C. (2019). *The history of philosophy*. Penguin Press, New York, NY, USA.
<https://dx.doi.org/10.7551/mitpress/7585.001.0001>.
- Joerges, B. (1999). Do politics have artefacts? *Social studies of science*, 29(3):411–431.
<https://dx.doi.org/10.1177/030631299029003004>.
- Kaiser, M., Millar, K., Thorstensen, E., and Tomkins, S. (2007). Developing the ethical matrix as a decision support framework: Gm fish as a case study. *Journal of Agricultural and Environmental Ethics*, 20(1):65–80.
<https://dx.doi.org/10.1007/s10806-006-9023-8>.
- Kermisch, C. and Depaus, C. (2018). The strength of ethical matrixes as a tool for normative analysis related to technological choices: the case of geological disposal for radioactive waste. *Science and engineering ethics*, 24(1):29–48. <https://dx.doi.org/10.1007/s11948-017-9882-6>.
- Krijger, J. (2016). *Schaamteloos handelen*. Master's thesis, Leiden University, Leiden, The Netherlands.
<https://openaccess.leidenuniv.nl/handle/1887/45041>.
- Mephram, B. (2000). A framework for the ethical analysis of novel foods: The ethical matrix. *Journal of Agricultural and Environmental Ethics*, 12(2):165–176.
<https://dx.doi.org/10.1023/A:1009542714497>.
- Mephram, B., Kaiser, M., Thorstensen, E., Tomkins, S., and Millar, K. (2006). *Ethical matrix manual*. Technical report, LEI, part of Wageningen UR, The Hague, The Netherlands.
<https://research.wur.nl/en/publications/ethical-matrix-manual>.
- Oosterlaken, I. (2015). Applying value sensitive design (vsd) to wind turbines and wind parks: An exploration. *Science and engineering ethics*, 21(2):359–379.
<https://dx.doi.org/10.1007/s11948-014-9536-x>.
- Palazzo, G., Krings, F., and Hoffrage, U. (2012). Ethical blindness. *Journal of business ethics*, 109(3):323–338.
<https://dx.doi.org/10.2139/ssrn.2212617>.
- Pereira, R. and Baranauskas, M. C. C. (2014). Value pie: A culturally informed conceptual scheme for understanding values in design. In *International Conference on Human-Computer Interaction*, pages 122–133, Cham. Springer International Publishing.
https://dx.doi.org/10.1007/978-3-319-07233-3_12.
- Rokeach, M. (1973). *The nature of human values*. Free press, New York, NY, US.
<http://www.worldcat.org/oclc/984471799>.
- Schroeder, D. and Palmer, C. (2003). Technology assessment and the 'ethical matrix'. *Poiesis & Praxis*, 1(4):295–307. <https://dx.doi.org/10.1007/s10202-003-0027-4>.
- Schwartz, S. H. (1992). Universals in the content and structure of values: Theoretical advances and empirical tests in 20 countries. In *Advances in experimental social psychology*, volume 25, pages 1–65. Academic Press, Orlando, FL, USA.
[https://dx.doi.org/10.1016/S0065-2601\(08\)60281-6](https://dx.doi.org/10.1016/S0065-2601(08)60281-6).
- Schwartz, S. H. (2012). An overview of the schwartz theory of basic values. *Online readings in Psychology and Culture*, 2(1):2307–0919.
<https://dx.doi.org/10.9707/2307-0919.1116>.
- Spiekermann, S. (2015). *Ethical IT innovation: A value-based system design approach*. CRC Press, Boca Raton, Florida, USA. <https://dx.doi.org/10.1201/b19060>.
- Thaler, R. H. and Sunstein, C. R. (2009). *Nudge: Improving Decisions about Health, Wealth and Happiness*. Penguin Books, London, UK.
<https://dx.doi.org/10.15695/hmltc.v33i2.3307>.
- Thomson, J. J. (1985). The trolley problem. *The Yale Law Journal*, 94(6):1395–1415.
<https://dx.doi.org/10.2307/796133>.
- Turilli, M. and Floridi, L. (2009). The ethics of information transparency. *Ethics and Information Technology*, 11(2):105–112.
<https://dx.doi.org/10.1007/s10676-009-9187-9>.
- van den Hoven, J. (2017). Ethics for the digital age: Where are the moral specs? In *Informatics in the Future*, pages 65–76. Springer International Publishing, Cham. https://dx.doi.org/10.1007/978-3-319-55735-9_6.
- van der Stappen, E. and van Steenberghe, M. (2020). The ethical matrix in digital innovation projects in higher education. In *33RD BLED ECONFERENCE*.
<https://dx.doi.org/10.18690/978-961-286-362-3.33>.
- Vinnari, M., Vinnari, E., and Kupsala, S. (2017). Sustainability matrix: Interest groups and ethical theories as the basis of decision-making. *Journal of Agricultural and Environmental Ethics*, 30(3):349–366.
<https://dx.doi.org/10.1007/s10806-017-9670-y>.

About the Authors



Marlies van Steenbergen is an experienced enterprise architect. She works as an advisor and coach in the field of enterprise architecture and has guided many organisations from all industry sectors in the development of an effective enterprise architecture practice. In 2017 she became

Professor Digital Smart Services at the university of Applied Sciences Utrecht. Marlies is co-founder of DYA (Dynamic Architecture), offering a pragmatic approach to architecture based on best practices. Nowadays, Marlies focuses on the link between architecture and innovation and on Digital Ethics.

[LinkedIn](#) | [Sogeti Profile](#)



Ton Eusterbrock is an experienced enterprise architect. He works as an advisor and coach in the field of enterprise architecture and has guided many organisations from all industry sectors in the development of an effective enterprise architecture practice. Ton is one of the

promoters of DYA (Dynamic Architecture), offering a pragmatic approach to architecture based on best practices. He is also a speaker on seminars and lectures at universities and coaches students teams. Ton's focus is in architecture principles and how to organise architecture governance in modern ways of working.

[LinkedIn](#) | [Sogeti Profile](#)



Hans Nouwens is an experienced enterprise architect with 20+ years of practical experience in the field of ICT, infused with rigorous academic learning. He works as an architect and trusted advisor mainly for Higher Education institutes. Enterprise engineering, enterprise

architecture and enterprise governance are his specialities that come with DEMO and CGEIT certifications. Hans has an academic interest in systems theory and cybernetics. Hans volunteers as board member for the interest group architecture of the KNVI and as board member of the Dutch society of systems thinkers (SCIO-NL). He regularly gives guest lectures on enterprise architecture and coaches students teams.

[LinkedIn](#) | [Sogeti Profile](#)



Edzo A. Botjes is an experienced all-round (IT) professional and architect with 25+ years' experience. He works as an enterprise, program and solution architect for companies from all industry sectors. He does this by combining various expertise's to identify and resolve root-cause

issues in the Enterprise Governance domain and in the realisation domain of an organisation. Edzo is a creator and loves sharing his passion via blogs, workshops and presentations. His focus nowadays is on the evolution of organisations into antifragile organisations and the role of communication within the identity of organisations. He is Antifragility Architect at Xebia.

[LinkedIn](#)

About Sogeti

Part of the Capgemini Group, Sogeti operates in more than 100 locations globally. Working closely with clients and partners to take full advantage of the opportunities of technology, Sogeti combines agility and speed of implementation to tailor innovative future-focused solutions in Digital Assurance and Testing, Cloud and Cybersecurity, all fueled by AI and automation. With its hands-on 'value in the making' approach and passion for technology, Sogeti helps organizations implement their digital journeys at speed.

A global leader in consulting, technology services and digital transformation, the Capgemini Group is at the forefront of innovation to address the entire breadth of clients' opportunities in the evolving world of cloud, digital and platforms. Building on its strong 50-year heritage and deep industry-specific expertise, Capgemini enables organizations to realize their business ambitions through an array of services from strategy to operations. Capgemini is driven by the conviction that the business value of technology comes from and through people. It is a multicultural company of over 290,000 team members in more than 50 countries. The Group reported 2021 global revenues of EUR 18 billion. *People matter, results count.*

Visit us at

www.sogeti.com

