Chapter 10
Artefacts, Agency, and Action Schemes

Christian F.R. Illies and Anthonie Meijers

Abstract Artefacts affect users in many ways. In this paper we develop an account of the moral status and relevance of artefacts. We argue in favour of an active role for artefacts, without introducing radically new moral agency concepts. We develop a tool for the ethical evaluation of artefacts: the 'action scheme'. An action scheme is the repertoire of possible actions available to an agent or group of agents in a given situation. Each of these options has a certain degree of attractiveness. There are many influences on an agent's action scheme – we distinguish between physical, intentional, and social contexts. When artefacts are introduced, they alter an agent's action scheme; new options become available, and some are made more, some less, attractive. Our tool allows designers to analyse and evaluate the effects of artefacts on users in a systematic way; it can show them in what ways artefacts can influence what agents are likely to do. The agent remains, of course, responsible for what he

---

Parts of this chapter are taken from Illies and Meijers, "Artefacts without Agency" (*The Monist* 92/3 (2009), 422–443), reprinted here with permission of *The Monist*. The chapter presents an advanced development of the ideas of that article, which was mainly set up as a critical response to Peter-Paul Verbeek's thesis of a (limited) moral responsibility of artifacts. The present chapter has a more systematic ambition. It presents two key elements of a general framework for analyzing the moral status of artefacts: 'action schemes' and 'second-order responsibility'. In response to helpful critiques that we have received of our *Monist*-paper, we have modified our ideas in several ways. Among other things we emphasize that the 'action scheme' is a conceptual tool, not a revised ontology. Furthermore, the influences on an agent's action schemes are more clearly analyzed and described. To show the practical relevance for ethics, an elaborated example from architectural ethics has been added.

C.F.R. Illies (✉)
Chair of Philosophy, University of Bamberg,
An der Universitaet 2, D - 96045 Bamberg, Germany
e-mail: christian.illies@uni-bamberg.de

A. Meijers
Chair of Philosophy and Ethics of Technology, Eindhoven University
of Technology, P.O. Box 513, NL - 5600 MB Eindhoven, The Netherlands
e-mail: a.w.m.meijers@tue.nl

P. Kroes and P.-P. Verbeek (eds.), *The Moral Status of Technical Artefacts*, Philosophy of Engineering and Technology 17, DOI 10.1007/978-94-007-7914-3_10,

or she does. But the designer (and others involved in the creation of artefacts) has what we call a 'second-order responsibility' for changes in the user's action scheme. We argue that the action scheme and the related concept of second-order-responsibility are two conceptual tools which enable us to look at artefacts in a way more promising than alternative ethical accounts.

10.1 Introduction: Two Debates on Artefacts

Technological artefacts and systems can influence human actions in profound ways. They make new kinds of action possible, for example: communicating at a distance, moving at a speed well beyond natural human capabilities, intervening in the human body and brain on an unprecedented scale. Artefacts can also alter our behaviour and make some actions more or less attractive. The physical characteristics of a house, for example, can invite people to feel responsible for their residential environment and act accordingly – or they can demotivate them from so doing. Technological artefacts also enter into the process of decision-making, as, for example, when an aeroplane flies independently, or when computer-based decision support systems are used in medicine, the legal domain, or by the army.

The point of this paper is to analyse and interpret these profound effects artefacts have over human life. We do so with an ethical question in mind: What is the moral status of artefacts? How should we understand their moral relevance?

Various ways of accounting for the role of artefacts have been put forward. On the one hand it is argued that artefacts are simply tools for actions and thus morally-neutral means to (moral) human ends. According to this theory, artefacts have no moral relevance and human agents alone can be held responsible for actions accomplished with the use of artefacts. Artefacts are seen as being categorically different to agents. On the other hand there are theories that attribute agency to artefacts, thereby rejecting traditional conceptions of artefacts as morally neutral. Bruno Latour's well-known actor-network theory states that technological artefacts 'act' and that together with human agents they are grouped in the same category of 'actants' (this is the principle of generalised symmetry).[1] These theories often claim that artefacts are also in some sense morally accountable for their effects.

Upon closer analysis there are actually *two* debates here:

- The first debate relates to the ways in which technological artefacts influence our world; to whether they actively determine their effects in a self-guided way or whether they have a more passive role (as mere extensions of the human body). We shall term this the 'Autonomy Debate', because what is really at stake here is whether the artefact's influence is fully explicable in terms of designer and user intentions or whether such influence extends beyond designer and user control in gaining some degree of autonomy. There are two extreme positions taken in this debate. Some regard artefacts as *mere instruments* of human agency; this is

[1] See, among many other publications, Bruno Latour (1987).

dubbed the 'Instrument Position'. Others grant artefacts a degree of *autonomy*. In its extreme this position holds that artefacts are on a par with goal-directed autonomous human agents; we call this the 'Agency Position'.

- The other debate relates to the moral relevance of artefacts; we therefore call it the 'Moral Relevance Debate'. Something has moral relevance in our definition if it substantially affects the moral evaluation of a situation or the 'oughts' of the agents involved. In general this requires that artefacts are directly or indirectly linked to intentional actions and that they have an impact on basic moral goods, values, rights, etc., either by promoting or inhibiting their realisation. There are here also two opposing views. The 'Neutrality Thesis' states that artefacts are morally-neutral means to various ends pursued by human beings. In this case artefacts are not by themselves seen as morally relevant. This is aptly exemplified in the statement "it is people who kill people, not guns". At the other extreme there is the "Moral Responsibility Thesis", according to which artefacts (or human beings in combination with artefacts) are considered to be morally responsible.

In the Moral Relevance Debate both views are closely linked to the two positions taken in the Autonomy Debate: the Neutrality Thesis places all moral weight on the intentionality of the users and/or designers of technological artefacts and sees the artefacts themselves as mere transmitters of these intentions – it is therefore akin to the Instrument Position. By contrast, the Moral Responsibility Thesis presupposes the Agency Position. Only if artefacts are agent-like, that is to say, the origin of certain morally relevant effects, and not mere transmitters, can they be regarded as morally accountable or even responsible for the subsequent effects.

In this paper we set out to elucidate the role of technological artefacts in human affairs by examining both debates. Our aim is to give an account of artefacts which does justice to their sometimes unexpected influence on what we do (the Autonomy Debate) and to their significance in morally relevant matters (the Moral Relevance Debate). In both debates we will argue in favour of an active artefact role *without* introducing radically new moral agency concepts. We intend to analyse these issues primarily from the perspective of those who are responsible for the design, creation, or production of new technological artefacts. Our concern is ultimately the ethical responsibility they might bear for the effects of these artefacts.

We shall start by discussing in more detail the Moral Responsibility Thesis and its problems (Sect. 10.2). The discussion will form the background to our own account, which will be unfolded in two steps. In Sects. 10.3 and 10.4 the perspective switches from artefacts and actions to what we call *action schemes*. An *action scheme* is the repertoire of possible actions or options available to an agent in a given situation where each such option has a certain appeal to the agent. The notion of action scheme is discussed here in some detail and the formation of action schemes is looked at. On this basis we go on to develop, in Sects. 10.5 and 10.6, a notion of *second-order responsibility*, which allows us to analyse the moral relevance of artefacts in greater detail. In Sect. 10.7 we apply the action scheme to a concrete case: architectural design. Finally, in our conclusion, we will take up the question of whether or not the resulting position is stronger than other positions

(Sect. 10.8). This will depend very much on the criteria for a successful account of the role of artefacts. We will briefly argue for some criteria in this section, before concluding that our position is more promising than others.

10.2 The Moral Responsibility Thesis and Its Problems

It is quite common to attribute agency to artefacts. For example, we naturally tend to refer to computers as thinking and acting entities. In an empirical study B. Friedman and L. Millett showed that 83 % of all computer science students attribute some aspects of agency, like decision-making or intention, to their computers – and 21 % even implied that computers have moral responsibility for errors ("It is the computer's fault").[2]

Why should a philosopher conceptualise the activities of artefacts in terms of agency let alone moral agency? Such a move evidently blurs distinctions in moral philosophy that have proven to be useful for a long time. The question is hard to answer in general. Let us therefore look at a concrete defence of artefact-agency: Peter-Paul Verbeek's version of the Moral Responsibility Thesis.[3]

Verbeek draws our attention to the fundamental ways in which artefacts *actively* shape the way we interact with the world by changing our perceptions and actions. This process is called *mediation*. Verbeek distinguishes two types of mediation of our perception.[4] Artefacts can extend the sensory capacities of our body, and artefacts can generate new representations of the world we live in. There are also several ways in which artefacts *actively* shape (or mediate) our actions. They do so by having an 'invitation and inhibition' structure and by delegation (the phenomenon of actions being transferred to other (types of) agents).[5]

Verbeek's position can be located within the two debates mentioned above. In the Autonomy Debate he rejects (with Latour) the *a priori* dichotomy between human and non-human actors as well as the idea that artefacts are merely tools in complete control of human agents. He defends the Agency Position by arguing that certain essential conditions for agency, if interpreted in the right way, apply also to artefacts. They *actively* shape our relationship with the world, their mediating role is "fundamentally unpredictable" (Verbeek 2008b, 100) and "their mediating role cannot be entirely reduced to the intentions of their designers and users" (ibid., 95). Verbeek even argues that artefacts have intentionality: "It seems plausible, then, to attribute a specific form of intentionality to artefacts. This 'material' form of intentionality is quite different from human intentionality in that it cannot exist without being supported by human intentionality. Only within the relations between human

[2] See "'It's the Computer's Fault' – Reasoning About Computers as Moral Agents", http://www.sigchi.org/chi95/proceedings/shortppr/bf2_bdy.htm (accessed September 2011).

[3] See for a more extended discussion of Verbeek's position Illies and Meijers (2009).

[4] Verbeek explicitly refers to Don Ihde (1979, 1991).

[5] Verbeek (2005), Chap. 5.

beings and reality can artefacts play their 'intending' mediating parts" (ibid., 95). Along the same lines Verbeek defends the view that artefacts are able to have non-absolute freedom by stating that they can enter into associations with agents who enjoy certain forms of freedom. "Just like intentionality, freedom also appears to be a hybrid affair, most often located in associations of humans and artefacts" (ibid., 98).

In the Moral Relevance Debate Verbeek seeks to eradicate the view that only the intentions of designers, producers, or users of artefacts can be evaluated in moral terms. In his view technological artefacts themselves are morally relevant, because of their mediating role. They affect the quality of our lives, they make us aware of morally relevant distinctions or phenomena, and they even force decisions upon us. If this line of reasoning is combined with the Agency Position, according to which artefacts *actively* influence our relation to the world and have some form of autonomy, then the Moral Responsibility Thesis formulated above follows. Verbeek does not go so far as to argue that artefacts *are* moral agents (though some of his formulations come very close to doing so). Instead he states that "moral agency is distributed over both humans and technological artefacts".[6] Thus hybrids of humans and artefacts are morally accountable. They have intentionality and freedom and can therefore be seen as fulfilling the necessary conditions of moral agency (Verbeek 2008b, 93 and 98). On the basis of this concept of hybrid agency he also transforms the notion of *human* moral agency: Moral agency, intentionality, and freedom are always embedded in a material context. "Intentionality is hardly ever a purely human affair, but most often a matter of human-technology associations" (ibid., 99).

10.2.1 Some of the Problems of the Moral Responsibility Thesis

What makes Verbeek's account attractive is that he takes the unpredictability of artefacts very seriously and acknowledges that their effects can go far beyond the intention and control of designers and users. In this sense one can certainly talk of an 'active role' of artefacts or perhaps even of them being 'autonomous' in the sense of independent from human intentions. (Though this would elicit the aspect of self-determination which is normally included in our understanding of autonomy – see below.)

Verbeek's arguments in favour of the moral relevance of artefacts are equally appealing. Many of his artefact examples raise moral questions that did not exist before. Artefacts can change our perceptions and actions, and in so doing they ultimately change us and our relations to the natural and social world. This is obviously an issue of great moral significance. Verbeek presents a strong case against the Neutrality Thesis, according to which artefacts are simply morally-neutral means to the ends pursued by agents.

[6] See Verbeek (2008a, 24). See also Verbeek (2008b).

However, there are also very good reasons *not* to adhere to Verbeek's conclusions in both debates.[7] In his analysis of the moral relevance of artefacts Verbeek simply ignores elements of moral agency which, in extensive philosophical analyses, have been shown to be of great importance. Many philosophers would argue, for example, that moral agency not only requires intentionality and freedom but also the ability to understand the moral options and moral demands of a particular situation. It also requires the ability to reason and to perform actions for good moral reasons and possibly even the capacity for empathy and for moral sentiment. The introduction of an undemanding notion of 'moral agency', as advocated by Verbeek, seems of no further use: nothing is gained but much is lost in this move, namely a useful category for action theory and ethics. Of course, one can also define moral agency in minimal ways to include artefacts but then the richer concept of full-blooded agency falls away, where goals are consciously adopted "on the basis of an overall practical assessment of the options and opportunities."[8]

There are similar concerns surrounding 'associations' of artefacts and human beings, as Verbeek calls them (following Latour). If what is meant by 'association' is a new *unity*, then the emergent properties of that unity should include the properties relevant to moral agency. Verbeek sets out to show that these associations have freedom and intentionality but he does *not* take into consideration other properties of moral agency, such as the ability to reason. This makes his attribution of *moral agency* and *moral accountability* to these associations highly problematic. If what is meant by 'association' is a *hybrid* of artefacts and humans, as several of Verbeek's formulations suggest, then the conclusion will be no different. In a hybrid the properties of moral agency will be located in one of the two constituting elements and it would be a mistake to attribute moral agency to the hybrid as a whole.

Let us give an example. In the case of a man using a pistol Verbeek would argue that the two form an association and that the man-pistol association has moral agency and is accountable. The association as such becomes blameworthy. That however, blatantly contradicts our practice of blaming and punishing. We do not (and we should not!) put the murderer *plus* his pistol, or the hacker *plus* his computer, in prison. In such cases it is the human agent alone who, according to standard moral practice, is blameworthy. (If the artefact and human being association is conceived of as a hybrid then there need not be a conflict. Then the human being remains the locus of moral agency and accountability.) Artefacts may *diminish* the moral responsibility of humans by being beyond their full control ("He did not know that the new car accelerated so quickly") – even Aristotle reminded us that ignorance limits responsibility.[9] But the responsibility is not partly 'taken over' by artefacts. That would be an inflationary understanding of accountability (or even responsibility) which would render most of our traditional ethical concepts useless

[7] We will focus here on the Moral Relevance Debate. The Autonomy Debate is taken up again in Sect. 10.6.

[8] Wilson (2007).

[9] See his discussion in the third book of *Nicomachean Ethics*.

and would disconnect accountability from praise and blame or any adequate reactive attitudes. Moral responsibility would then become a rather empty notion.

What we need, in our view, is an account of artefacts that

(i) explains their fundamental role in what we perceive and do
(ii) can be used in the moral evaluation of artefacts
(iii) does not revoke useful notions such as full-blown human agency and moral responsibility.

As will be shown below, this will be made possible by introducing a level of analysis which we call the 'action scheme' level.

10.3 Changing the Perspective: From Action to the Action Scheme

In order to gain clarity at a theoretical level it is often useful to look at practical cases. Let us take the often discussed example of the speed bump, which forces car drivers to slow down. Here an artefact seems to prescribe a certain course of action. We can also present this as a conflict for the driver: if she drives slowly, the car will be fine though she might arrive late for work. If she does not slow down, then she might be on time but her car will be at risk – thus making her potentially *very* late. Still, the woman has a choice. Yet one of the things she could have done without the speed bump in place, namely driving fast in order to arrive in time, has become much more unattractive due to the introduction of the speed bump. We summarise the situation as follows: (1) without the speed bump the driver has two (relevant) options for action and (2) due to the artefact, the attractiveness of one of the options has changed.

Rather than looking at how artefacts influence individual actions we now focus on how artefacts affect the *repertoire of actions* available to the agent. In what follows we shall conceptualise this as 'action scheme'. It is defined as follows:

> An *action scheme* is the *repertoire of possible actions (each of which has a certain degree of attractiveness) which is available to an agent, or group of agents, in a given situation.*

The specific attractiveness of an action results from many factors: it is influenced by the degree to which, in a certain context, the action corresponds to the desires, inclinations, or talents of an agent, with her previous history, her convictions, ideas, intuitions, and character.[10]

Technological artefacts influence action schemes. Not only do they affect the agent directly but also indirectly by modifying the repertoire of possible actions available to her, *including* their attractiveness. For example, the introduction of the

[10] We fail to see why Selinger et al. characterize our position here as "attractiveness appears to be a feeling" (p. 84). The attractiveness of an action to travel by car, for example, is determined by its cost, its fuel consumption, the time it takes, and so on, in addition to its emotional characteristics.

mobile phone has extended our range of possible communicative actions (I can contact my wife in Utrecht while walking in the Black Forest). And the speed bump takes away the attraction of driving fast for the woman because she does not want to ruin her car. The action scheme, however, is part of a bigger story: agents and their actions are always embedded in a *dynamic context* and the action scheme is relative to this context. A woman who hates her husband might see speed bumps as a welcome opportunity to ruin his precious car; that will make her accelerate rather than slow down.

Let us therefore try to give a general account of how an action scheme is formed.

10.4 The Formation of Action Schemes

Analytically there are three *types* of contexts that shape an agent's action scheme: the intentional context, the physical context and the social context. They provide possibilities and set boundaries for the actions available to an agent and give them a certain attractiveness. Together these three form the overall context of an agent's action scheme. Let us look now at the three contexts in more detail.

The physical context consists of the physical make-up of the agent and the physical properties of her situation. Physical is meant here in a broad sense: it contains everything that is described by the natural sciences, including biology. The driver of a certain car might be an average-sized, dark-haired woman of 32. The car may be aerodynamically well-shaped, accelerating and braking quickly. The speed bump has a certain length and height. There is also a wider physical environment which includes the weather and even the most general physical possibilities and impossibilities as described, for example, by the laws of gravity (without which speed bumps would have rather different effects). Precisely which of these physical properties are relevant will depend on the particular situation.

The social context consists of the social role, status, and rights of the agent involved, of the social characteristics of her situation, and the wider social environment. Traffic-rules belong to the social situation but also the costs of repairing a car (prices are social arrangements). The broader social environment will also include the institutions of the country, its laws, communication patterns, family structures, and so on. Let us assume that the driver is a paediatrician who is on her way to the hospital to do her night shift but had a row with her husband before leaving home.

The third context is the intentional one. It consists of the intentional make-up of the agent, that is to say, her beliefs, desires, emotions, experiences, expectations, and memories (for example of her husband shouting at her before she left the house). Intentional states are never isolated but are always embedded in a web of other intentional states. The woman might consider the car to be rather expensive and she knows what the car means to her husband. In order to know what a particular belief does to the action scheme, one would have to know how that belief relates to other beliefs, intentions, and desires (she might be furious with him but also afraid of arriving late at the hospital). Perceptions are also part of the intentional context,

constituting reasons for the agent to hold certain beliefs and have certain intentions (the driver having found a long blond hair on her husband's overcoat the day before harbours all sorts of concerns). Obviously, intentionality always operates against a background that is often not fully conscious to the agent in a given situation.

When we act, we 'choose' an option from the action scheme, i.e. from the repertoire of possible actions that appear available to us and that have a certain attractiveness in a given situation. If we act consciously, like the driver who wonders on what to do, then our action will be based on deliberation. The degree of attractiveness of an option indicates the probability that the agent will choose it if no other factors play a role in the deliberation; it does influence but not fully *determine* what the agent will do. However, choices are often made without reflection; the agent may use established routines, or may simply take what seems to be the easiest or most obvious course of action. In the latter case the attractiveness of the actions will be decisive; the most attractive action from the action scheme will be equivalent to default behaviour.

It should be added that the three contexts that shape action schemes will not be equally relevant in all situations; which ones come into play and how that happens depends on the particular situation. The typology we have introduced should also not be taken as a systematic aetiology, but should rather be seen as a pragmatic way to account for the ways in which artefacts and other things affect action schemes. The typology is also a simplification in the sense that many things will be parts of different contexts all at once. Technical artefacts, for example, are not simply physical objects, but objects with a function for users. They are thus related to physical, intentional and social contexts alike. They can be seen as mind-dependent objects, as *objects made for action*. A car, for example, is not just a physical object; it is also linked to human intentions by being a means to an end, or by being an object of desire. Cars are also related to social contexts. By having a car, the paediatrician can live in a green suburb rather than close to her workplace; and the make of car might lend a certain social status. Ultimately it was the danger posed by cars that had led to local government decisions to introduce speed bumps in the first place.

A further clarification concerns the way in which the three contexts "influence", "shape", "form", or "determine" the options for action in the action scheme. These expressions are intended to cover a broad range of influences. A traffic regulation with high fees creates a *reason* for the driver to consider driving with reduced speed, whereas a flat tyre effectively blocks the option to go by car in a causal way.

The introduction of action schemes in the moral debate about artefacts is not meant to introduce a new ontological entity. As we said before, an action scheme is the repertoire of possible actions that appears available to an agent in a given situation. This repertoire is just a simple list of options for action in a situation accounted for in a systematic way. Ontologically speaking we have not introduced anything new.[11]

[11] It is therefore a misunderstanding of our position to conceive action schemes as separate ontological entities that have causal powers to motivate agents, as Selinger et al. (2011, 84) do. What motivates agents is their beliefs, desires, emotions, and so on. The knowledge of new options for

What we have introduced, however, is a different level of analysis in ethics. In our view we should not analyse individual actions alone, but we should also systematically analyse the *repertoire* of actions available to an agent in a given situation. In ethics this repertoire is usually taken into account when an 'all things considered' moral judgment is made. Given the alternatives a certain action is identified as the best moral action in that situation. So it *seems* that traditional ethical theories already include the action scheme in their analyses. The difference, however, is that these theories take an agent's action scheme *for granted*. The move we make is to regard it as an explicit and distinct object of analysis in moral evaluation. We want to address questions such as: is this repertoire large enough in a given situation, is it adequate for the specific characteristics of the agent, does it contain enough options that are morally attractive, and so on. Making action schemes the object of analysis in ethics is especially important when analysing new artefacts such as buildings, smartphones, or brain implants.[12] Artefacts are *standing possibilities for action*, they make actions possible.

Action schemes are always perspective-bound. The options for actions available to an agent can be different if seen from the first-person perspective or the third-person perspective. Many people have smartphones which can be used as phone, calendar, means to communicate via email, navigator, torch, and so on, depending on the applications installed. Few people know *all* those functions or are able to use them. Thus from a first-person perspective the influence of such a phone on the action scheme can be very small, but from the perspective of someone else, for example the designer, it may be very large. What matters, however, *when acting* is the first-person perspective and the options available to the acting agent. Unknown options for actions do not belong to the action scheme of an agent.[13] In addition, limitations of the action scheme can also depend on other things, such as emotions:

---

action, including new actions made possible by artefacts, may also motivate agents to act in certain ways. There is, however, ontologically nothing mysterious about this. The repertoire of possible actions is only made larger and their attractiveness for the agent changed, which might result in a different outcome of deliberation. There is no reason to assume that because of the influence of technological artefacts on actions schemes, we have to assume that "action schemes are metaphysically real and must be found somewhere" (p. 85). In a more radical spirit, Peterson and Spahn (2011) argue that Ockham's razor would apply to the unnecessary ontological claims we make by introducing the notion of 'set' in the actions scheme discussion. We agree that we do not need these unnecessary ontological claims. Our initial phrasing of action schemes in terms of *sets* of possible actions may have added to the misunderstanding, the notion *set* was intended there in an everyday sense. See also Koller (2011) for useful suggestions about the possible readings of the notion of a *set*. We believe, however, that our account is compatible with various ontological readings of the notion of 'repertoire of actions', as long as this allows for an evaluation of such a repertoire in terms of moral preferences.

[12] This not only applies to *technological* artefacts but also to *social* artefacts, such as laws, organizations, and institutions. The possible application of the action scheme approach is thus much wider than discussed in this paper.

[13] As mentioned before, this does not mean that the agent always needs to be *consciously* aware of these options.

a very fearful person might not see certain options as options *for him*; he simply does not dare to choose them.

Before we return to our two debates and to the role of artefacts, one further point must be stressed: action schemes are to be understood as *dynamic*. They are open to changes. These changes do not simply happen to us, we are not just passive in this respect, but we influence these schemes ourselves (either our own or the action schemes of others) – and we do this consciously or unconsciously, intentionally or unintentionally. Politicians, for example, actively introduce rules and regulations in order to promote certain actions and discourage others. Changes can also occur at an individual level when we modify our social context, by, for example, being friendly to someone, thereby making his option to be similarly friendly to us more attractive to him. Furthermore, our past actions codetermine our scheme of future actions in several ways. A decision selects and excludes options, but it can also pave the way for future actions by opening up new opportunities. And action schemes can be mutually exclusive: alternative designs of an artefact might lead to alternative actions schemes – a building is either accessible or not accessible to wheelchair-users. There are many more ways of shaping an action scheme: by education, by setting example, by initiating a habit. It should be stressed that many changes in action schemes are neither intended nor controlled: the blond girl on her father's arm (who happened to lose a long hair when standing next to a stranger in the tube) had no idea that this would result in 'trashing the car' being an attractive option for a young paediatrician in a green suburb.

Action schemes are useful for ethical analyses because they help us to articulate and account for moral differences. We might say, for example, that it is morally preferable for supermarkets to sell fair-trade coffee rather than not. This can be expressed as follows: an action scheme A1 is morally preferable to an action scheme A2 if the only difference between A1 and A2 is that A1 contains an additional option for action that is morally preferable to the other available options for action. To provide cars with first aid kits allows people to help others efficiently after an accident; which seems morally preferable to not offering this option.

It is here that one might want to compare the action scheme approach with Amartya Sen's 'capability approach'. According to Sen, we should look at the concrete capabilities that are open to an agent – we must ask what he can do on the basis of circumstances, resources etc. Sen argues that we should not evaluate simply the goods or resources that situations, policy-making etc. provide, because different people cannot always use them in the same ways.[14] The focus should be on the *actual capabilities* (or freedoms) of real people in some situation. Their individual capabilities should be increased so that everyone can achieve fundamental 'functionings' (i.e. basic states and activities of human beings, such as being well-nourished or being able to vote in an election). All of this is highly compatible with the proposed action scheme approach, which can also be seen as a more precise articulation of some of the ideas of the capability approach. Similar to the capability approach, the action scheme approach is sensitive to the different ways

[14] Sen, Amartya (1982).

in which people can make use of opportunities. Sen talks about "conversion factors"[15] as the degree to which a person can make use of resources and transform them into functioning; the very same good or resource can bring very different kinds of freedom to people. (A car, for example, does not offer travel to someone who cannot afford petrol.). The same is captured by the first-person perspective on actions schemes; we can ask what options for action an artefact can provide for real people in real situations.

Moral dilemmas concern situations where there is no action possible that does *not* violate some fundamental norm or value. These dilemmas cannot be expressed by referring to the moral properties of *single* actions. They need a reference to the *repertoire* of actions available in a given situation, to the action scheme. Antigone was confronted with a dilemma because her action scheme contained only two options, and they were sacred duties of which the one could only be realised at the expense of the other: whatever she does, she will be guilty.

Another morally relevant feature of the action scheme lies in the varying attractiveness of different options: moral education might be construed as a process of widening the range of options (developing new skills and sensitivities means having new options for action) *and* making the morally good choices more attractive (self-discipline leading to the reduced attraction of options that should be avoided).

What exactly are the criteria for preferring action scheme A1 to action scheme A2 from a moral point of view? Different ethical theories will express different ideas about the criteria we use to evaluate action schemes. Since we do not want to argue in favour of any specific ethical theory, this can be left open. The notion of an action scheme is an analytic tool to express morally relevant differences at the level of the repertoire of actions available to an agent in a given situation, not an explanatory or normative theory. As such, it is neutral with respect to ethical theory – and compatible with different theories.

It is obvious that different ethical theories give very different answers to the question: *What is good?* However, in most cases there remains a link to actions: for the core function of ethical theories is to offer a framework for the evaluation of what to do from a moral point of view – by clarifying what is good and what should be supported or avoided. If nature has intrinsic value, then do not destroy the rain forest! If autonomy is of prime importance, then respect human beings and their basic rights! It is here that the suggested tool finds it application: it is not linked to a particular ethical approach but helps to clarify the ways in which the introduction and use of artefacts can influence what people are likely to do. The key of our proposal is to extend the traditional ethical reflection with an analysis of the effects of an action on somebody else's action schemes.

This result can be phrased in a more consequentialist language (an action is good if it brings about a better action scheme), or in a more deontologist phrasing (act so that you promote the freedom of others to act by providing them with better action schemes). The action scheme might not be a helpful analytic tool for *all* ethical

[15] Sen (1992, 19–21, 26–30, 37).

theories, but we believe that it can assist in many situations, in particular with the ethical analysis of artefacts and their moral impact.[16]

10.5 Action Schemes and Second-Order Responsibility

Normally we hold someone responsible if he is likely to be blamed (if what he did is bad) or if he is a candidate for approval or praise (if what he did was good). Furthermore, if an agent is held blameworthy he will have to satisfy certain conditions of agency. A prime condition is that the action in question was performed voluntarily. Two specifications of this condition come down to us from Aristotle. Firstly, the action must be under the agent's control; it must be up to him whether he performs the action. Secondly, the agent must know what he is doing; that is to say, he must be aware of the action and its consequences.[17] The driver is normally responsible for the speed of her car; if she knows about the effects of speed bumps but fails to slow down, then her husband will rightly blame her. Regarding *moral* responsibility there is a further condition that needs to be stressed, namely

[16] Peterson and Spahn (2011) raise an objection which seems to undermine our claim that the notion of an action scheme is neutral with respect to ethical theories. In their consequentialist view it is a "category mistake" to attribute moral properties to *action schemes* or *sets* of actions. Doing so would be "a radical departure from one of the most basic assumptions in moral philosophy", viz. that only actions are the true bearers of moral properties (ibid.). A number of observations have to be made here. First, their claim is factually incorrect. Virtue ethics, for example, is not about actions but about the moral traits of a person's character. But the real issue is of course whether consequentialism is compatible with our action scheme approach. If we take consequentialism to be the general claim that the moral properties of X depend only on its consequences, then even within consequentialism this allows for different types of X and also for what could be conceived as relevant consequences. In the history of consequentialism the X that is the object of moral analysis has not only been *actual* or *concrete* action but also *abstract entities* such as possible actions, intended actions, likely actions, or counterfactual actions. Therefore, the fact that action schemes are abstract entities does not make them incompatible with consequentialism. Moreover, not only actions but also *motives*, *virtues* or *character traits* have been put to consequentialist analysis. Thus a philosopher defending (direct) consequentialism about motives holds that the moral qualities of a motive depend on its ultimate consequences in the world. A consequentialist stance on virtues holds that the moral qualities of a character trait depend on the consequences of that trait. Given this plurality of possible approaches within consequentialism we see no reason why action schemes cannot be relevant to a consequentialist moral analysis. It seems perfectly possible for a consequentialist to say that an action scheme that contains a dilemma (two options for action that have equally negative moral consequences) is morally inferior to an action scheme that contains a third option for action that has positive moral consequences. Finally, the objection by Peterson and Spahn that it makes sense to attribute moral properties only to something that is under our control seems to be too strong. It would rule out moral judgments about situations that are not under our control where these judgments seem to be perfectly natural. We fail to see, for example, why a consequentialist cannot make the judgment that a situation in which an agent finds herself in a trolley car with failing brakes and only two options for action (which both involve killing people), is morally inferior to one which contains a third option for action in which nobody is killed.

[17] See *Nicomachean Ethics* III.1–5 (1110a–1111b4).

awareness of the relevant norms or values in a given situation. We place moral blame on an agent only if it is clear to her that she *should not* have performed the action from a moral point of view.

If people are physically or psychologically forced to do something they are generally not blamed or praised. They have not 'acted' in the full sense of the word. In terms of action schemes: a responsible agent is someone whose action scheme offers him *different* possible actions. He is responsible in a given situation only for the choice between those possibilities. The repertoire of available possibilities also denotes the limits of his responsibility; no one is to be blamed or praised for *not* having chosen *impossible* actions. (This follows from the first specification of the condition 'voluntary'.)

As already said, action schemes are dynamic and shaped by many frameworks – and these frameworks will partly depend on what other agents actually do. There are two possibilities here. For agent A and agent B, $action_1$ and $action_2$, and time t_0 and time t_1 we can say:

1. Agent A can influence with $action_1$ at t_0 the action scheme of agent B at t_1
2. Agent A can influence with $action_2$ at t_0 his own action scheme at t_1.

(Obviously, in this case the actions that are necessary to influence B's or A's action schemes at t_1 are part of A's action scheme at t_0).

The fact that action schemes are not simply given but can be influenced gives us responsibility for them to the extent that they can be shaped by us. This allows us to distinguish between two ways of being responsible for actions.

Either

(1) We may consider the responsibility of agents for their actions in the more traditional sense. In such cases we look at the actions and their outcomes in general; we ask what effects an action has had on the world or on other human beings, whether the action was in accordance with moral rules, and so on.

or

(2) We may focus on the ways in which our actions affect the action schemes of others (and ourselves). In these instances we look at the ways our actions influence the repertoire of future actions that agents have at their disposal.

We call responsibility in case (1) a "first-order responsibility" and in case (2) a "*second-order responsibility*", the difference in order reflecting the change of perspective from action to action schemes.[18] The second-order responsibility widens

[18] The distinction between first-order and second-order responsibility does not correspond to the distinction between direct and indirect responsibility. We can bear direct and indirect responsibility for actions as well as action schemes. The distinction between direct and indirect responsibility reflects the degree to which my actions *causally* contribute to the realization of a certain effects. Some effects will be the direct result of my action, others will be realised only if other contributing causal factors are in place.

the realm in which we hold agents and ourselves responsible, but does not make it too broad.[19] All influences on the action scheme which remain outside human control (an earthquake for example, as part of the physical framework) are not something we are responsible for – but we are responsible for designing nuclear power stations in such a way that we have sufficient options for action when an earthquake damages a nuclear power station.[20]

Let us look at an example to illustrate the distinction we have in mind.[21] If a doctor makes an ultrasound image of an unborn child, we may focus on the effects of that very action on the mother, father, and child. The image gives information about the well-being of the child, its development and so on, all of which may be variously reassuring or alarming for the parents. The doctor has first-order responsibility for this action. We may also focus, though, on the way that making such diagnostic images changes the very options that the parents have at their disposal. Suddenly they may have to consider actions such as prenatal cures or even abortion, actions that did not need to be taken into account before. In the long run we may expect the practice of caring during pregnancy to change, because being a morally good parent may then be seen to involve making ultrasound images of your unborn baby in order to be informed about its health status. Doctors, but also the engineers who develop these types of imaging devices, can be said to have second-order responsibility for changing parental action schemes.

Looking at responsibility from this second-order perspective does not mean having to hold people responsible for what others *actually* do; no one is to be blamed or praised for the choices of others.[22] The second-order responsibility of A does not diminish the normal (or first-order) responsibility of B; B remains fully responsible for her choice on the basis of her action scheme at a certain time. But we do hold A

[19] The notion of second-order responsibility is different from the notion of meta-task responsibility, as discussed in Van den Hoven (1998). Meta-task responsibility is defined by him as: "A user A has a meta-task responsibility concerning X means that A has an obligation to see to it that (1) conditions are such that it is possible to see to it that X is brought about and (2) conditions are such that it is possible to see to it that no harm is done in seeing to it that X is brought about" (Van den Hoven 1998, 103). The idea that agents are not just responsible for a task but also for the conditions that make it possible to carry out that task in a responsible way differs from the idea developed here. Second-order responsibility implies that agents are in some sense not only responsible for their actions but also for the repertoire of actions available to them and others. That involves much more than securing enabling conditions for a certain task. Both notions have in common, however, that they widen the responsibility of agents beyond a specific task or action.

[20] Peterson and Spahn (2011) have argued that the action scheme model does not allow for a sharp distinction between human influence via artefacts and natural phenomena affecting the action scheme: these phenomena "are at least as unpredictable and difficult to control as are new technologies." (p. 12). Yes they are – but responsibility only comes into it when an event or phenomenon is directly or indirectly linked to intentional action. To the extent that physical events are outside human control, there is no point in regarding any human being, let alone the events, as morally responsible for changing an action scheme.

[21] The example is taken from Verbeek (2008a) and adapted for our purposes.

[22] Cases of coercion are no exceptions to this rule: if we force someone to do something, we (and not she) are responsible for the harm we did to her *and* for the action she performed.

responsible for having influenced B's action scheme. It follows that we can regard it as a *moral task* to foster good action schemes, both for ourselves and for others who are dependent upon us.

It should be noted that second-order responsibility is not necessarily a *weaker* form of moral responsibility; it might be quite the contrary. It is often particularly wrong to corrupt the action schemes of others. Fagan is certainly worse than Oliver Twist, his pickpocket pupil. This might also be the case with someone who corrupts his own action scheme, by, for example, taking drugs. Aristotle demands that "penalties are doubled in the case of drunkenness", because the drunkard "had the power not to get drunk and his getting drunk was the cause of his ignorance."[23] Although Aristotle does not give a satisfactory explanation for the doubled penalty,[24] we can support his point by action scheme analysis. If someone gets drunk deliberately, he alters his entire action scheme and thus also the basis of *many* future choices. Thus getting drunk is a bad action that will easily multiply and lead to many more bad actions. If we allow some consequentialist reasoning to enter ethics, we will regard this as worse than simply failing once.

10.6 The Moral Agency of Artefacts Revisited

Let us return to our original question. Given the profound effects of technological artefacts on human affairs, how can we understand their role and evaluate their moral significance? The two notions we have introduced, 'action scheme' and 'second-order responsibility,' are analytic tools designed to clarify the ways in which human agents are affected by artefacts, but also to show how *designers* can affect other agents by the ways in which they craft artefacts. The two concepts will also enable us to render more precise the moral responsibility designers have, and the extent to which artefacts themselves can be said to possess characteristics of moral agency. The crucial step in this understanding of artefacts is the move from action to action scheme. Artefacts *do* matter for our actions, obviously, but we cannot fully understand how profoundly so long as we ignore their influence on the repertoire of actions available to an agent in a given situation, where each option is presented in a certain attractive light. As functional objects artefacts are part of the physical, the intentional, and the social contexts of actions discussed above.[25] And therefore, what agents can do depends often essentially on artefacts.

[23] See *Nicomachean Ethics* III.5.

[24] Aristotle justifies the harsher punishment by saying that "the moving principle [for his ignorance] is in the man himself" – but we might remark that the sober man also possesses the moving principle for committing a crime himself. This simply leaves open the question why should it be worse to drink (and thereby make oneself ignorant) *before* doing something wrong rather than doing something wrong straight away.

[25] For an analysis of the dual nature of artefacts see Kroes and Meijers (2006).

10.6.1 The Autonomy Debate Revisited

Proponents of the autonomy of artefacts often base their claim on the difficulty of predicting or directing the effects of artefacts. Verbeek even argues that they are "fundamentally" unpredictable; that is his main reason for attributing some form of autonomy to artefacts. There are indeed limits to our foresight and to our control. It is our contention that considering artefacts (or associations) to be agent-like entities merely re-phrases the riddle in metaphorical terms and does not help elucidate it. It is more helpful to look in detail at the complex ways in which artefacts influence action schemes. As we have seen, these schemes are the result of the mutual interactions between the intentional, physical, and social contexts. The very complexity of this interaction is, we contend, what makes it so difficult to predict the effects of artefacts.

Designers and engineers have to confront this complexity. They need to know (in so far as it is possible to know) how these frameworks jointly shape the action schemes of potential users. The intentional make-up of users is notoriously difficult to anticipate, and the effects of the social context on the action scheme are often far from obvious. Artefacts may come to have effects very different from those originally intended by their designers. For instance, energy-saving light bulbs were introduced to reduce the overall consumption of energy, but these bulbs seem to have encouraged people to change their behaviour; the availability of the new bulbs has led many to keep lights on longer than previously. It was wrong to assume that the new bulb would be neutral with regard to people's behaviour. This becomes apparent when we analyse the bulbs in terms of the action scheme. The previous option 'to leave the light bulb switched on' was not very attractive, because it was costly. In the new scheme the energy-saving light bulbs changed the attraction of this option because it became a cheap alternative; so people were no longer so bothered about switching off the lights. This unintended effect of the new bulbs can be best explained by regarding it as an altered action scheme that had not been properly anticipated.[26]

The unpredictability phenomenon is not unique to artefacts. We encounter the same difficulty when we look at other ways of affecting human behaviour. Politicians, for example, are no better off when they want to influence people using law, sanction, or propaganda. Chamberlain's famous claim that he had secured 'peace in our time' at the time of the signing of the Munich Agreement revealed a poor understanding of Hitler's action scheme (i.e., the options that were attractive *for Hitler*). Churchill seemed to have grasped Hitler's scheme much better. But should we blame Chamberlain? It is always easier to explain the choice of an action

[26] What would have been the right way to make people actually save energy? It would have been to increase the attractiveness of the action 'switch the light bulb off', for example by environmental education (to create an incentive to save energy), or even a rather drastic law banning excessive illumination of houses (with legal sanctions making it unattractive to leave lights on). One could also design smart light-bulbs which switch off automatically if no one is in a room. In that case the 'leave the light bulb switched on' option would simply be removed from the action scheme.

ex post than to foresee the deliberative process leading to them *ex ante*. To make matters worse, even if Chamberlain had had a more realistic grasp of Hitler's personality, it would have been very difficult for him to take steps to allay the actions of a maniac. The directing of future actions not only requires a profound understanding of the relevant options for action at a certain point in time but *also* of how it appears to the agent and, further, counterfactual knowledge of the possible modified schemes in which the desired action is a very attractive option. Such knowledge is often not available.

To conclude, we do not need to attribute mysterious forms of agency to artefacts in order to account for the unpredictability of their effects. We maintain that such unpredictability is largely due to the fact that artefacts influence action schemes through various contexts in highly complex ways.[27]

10.6.2 The Moral Relevance Debate Revisited

If there are no compelling arguments for attributing agency to artefacts then the same is true of moral agency. The Moral Responsibility Thesis finds no support. The other extreme standpoint in the debate, the Neutrality Thesis, which holds that artefacts are merely neutral means to the ends agents pursue, seems also implausible. Because of their effects on the actions of users, artefacts can hardly be denied some moral relevance. They are able to change our relationship to the world in quite fundamental ways and to introduce (potentially) serious moral consequences which go beyond those of their designers' intentions. The challenge, then, is to formulate an intermediate position that attributes moral relevance to artefacts without making them morally responsible or morally accountable for their effects.

Looking at action schemes and second-order responsibility (i.e., attributing responsibility to *human* agents for changes in the action schemes of agents) allows us to analyse artefacts' moral relevance more precisely. There are many ways in which we can shape action schemes. Introducing a traffic rule, for example, is an institutional way of changing action schemes. Putting a thief behind bars is a physical way. Convincing somebody to stop smoking is an intentional way. Artefacts also alter action schemes, and this explains their moral relevance. That is why the design, production, introduction, and use of artefacts brings with it second-order responsibility for the effects artefacts have on the action schemes of agents. This responsibility is often indirect and partial since the causal chain leading to these effects is complicated and involves other agents as well.

[27] There is a *caveat*. Certain high-tech artefacts are increasingly acquiring properties that are agent-like. In future there may be a need to develop agency-concepts that reflect these properties. A modern computer may pass the Turing test under certain well-defined conditions. A missile may be said to have goal-directed behaviour. Research into artificial intelligence aims at developing non-human agents. Whether or not we will attribute agency, or even moral agency, to artefacts or systems in the future remains an open question. This issue should not, however, be confused with the issue of unpredictability discussed in this paper.

New options which artefacts open to us have sometimes been the topic of ethical debate, in, for example, spectacular cases relating to nuclear devices. The action scheme perspective allows us to evaluate these effects in ordinary cases and in a much more systematic way. It will shed new light on the responsibilities engineers, researchers, developers, designers, and the producers of all sorts artefacts have. These parties usually limit their responsibility to the well-functioning of the artefact together with accounting for the risks involved in using the artefact on a certain scale. They do so by offering a use plan.[28] This is usually a rather narrow set of instructions that need to be followed in order to realize the function of the artefact. Such a use plan is different from, and much more limited than, an analysis of action schemes. Focusing on action schemes broadens the responsibility issue considerably; it implies that engineers not only have first-order responsibility for the well-functioning of artefacts, but also that they have second-order responsibility for how such artefacts may influence action schemes.

10.7 Analysing Action Schemes: Applications from Architectural Design

Let us turn to architectural design as an example of the explanatory and evaluative use of action schemes. The point is to demonstrate that our approach allows for a detailed ethical appreciation of architecture which includes hitherto much-neglected aspects of moral relevance. It enables us to make ethical judgements on the basis of architecture's influence on human behaviour, and it allows us to critique existing buildings (and also architectural plans), and is therefore a useful tool in the hand of designers who desire to design and build in an ethically better way.

Ethics of architecture is, admittedly, a young branch of ethics, but is often severely limited in scope; it focuses mainly on environmental issues.[29] In particular, the ecological crisis that came to people's awareness in the 1970s has triggered concerns about the 'ecological footprint'[30] of architecture and has given rise to debates about sustainable ways of building – a movement that has gained new importance because of concerns about global climate change. After all, the impact (on the environment and climate) of *building* is hardly equalled by any other human activity.[31]

But there is much more to be said about the moral relevance of architecture.[32] The way in which we build is of great importance to human well-being (safety, health, psychological well-being etc.), and provides cultural and symbolic meaning

[28] See Houkes and Vermaas (2004).

[29] See, for example, the important collection of articles by Warwick Fox (2000).

[30] Rees (1992).

[31] See Illies (2009b).

[32] For this see also Illies and Ray (2009).

that can be of ethical interest. It also influences and guides human behaviour. The cultural theorist Edward Hall was one of the first to emphasize this aspect and goes so far to claim (in a title co-authored with Mildred Reed Hall) that the built environment itself is "a greater determinant of behaviour than personality."[33] In what follows we will turn to this influence in order to show the applicability of our approach. The action scheme approach can make this effect on behaviour more obvious. It enables us to analyse the options for action a building offers in a systematic way – and also their attractiveness (at least for a specific group of users at a certain point in history).[34]

Let us begin by looking at some examples of building's influence on human behaviour.[35] Small well-lit rooms with comfortable furniture, for example, can support social exchange in residential accommodation for the elderly. In 1957 the psychiatrist Humphry Osmond (1917–2004) labelled this capacity "sociopetality" and characterised it as "that quality which encourages, fosters, and even enforces the development of stable interpersonal relationships such as are found in small, face-to-face groups."[36] Another example is provided by A.W.N. Pugin's designs for English convents: designs which break with the historical tradition. Rather than having square cloisters or a hall in the centre, as in medieval convents, his buildings contained exaggerated, long, internal corridors that meandered through the building, sometimes even demanding that the residents go forward and backward on different floors before reaching a room. What seems an unnecessary and extensive circulation space for low-budget buildings is powerfully explained in an analysis by Timothy Brittain-Catlin: Pugin suggests a certain ideal of life (constituted by certain actions). This ideal had been proposed in the Catholic revival of his time, most importantly emphasising that one should separate different activities (praying, eating, social exchange, etc.) in order to do them more self-consciously. And it is this way of life (and its accompanying action scheme) that is encouraged by the design.

The architect and city planner Oscar Newman observes in a study of housing in New York that high-rise apartment buildings occupied by many people show a higher crime rate than lower buildings. He explains it by the fact that in the low-rise buildings, residents show a greater personal responsibility for their environment. Based upon this research, Newman develops the concept of *Defensible Space* (1972) suggesting a form of crime prevention (and increased public health) through community design.[37]

---

[33] Hall and Hall (1975, 42).

[34] It has been debated whether architecture can actually influence the behaviour of its users and inhabitants in any significant way. Alice Coleman (1990), on one hand, argues for a strong influence of urban structures upon behaviour – similar arguments are made, at least implicitly, by many defenders of New Urbanism. Others, on the other hand, disagree, and consider social factors more important than physical ones. Bill Hillier (1986) and others argued that many of Coleman's results were statistical artefacts and that the same forms might have been perfectly suitable for different inhabitants. For a general overview see Mikellides (2007).

[35] Brittain-Catlin, T. (2006).

[36] See Osmond (1957).

[37] http://www.defensiblespace.com/art.htm (accessed September 2011). It should be added that the well-documented physical and mental illnesses associated with poorly designed social housing

In all these examples, the chosen structural features of the built environment (the shape of rooms, form of cloisters, etc.) make the occupants behave in certain ways; or, at the very least, they incline a person to one behaviour rather than another. With the help of action schemes, we can account much more precisely what these effects on human behaviour are (for users of a certain type, time, and culture etc.). In order to do so we need to look at the two aspects separately:

(a) *What options for action are offered by the architectural structures?* A door between two rooms, for example, enables occupants to have encounters while walls "wall" them off. A room without windows does not allow users to work there without electric light. And a highway through an urban settlement will limit walking options for pedestrians but will provide new options for quick access by car. An action scheme analysis of a building will have to list relevant options for actions that the built space provides.
(b) *Which options for action are made attractive and which are made unalluring by the architectural structure?* Because they are less mobile, and perhaps burdened with various physical infirmities, many elderly people feel vulnerable, so that they prefer to be in smaller rooms rather than in big halls. Thus the option of gathering in a small room and talking to each other is much more attractive than gathering in big rooms.[38] Any such analysis must obviously take the specific features of the user into account; a place that is attractive for a gathering of elderly people might be of little interest for a student-party or a family assembly with children. The range will vary. Some features might add to the attractiveness of a certain activity for all possible users (a library must be well lit, to allow people to read, irrespective of their age, sex, religion etc.), while others are dependent on the cultural setting (today's students might find it impossible to work in a library without Internet access) or on age, traditions, health (can people in wheel-chairs access the library?), family structure, or even individual priorities (Jane Austen was happy to write her novels on the kitchen table, but Virginia Woolf needed a room of her own). And even though it is hard to quantify attractiveness we *can* ascertain whether a certain room makes it easy or awkward to perform particular actions.

Let us look at the examples again. With Pugin's buildings we could list which rooms are accessible and from where; and we can also see what actions should be performed in which rooms. Such a list might then look like:

---

projects are often caused primarily by economic and social deprivation, the impoverished quality of the architecture merely illustrating the problem and inevitably compounding it.

[38] New kinds of behaviour can also be opened up in subtle ways – for example, by making people think about new issues, or about old ones in new ways. Frank Lloyd Wright, for example, designed most of his so called *Prairie Houses* around a fireplace or hearth to express family life and its values, especially unity, harmony with nature, and the simple life. Expressed in terms of action schemes we might say that having such a fireplace in a house can lead to different kinds of behaviour by fostering the attractive option of sitting together around a fire-place. And this might trigger reflections about the fundamentals of family life etc.

Room A offers options:

1. direct access to rooms B, C, and F; slow access from E (long corridor) etc.
2. ø-ing in the room is attractive (room size, lay-out etc. encourage people to ø).

Such an inventory allows for an evaluation on the basis of a list of desirable actions that should be performed easily in these rooms. If it is positive for people to ø in room A, then it is a good room according to this standard. If there is a moral demand to ø in room A then it is morally praiseworthy to design room A in this way.

If, for example, the ideal of Catholic revival is to become more conscious of what you are doing by keeping different activities apart, then separate rooms for gathering and work, and possibly long passages between them, makes the option of doing so more attractive. In the spirit of the revival movement, it is a good building because behaviour is guided in the right direction. This example might be regarded as morally neutral – at least it needs further argument to acknowledge the standards of Catholic revival as morally demanded. But when we look at the retirement-home, we probably agree that it is morally demanded of us to make the elderly feel at ease in their home and to give them the chance of social exchange. Constructing the built environment in such a way that there are action schemes with attractive options for gathering is, then, a moral quality (and even requirement) of such a building.

For Defensible Space studies, the action scheme would also be useful as a tool for identifying general patterns. One could, for example, make a matrix with the attraction of certain actions in certain settings for specific groups and use them systematically for the evaluation, but also for the planning of settlements. After all, action schemes are not merely a tool to evaluate given structures according to some standard, they allow also to compare buildings and to make design choices.

Let us look, for example, at the infamous Pruitt-Igoe housing project for the socially disadvantaged, designed in 1951 by Yamasaki, the architect of the former World Trade Centre. He constructed 11-story buildings which totaled 2,870 apartments. They were originally heralded for their innovations. But later on, their 'impersonal structures' have been blamed for having generated vandalism and crime – so much crime, in fact, that no one wanted to live there. The complex was demolished after just 20 years, a moment famously baptized by Charles Jencks as 'the death of modern architecture', arguing that this architectural style was unable to provide livable environments (at last for poor people who could not make sense of the architectural language used).[39] It seemed that for the people living there (mostly extremely poor African-Americans), the buildings looked like prisons and they could never feel at home there or develop a sense of community. Others, however, have argued that the situation that ultimately led to depleting the houses and demolishing them had nothing to do with the architectural style; but was a consequence of the mediocre quality of the buildings in combination with "the interaction of paternalistic regulation, racist segregation, and family-destroying welfare law [that] made the project itself an unsafe, unfriendly environment."[40] An action

[39] Jencks (1987).

[40] Birmingham (1998).

scheme analysis allows us to compare systematically different structures, or similar structures in different architectural styles, that are inhabited by comparable groups, ideally living under the same laws and regulations, so that we can specify the contribution of the built environment to their behavior.

An action scheme analysis might also be helpful in expanding Newman's scheme. One of his principles of "defensible" architecture is that buildings and structures should be suited to different resident groups so that they (given their ages, habits, culture, socializing proclivities, family-structure etc.) are able to control and utilize them optimally. This requirement can be combined with an action scheme analysis by asking systematically which options are attractive for a certain group of users (a differentiation Newman had neglected). Young families, for example, find it more attractive to use open common ground between apartments as a playground for their children while elderly people desire more quiet areas. Such an approach might give rise to insights far beyond what Newman envisioned in his crime-prevention analysis; it might actually help us to build an "architecture for happiness" (to borrow a title from Alain de Botton); and happiness is, at least in some classical ethical systems, a thing to be encouraged.

Let us finally turn to what is as yet unbuilt – and thus to architects, contractors, and all those who have influence upon the design and structure of the built environment. If we take the moral relevance of architecture's influence on human behaviour seriously, it will obviously have far-reaching implications for the second-order responsibility of designers. Architects and planners should build with the awareness of the possible effects on the behaviour of residents and users. The action scheme analysis provides knowledge that can be used systematically for this purpose; architects could use approved sets of attractive actions (expressed in standard action schemes) as a kind of blue-print for their buildings. If they want to build a public square, they should investigate which actions it should allow – and whether the planned action scheme is likely to make the (ethically, socially etc.) desired actions easy. This will not by itself constitute a proposal for a specific design or architectural style; in most cases there will be many possible ways to create good action schemes. (Consider Siena's *Piazza del Campo* and the *Place des Vosges* in Paris – very different ways of creating a highly attractive set of social options.)

It should be added that this is a long term task and not easily achieved. A lot of empirical studies will have to be performed to establish a useful list of action schemes for standard architectural challenges – but any such general list will have to be completed by looking always at the particular situation.[41] Furthermore, the mere investigation of expected action schemes does not suffice to tell the architect how to build; there must always be space for a critical perspective within architecture, and the possibility of opening new ways, not yet envisaged in any known action scheme. After all, it is very difficult to say what actions should be promoted by architecture, and what means are morally acceptable in the pursuit of these

[41] Some work in this direction, though without the concept of action schemes, has already be done in, for example, the context of "evidence based design". It is, however, very much limited to hospitals, and looks at very few options for action.

actions. Both reflections will be very difficult: a normative theory for architecture is still needed, a theory that provides well-justified ideals, values, or goods for the different areas – and a theory that makes suggestions on how to deal with conflicting demands, both ethical and other, in specific cases. It is not clear yet what this theory might look like.[42]

Action schemes, however, promise to be at least a first step; they can provide a useful tool for analysing the actual effects of buildings on users in a way that allows us to grasp this much more precisely than other approaches. Taking action schemes seriously will also make it more obvious in which ways building is an ethical task. Architects have second-order responsibility to look at building-designs with regard to their effects on the action schemes of future users.

10.8 Conclusion

In this chapter we introduced the notions of action scheme and second-order responsibility in order to understand and evaluate the moral relevance of technological artefacts. Using these conceptual tools we then developed a position within the Autonomy Debate and the Moral Relevance Debate which avoided the problems associated with current views.

Our position seems plausible in the light of crucial criteria. Firstly, it allows us to address the profound effects that technological artefacts have on human beings, including on our perceptions and actions as described by Ihde and Verbeek. Secondly, this account, and these categories, support important concepts and distinctions which have shown their usefulness in moral debates. Latour's use of the general notion of 'actant' as a replacement for agent, for example, blurs these distinctions and makes it impossible to reconstruct relevant differences between human agents and artefacts in ethical analyses.[43] Verbeek's extension of the notion of moral agency to artefacts (or hybrids) is equally problematic. Ultimately it is our contention that human agents remain morally responsible. Thirdly, the position avoids the Neutrality Thesis. We agree with many authors who claim that artefacts do have moral relevance. Fourthly, the account is applicable to particular cases, as we have seen in our discussion of architecture, thus allowing one to understand the effects of a particular artefact in a specific context. It can also be used to analyse the moral responsibility of engineers and designers. Finally, the account is not biased towards (let alone based upon) any specific theory of action or ethical theory. It is perfectly general and can be combined with specific ethical analyses. On the basis of these criteria we conclude that our position is more promising than the rival positions discussed.

[42] On the problems of a general philosophy of architecture see Illies (2009a) and Illies and Ray (2009).

[43] We should mention here that the notion of actant was not developed for an ethical analysis of the role of technology but for purely sociological analysis.

Acknowledgments The authors would like to thank Marcus Duewell, Stefan Koller, Peter Kroes, Martin Peterson, Andreas Spahn, and the participants of the NIAS workshop on Moral Agency and Technical Artefacts in Wassenaar/The Netherlands for their stimulating comments on earlier versions of this paper.

References

Aristotle, N. (1998). *Nicomachean Ethics* (J. L. Ackrill & J. O. Urmson, Ed., D. Ross, Trans.). Oxford: Oxford University Press.

Birmingham, L. (1998). Reframing the ruins: Pruitt-Igoe, structural racism, and African American rhetoric as a space for cultural critique. Positionen 2.2. (1998). http://www.tu-cottbus.de/theoriederarchitektur/Wolke/X-positionen/Birmingham/birmingham.html

Brittain-Catlin, T. (2006). A.W.N. Pugin's English convent plans. *Journal of the Society of Architectural Historians, 9*, 356–376.

Coleman, A. M. (1990). *Utopia on trial: Vision and reality in planned housing*. London: Hilary Shipman Ltd.

Fox, W. (Ed.). (2000). *Ethics and the built environment*. London: Routledge.

Hall, M. R., & Hall, E. T. (1975). *The fourth dimension in architecture. The impact of building on behaviour*. Santa Fe: Sunstone.

Hillier, B. (1986). City of Alice's dreams. *Architect's Journal, 9*, 39–41.

Houkes, W. N., & Vermaas, P. E. (2004). Actions versus functions: A plea for an alternative metaphysics of artifacts. *The Monist, 87*, 52–71.

Ihde, D. (1979). *Technics and praxis*. Dordrecht: D. Reidel Publishing Company.

Ihde, D. (1991). *Instrumental realism: The interface between philosophy of science and philosophy of technology*. Bloomington: Indiana University Press.

Illies, C. (2009a). Philosophie als Architektur – Philosophie der Architektur. *Aus Politik und Zeitgeschichte*, *25*, 3–6.

Illies, C. (2009b). The built environment (section technology & environment). In J.-K. Berg Olsen, S. Andur, & V. F. Hendricks (Hrsg.), *A companion to philosophy of technology* (pp. 289–294). Oxford: Blackwell.

Illies, C., & Meijers, A. (2009). Artefacts without agency. *The Monist, 92*, 420–440.

Illies, C., & Ray, N. (2009). Philosophy of architecture. In A. Meijers (Ed.), *Philosophy of technology and engineering sciences* (Handbook of the philosophy of sciences, Vol. 9, pp. 1121–1174). Oxford/London: Elsevier Science.

Jencks, C. (1987). *The language of post-modern architecture* (5th ed.). New York: Rizzoli.

Koller, S. (2011). *Action schemes and agent autonomy*. Part I. Unpublished manuscript.

Kroes, P., & Meijers, A. (Eds.). (2006). *The dual nature of technical artefacts* (Special issue of studies in the history and philosophy of science, Vol. 37, pp. 1–158). Amsterdam: Elsevier.

Latour, B. (1987). *Science in action: How to follow scientists and engineers through society*. Milton Keynes: Open University Press.

Mikellides, B. (2007). Architectural psychology 1969–2003, theory, practise and education. *Brookes eJournal of Learning and Teaching.* http://bejlt.brookes.ac.uk/article/architectural_psychology_19692007/

Osmond, H. (1957). Function as the basis of psychiatric ward design. *Mental Hospitals, 8*, 23–29.

Peterson, M., & Spahn, A. (2011). Can technological artefacts be moral agents? *Science and Engineering Ethics, 17*(3), 411–424. Online first publication 7 October 2010.

Rees, W. E. (1992, October). Ecological footprints and appropriated carrying capacity: What urban economics leaves out. *Environment and Urbanisation, 4*(2), 121–130.

Selinger, E., Aguilar, J., & Whyte, K. (2011). Action schemes: Questions and suggestions. *Philosophy and Technology, 24*(1), 83–88.

Sen, A. (1982). *Choice, welfare and measurement*. Oxford: Basil Blackwell.

Sen, A. (1992). *Inequality re-examined*. Oxford: Clarendon.
van den Hoven, J. (1998). Moral responsibility, public office and information technology. In I. T. M. Snellen & W. B. H. J. van de Donk (Eds.), *Public administration in an information age: A handbook* (pp. 97–112). Amsterdam: Ios Press.
Verbeek, P. P. (2005). *What things do*. University Park: Pennsylvania State University Press.
Verbeek, P. P. (2008a). Obstetric ultrasound and the technological mediation of morality: A post-phenomenological analysis. *Human Studies, 31*, 11–26.
Verbeek, P. P. (2008b). Morality in design. In P. Vermaas et al. (Eds.), *Philosophy and design. From engineering to architecture* (pp. 91–103). Dordrecht: Springer.
Wilson, G. (2007). Action. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy.* http://plato.stanford.edu/archives/spr2008/entries/action/. Accessed 8 Sept 2011.