Moral Mid-Level Principles in Modeling

Sven Diekmann (s.diekmann@tue.nl)

Abstract

Modeling plays a substantial role in decision-making and therefore many models raise moral issues. Modelers, especially in operations research, become increasingly aware that this poses new problems to modeling. This paper addresses two questions: (1) How do moral issues in modeling arise? (2) How can these moral issues be resolved? In this article I trace the origins of moral issues in modeling back to (a) the beneficial or harmful consequences of using a model, to (b) the sometimes doubtful purposes of a model, and to (c) the target systems that include concerns about good human behavior. Furthermore, I propose a framework for good modeling practice in the form of four *moral mid-level principles*. The first is the principle of *transparency*, which codifies the obligations to explain, describe, and document the structure, assumptions and further properties of the model. The second is the principle of *integrity*, which refers to the responsibility of modelers to follow professional and scientific standards. The third is the principle of *comprehensiveness*, which stresses that modelers should acknowledge all moral concerns of all stakeholders thoroughly. The fourth and last one is the principle of *evaluation*, which states that models should evaluate all moral issues with respect to their consequences and implications. In summary, the article provides tools to assess whether a model raises moral issues, and it develops a basic framework for solving practical moral problems.

1. Introduction

Models, especially in operations research, often play an important role in decision-making processes. The construction and use of models can therefore raise explicit or implicit moral issues. Operations researchers become increasingly aware of this problem.[1] However, the debate about how modelers should deal with moral issues is still in its infancy. Besides a recent approach that uses discourse ethics (Le Menstrel & van Wassenhove, 2004; Drake et al., 2011; Kunsch et al., 2009; Mingers, 2011), there are no systematic approaches for how modelers should tackle moral problems. In this article I seek to provide a basis for more thorough moral discussions and develop a framework for dealing with practical moral problems that arise in modeling. My aim is to address two fundamental questions: (1) How and when do moral issues arise in modeling? (2) How can these moral issues be resolved?

In operations research, and other disciplines such as engineering, several authors have touched upon a number of moral issues that arise in modeling. To name some, Wenstøp and Magnus (2001) describe a model for decision-support in the Norwegian HIV policy. Among other criteria, this modeling process had to address the extent of future HIV cases, privacy, general and personal anxiety, life quality reduction, and stigmatization of individuals and groups (Wenstøp & Magnus, 2001, pp. 61-62). Rabins and Harris (1997) describe modeling processes in engineering that resulted in fatal consequences. They conclude that every modeling process should respect moral standards such as the Golden Rule, informed consent, and utility. Kleijnen (forthcoming) names several morally problematic modeling cases, e.g. the assessment of a plant to store nuclear waste, the evaluation of a storm surge barrier, or a model that supports the search for explosive sea mines. Taket (1994) analyzes moral problems she encountered during her work as operations researcher, and proposes guidelines for morally concerned, critical (self-)reflection in modeling.

[1] Two milestones of this debate are Wallace (1994) and the special issue "Ethical Issues in OR/MS" in Omega, see Le Menestrel and van Wassenhove (2009). Wenstøp (2010) provides a comprehensive overview of the field of operations research and ethics.

However, so far there has been little discussion in the literature about how, why and when moral issues arise in modeling. In many cases it is left to the modeler's intuition to recognize moral problems. Furthermore, it is unclear how moral issues are connected to models. Indeed, some authors seem to assume that moral problems are external to the modeling profession. For example, Walker (2009) states that the best way to deal with moral problems is to follow the best practices of one's profession. While the emphasis on best practices is obviously justified, focusing exclusively on them presumes that moral problems do not differ from any other problem in modeling. In the following I show that this presumption is incorrect: By considering basic moral theories such as consequentialism, deontological ethics, and virtue ethics, we are able to see that the origins of the moral issues in modeling lie within the modeling practice itself. Of particular importance are the consequences of using the model, the purpose of the model, and the properties of its target system.

A further concern one can raise about the existing literature is the apparent lack of grounding in moral theory.[2] Even though many authors contribute helpful discussions of how practical moral problems could be resolved in particular cases, the lack of theoretical grounding hampers the development of more systematic solutions. The moral mid-level principles proposed in this article can function as a basis for systematic approaches. The first is the principle of *transparency*, which codifies the obligations to explain, describe, and document the structure, assumptions and further properties of the model. The second is the principle of *integrity*, which refers to the responsibility of modelers to follow professional and scientific standards. The third is the principle of *comprehensiveness*, which stresses that modelers should acknowledge all moral concerns of all stakeholders thoroughly. The fourth and last one is the principle of *evaluation*, which states that models should evaluate all moral issues with respect to their consequences and implications. The four principles are derived from practical experiences and solutions of professionals, but also grounded in moral theory. In the final section 5, I explain in detail how moral mid-level principles are employed.

[2] Drake et al. (2011), Kunsch et al. (2009), and Mingers (2011) are positive exceptions here.

2. The Sources of Morality in Modeling

Moral issues in modeling arise on a big scale (e.g. when a model affects governmental decisions), as well as in isolated situations (e.g. when a model influences the design of a particular artifact). In order not to overlook relevant insights, I will take a pluralistic perspective to moral theories: I assume that all commonly acknowledged moral theories are able to point out morally problematic or noteworthy issues. However, since the broadest three perspectives in ethics -- consequentialism, deontological ethics, and virtue ethics -- subsume most moral theories, I shall focus on these for the rest of my inquiry.[3] As mentioned in section 1, my analysis focuses on practical moral problems, as they arise in the work of operations researchers.

First consider the consequentialist approach to moral theory. Consequentialists focus on the consequences of using a model. The consequentialist key idea is that an action should be preferred just in case it leads to optimal overall consequences. For example, models used for safety in engineering typically have morally good consequences. Crash models in the background of computer simulations help to decrease human harm in an efficient and cheap way; their use has therefore good consequences.

That said the use of models can also have bad consequences, as shown by an unfortunate event during the first gulf war. On 25th February 1991, a patriot defense system failed to intercept a Scud rocket, due to misusage of the system and its underlying model. Tragically, 28 soldiers died. (Blair et al., 1992)

Whether such usage is an intended use or a misuse is less important from a consequentialist perspective. In consequentialism, the consequences of actually employing a model are more important than the model itself or the intentions it was built with. Indeed, the model itself can be treated as a black box: The usage of the very same model is morally good in case the consequences are optimal, morally wrong in cases of negative triggered consequences. It is therefore that the consequentialist perspective discovers the origin of moral issues solely in the consequences of the use of models.

[3] The idea to use a pluralistic approach is inspired by van den Poel (2001).

Deontologists focus on the action as such, without paying any direct attention to its consequences. In what follows I use Kantian ethics as a paradigmatic example for a deontological moral theory. The key idea in the Kantian theory is the categorical imperative. Kant offers two main formulations[4]: (1) "[A]ct only in accordance with that maxim through which you can at the same time will that it become a universal law." (Kant, 1998, p.31) and (2) "[A]ct that you use humanity, whether in your own person or in the person of any other, always at the same time as an end, never merely as a means." (Kant, 1998, p. 38) The first formulation is the principle of universality. Good actions should always be performed regardless of the situation or the agent. For example, deceiving is morally wrong. Although there might be circumstances under which deceiving leads to a desirable outcome, deceiving undermines the act of communication. If we would deceive on a regular basis, communication would lose its function of transmitting reliable information. The second formulation tells us that humans have a dignity, which we should respect in all our actions. For instance, employers should never see their employees as pure resources, but always acknowledge that they are also persons who have further, e.g. social, needs.

When applied to modeling, both formulations restrict the purposes for which models should be employed. Independent of the consequences of the model, the purpose determines the moral evaluation in Kantian ethics. For example, models should never be used to deceive: Models are supposed to be elucidating instruments. This elucidating nature is the value of and the reason for employing them. If models were used to deceive in general, they would lose their elucidating nature and there would be no reason for using models. Moreover, models should never support exploiting people, because this would ignore that people should (also) be treated as ends. Even in cases where the final consequences might lead to optimal benefits, the elucidating nature is still undermined, or the people are still exploited. Thus, a model with such a purpose would raise moral problems for advocates of a Kantian moral theory.

Probably one can be rather sure that most models are developed for good purposes. However, there are also some cases in which models were developed with morally wrong intentions. Such an example

[4] Kant claims that these are equivalent.

is described by Fienberg (1994). He describes the process of choosing between two rival statistical models for the 1990's US decennial census. A problem at that time was that minorities were undercounted. In order to solve this problem, new statistical procedures were introduced. Besides the statistical model of the US Bureau of the Census, there were additional statistical tools developed by an independent group of mathematicians. A model that incorporated these additional tools would had been more accurate, than the model that had been employed by the Bureau of the Census. Fienberg claims that unofficial political interest of a few powerful white Americans turned down the more accurate model. Given that such a claim is true, the purpose of the model was to distort information for the sake of partial interests. This is a kind of deceit and thus plainly wrong.

Virtue ethicists focus on the good life and the pursuit of achieving it. The notion of a good life, or "eudemonia", was of fundamental importance to ethics in ancient Greek philosophy. Eudemonia is defined as the state of complete happiness, and it is the moral goal of one's life; virtues are seen as the means to reach this state. In particular "[h]appiness is an activity of soul in accordance with perfect virtue" (Aristotle, 1980, [1102a]). Models can relate to the pursuit of a virtuous life by evaluating visions about it. They can interrogate which human behavior would lead to a desirable outcome, and thus state which virtues help reaching a good life. For example, many people see it as part of a good life to consume meat on a daily basis. However, models forecasted the environmental impact of such lifestyle and predict substantial negative effects, if all humans consume meat in such amounts. (Stehfest et al., 2009) Thus, meat consumption on a non-daily basis is more virtuous, and models help us to understand this.

When models evaluate visions of a good life, models need to represent or describe these visions. In the literature, the represented or described entity is often referred to as target system. Therefore, when one seeks to know whether a model raises moral issues under a virtue ethical perspective, one has to look at the target system of that model. Whenever the target system includes a potential way of living or how one could improve a current way of living, the model raises virtue ethical concerns.

If one turns to the question whether the target system raises good or bad moral issues, virtue ethics also provides an answer. When pursuing the state of complete happiness is morally good, then constructing and employing a model that helps to attain that state is praiseworthy. Where models do

not anticipate this perspective towards a good life, they are not morally praiseworthy. In addition, if the target system would be chosen in such a way that the model misguides from the good, constructing or employing the model would be morally wrong.

There are also moral issues, which are not related to what the model is about or used for, but which are related to the physical properties of models. For example, when animals are used as models for the human body. In these cases moral issues arise, because animals have intrinsic value. Such concerns are not only related to living beings; most entities of high value seem to raise similar problems. Think for example of a model that would destroy the Mona Lisa as a means for predicting aging processes of artworks. Such a model raises value conflicts and arguably moral issues. In this analysis I leave such issues aside, because these moral issues rely on physical properties, which is neither specifically about modeling nor the construction or usage of models.

3. Three Moral Dimensions of Modeling

In the foregoing section I have identified some of the most significant origins of moral issues in modeling. However, there is more to be said about how moral issues are connected to models and modeling practice. In this section, I fill this gap. Furthermore, I discuss some actual cases. The aim is to clarify a bit more how knowledge about the origins of moral issues helps to analyze real-world cases.

Philosophers of science have been taking a number of attempts to clarify what scientific models are, how they relate to scientific theory, and how they relate to the phenomena under research. For the discussion of models in application-oriented sciences, such as operations research, the model account of Morgan and Morrison (1999) describes the modeling process most accurately. Their account treats phenomena as important as theory for the construction of models. In short, they claim that models are autonomous "instruments" that "mediate" between theory and phenomena, such that most models embody parts of the theory as well as parts of the phenomena. When I refer to the target system, I precisely refer to the phenomena the model is related to in this way.

Even though the usage and the purpose for using a model are eventually up to the model-user, modelers can influence the consequences of the uses and the purposes of their models. The account of

Morgan and Morrison includes several degrees of freedom in the construction of models; on the one hand different kinds of representation, and on the other hand a variety of model functions. These degrees of freedom allow for influences on the construction of models toward desired design and intended applications, without harming the scientific character of models.

Often the notion of representation[5] in modeling refers to a kind of mirroring a phenomenon -- Morgan and Morrison take a different approach:

"Instead, a representation is seen as a kind of rendering -- a partial representation that either *abstracts from*, or *translates into* another form, the real nature of the system or theory, or one that is capable of *embodying only a portion of a system*." (Morgan & Morrison, 1999, p.27, my emphasis)

The kind of abstraction that is aimed on, the language into which it is translated, or the choices of embodied properties are established in the construction process of models. Similar claims hold for the function of models. Morgan and Morrison speak of several functions models can fulfill, viz. theory construction and exploration, measurement devices or measurement structuring, and design and intervention. (Morgan & Morrison, 1999, pp.18-25) All these degrees of freedom can be used for constructing models for particular uses and purposes. On the one hand, such construction should ensure that morally good uses and purposes are likely. On the other hand, modelers can construct models such that unintended consequences are virtually impossible, or that the model hardly serves unwanted purposes.

The three origins of moral issues -- use, purpose, and target system -- can be applied to analyze models respectively their potential of raising such issues. Table 1 summarizes the results for six cases, I analyze briefly in the following.

[5] Some modelers claim that their models, in particular agent-based models, do not represent anything. Such models can however be neglected, since relevance for decision, i.e. moral relevance, presumes some kind of representation. There are also other kinds of models, like the Ford Model T, fashion models, etc. Since those and other models are not related to representation in the sense we are concerned with here, they are excluded from the discussion.

Table 1: Use, purpose, and target system are the three origins of moral issues. The analysis of six cases reveals differing potentials for raising moral issues. Aspects that raise moral concerns are bold.

No.	Case	Consequences of the Use	Purpose of the Use	Target System
1	Scud interception	**failure to prevent death of 28 soldiers**	prediction of missile trajectories for intercepting them	trajectories of Scud rockets
2	GIS landscape modeling	**neglected or destroyed wetlands**	mapping of topography and water streams	landscapes
3	1990's US decennial census	**probably disadvantageous for minorities**	**biasing results in favor of a minority**	number and type of US inhabitants
4	forecasting the effects of human meat consumption on climate change	description of the expected climate effects of different diets	explore the consequences of diet changes	**human way of living respectively effects on climate change**
5	Eyjafjallajökull ash cloud prediction	**numerous unnecessary flight delays**	**interrogating flight safety**	ash cloud of a volcano
6	the harmonic oscillator	understanding of oscillations	description and understanding	linearly swinging masses

Example 1: The patriot failure during the first Gulf War is an example of a model with bad consequences. A patriot defense system was employed to intercept Iraqi Scud missiles. For intercepting, the patriot software had to calculate the trajectory of the missiles. A trajectory was modeled by using data of velocity and time. Time values were saved as real values with 24 bits. The higher the values of time, the fewer decimals can be respected. Since time was modeled with increasing numbers, starting from 0 when the software booted, longer periods of use led to less

precision. This poses no problem for the originally intended short-term use of patriots. However, during the first gulf war, patriots were used for several consecutive days. In consequence, they first became less efficient and after 20 hours useless. The discussed patriot system worked for 100 consecutive hours at the incidence of failure. (Blair et al., 1992) When we employ the moral perspectives, we see that the target system of the model is morally neutral. The purpose is military but can still be seen as morally less problematic, because the interception of missiles is a form of protection. The very problematic case is the use of the model; it raises severe moral concerns from a consequentialist perspective: Since 28 soldiers died, using this particular model (in this particular incidence) was morally wrong. Instead another model should had been used.[6]

Example 2: Jenkins and McCauley (2006) discuss a problem with Geo Information Systems (GIS) that leads to the destruction of environmentally important wetlands. In GIS, the landscapes are represented by cells on a grid. The height of each cell is represented as an integer number. The model makes two assumptions. (1) Each grid-cell that is higher or lower than all its surrounding cells is assumed to be noisy data or a rounding error. In these cases, the height is lowered or elevated to fit its surrounding. (2) Water streams are assumed to have fractal geometry. Thus, all water flow is restricted to this pattern. Both assumptions disregard environmentally important wetlands. Wetlands are (i) usually small lowered areas (often smaller than one cell) that are part of the water system on a seasonal or occasional basis, but (ii) do not fit into a fractal pattern. Unfortunately "[t]hese wetlands … provide important ecological services such as flood mitigation, groundwater recharge, nutrient processing, and habitat for unique flora and fauna." (Jenkins & McCauley, 2006, p. 278) Nevertheless, GIS are used for decision-making in geo-engineering. Thus, decisions about shaping and changing the landscape are ignorant to those environmentally valuable areas. As bad consequence wetlands are destroyed, which is a moral wrong.

Example 3: The case of the 1990's decennial census (Fienberg, 1994) as discussed in section 2, shows moral problems in the purpose of the model deployment. The model was chosen, because it benefits a minority of powerful people by inaccurately assessing the number of US inhabitants. Since

[6] It's a tragic fact that an improved model arrived at the failing patriot system the next day. (Blair et al., 1992).

the census models are supposed to be as accurate as possible about the actual number and type of US inhabitants, Fienberg describes a case of deceit. Following Kantian ethics, one cannot want all models to deceive, because this would counter the elucidating nature of modeling. Thus deceiving with models is a moral wrong. (Since the actual consequences of this model use are unknown, one can only say that they were probably wrong.)

Example 4: Stehfest et al. (2009) employ the Integral Model to Assess the Global Environment (IMAGE) to forecast the impacts of human diets on climate change. They compare five diet scenarios to explore how changes in the use of animal products affect land-use and the emission of greenhouse gases. Their reference scenario is an income-driven increase in per capita meat consumption without any counter-acting climate policies. They predict that already a change to the healthiest diet[7] of meat consumption has a significant positive impact on land-use and emitted greenhouse gases. Even though there study is meant to be purely descriptive and provides only scientific claims, it touches the moral question of how we should live.

Example 5: The volcanic eruption of the Eyjafjallajökull in April 2010 caused the cancellation of 108,000 flights, affected 10.5 million travelers, and raised losses of 1.7 billion dollars for the flight industry. It is highly contentious whether the decisions made were correct or an overreaction. (Budd et al., 2011) Here a weather forecasting model was an essential part of the decision-making process. The split into the three dimensions reveals the underlying conflict: The purpose of employing the model was passenger safety, and therefore morally good from a deontological perspective. On the contrary, using the model for decision-making was wrong from a consequentialist perspective. The consequences raised costs and much psychological distress, while the actual increase in safety is still highly doubted (Budd et al., 2011).

Example 6: The model of the linear oscillator is one of the fundamental models in physics. It describes the behavior of a mass point on a swing that is in linear movement. I include this example as a representative of models that are usually not perceived as raising moral problems. The linear

[7] For the healthiest diet Stehfest et al. refer to Willet, WC (2001). *Eat, drink, and be healthy: the Harvard Medical School Guide to healthy eating*. Simon & Schuster: New York.

oscillator has usually an epistemic purpose; it discusses no human habits, but a physical object (a swinging mass); and although a use with morally wrong consequences is surely imaginable, by far most of its uses in physics do not directly touch on benefits or harms.[8]

The analysis of each case is rather straightforward and, since I kept myself brief, might miss out some details. Nevertheless the analysis fulfills its purpose to systematically approach morality in modeling and allowing modelers to question for moral issues related to models: (1) Can the consequences of using the model be morally good or bad? (2) Is the purpose for employing the model good or wrong? (3) Does the target system of the model include matters related to the human way of living?

4. Moral Mid-Level Principles in Modeling

In this section, I propose moral mid-level principles as a method for dealing with moral problems in modeling. Moral mid-level principles contrast to moral theories in their range of applicable cases. Moral theories claim to be generally applicable to most situations, while moral mid-level principles[9] are specialized prima-facie principles that hold only within a specific field. Therefore moral mid-level principles in modeling are designed to solve the moral problems respectively the particular circumstances of modeling.

In every field, professionals develop and practice specific norms and codes of conduct. For employing moral theories, one would have to abandon those norms and codes. Such abandonment presumes superiority of moral theories, and thus ignores the professionals' moral abilities. Moral mid-

[8] There are physical models that touch upon moral questions. For example models in quantum mechanics raise doubts in causality and thus in human accountability. (Thanks to M. Le Menestrel for pointing this out.) However, these concerns are beyond the focus on practical moral problems, taken in this article.

[9] They are called *mid-level* principles, because they stand in between unconditional principles and context-specific obligations. Unconditional principles, e.g. impartiality, are supposed to hold under all circumstances in any case. Obligations (e.g. "One should be quiet during a conference talk.") hold only in particular situations under specific circumstances. The obligations change when the situation or the circumstances change (e.g. "One should be talkative and raise questions *after* a conference talk.").

level principles however are based on these field-specific norms and codes. In addition to the field-specific basis of moral mid-level principles, I nevertheless refer to moral theories in my arguments. This reference makes the principles plausible from a general point of view, and, in consequence, strengthens the support of the moral mid-level principles. The methodology of moral mid-level principles goes back to Beauchamp and Childress (2008). Since moral mid-level principles are field-specific, I only use their concept of mid-level principles, not the particular moral mid-level principles that they have developed for biomedical contexts.[10]

It is difficult, if not impossible to derive the particularly best actions for a given situation from moral mid-level principles; neither is this the aim of this article. Rather the moral mid-level principles should demarcate a space of morally appropriate actions. It might be that there are more mid-level principles than the four presented here. Such additional mid-levels principles would not contest my claims, because my principles should provide orientation instead of final answers. Rather, additional principles would add precision to the approach. Finally, many moral mid-level principles are partly epistemically justified. Despite this, the mid-level principles are still moral principles, because epistemic motives -- such as accuracy, simplicity, or theoretical consistency -- do not suffice to support the mid-level principles presented here.

4.1 Transparency

The principle of transparency is grounded in the fact that no model result can be understood without some clarification of how the model and its result relate to the world. For example, there are many models that interrogate to which extent energy policy should aim on the use of coal, gas, nuclear or renewable energies. Several studies investigate this issue, employing different models with varying assumptions and basing on different data. How these different studies come to different and sometimes contradicting results is discussed in Sundqvist et al. (2004). They compare models about estimations of external costs for a variety of energy sources. The assumptions of the models used in

[10] Beauchamp and Childress (2008) promote Respect for Autonomy, Beneficence, Nonmaleficence, and Justice as moral mid-level principles in biomedical contexts.

the estimations differ in, (1) focus on energy technologies, (2) local characteristics of the researched sites, (3) scope of researched external cost, (4) implicit trade-offs among environmental effects, and (5) different methods to deal with complexity or uncertainty. (Sundqvist et al., 2004, pp. 234-235) Nevertheless, given the particular assumptions all models are scientifically feasible. However, all models are only valid under respectively these characteristics. For instance, many models are developed for Europe. Employing them for tropical regions certainly leads to inappropriate results and environmental damages. (Sundqvist et al., 2004, p. 236)

Such inappropriate results can cause misusage. Modelers or model-users may employ a model because they appreciate its result; not its appropriateness. On the one hand, this would be, an epistemic problem, because pseudo-knowledge would be generated. On the other hand, it also poses a moral threat. Wrong models can intentionally be used to deceive or unintentionally lead to bad consequences. In the former case, the moral wrongfulness is obvious; in the latter case, one-sided decision-making is likely to raise bad consequences, such as unnecessary harm.

Fulfilling *transparency* means to state assumptions clearly and to make the design of a model understandable to model-users. That means, a model can be called transparent when the model-user knows how the model works, which assumptions it involves, on which problems it can be applied, and which restrictions the model makes.[11] In the literature transparency is widely accepted. For example, the contributors to the book Ethics in Modeling "agree on some aspects of ethical conduct, such as … to make clear to the model-user what the model can do and what its limitations are." (Wallace, 1994, p. 8) Ross and Harris (1994) see the pro-active education of the model-user as one of the main responsibilities of every modeler. Besides providing thorough documentation and encouraging model-users to ask questions "[a] model builder has a professional responsibility to explain not only the strengths of a particular model but its inadequacies ..." (Ross & Harris, 1994, p.161). Rabins and

[11] I do not advocate that model-users should have the same knowledge about the models as the modelers. That would imply that only modelers are good model-users, which is not feasible in a decision process. Rather, transparency wants the modeler to communicate actively, and to be open for doubts and questions of model-users, such that the model-users do not treat the models as black-boxes (Ross & Harris, 1994).

Harris (1997) see informed consent as one of the major moral principles in modeling. Drake et al. develop a stakeholder approach for socially responsible modeling, based on discourse ethics. In their discussion, they claim that transparency is a duty in modeling: "… the way in which models are created, what problems they address, and how they are implemented should be made available to stakeholders of the research activity." (Drake et al., 2011, p.10)

Besides the agreement in literature, there are also moral arguments for pursuing transparency. Transparency is supposed to provide relevant information to model-users for making rational decisions. From a Kantian viewpoint, acting rationally entails acting morally. Therefore providing relevant information is a moral obligation for modelers, when treating stakeholders also as ends. Also the literature refers to obligations that ensure autonomy. For example by explaining model-users the strengths and inadequacies, as Ross and Harris (1994) suggest; or by informed consent, as suggested by Rabins and Harris (1997). The stakeholder approach, as suggested by Drake et al. (2011), goes even further. Here, not only the agency as such should be respected, but the interests of all affected people. Even though Drake et al. give a discourse ethical justification for their approach, one can easily provide a consequentialist justification. One may argue that a stakeholder approach promotes optimized maximized beneficial outcomes. Since stakeholder approaches presume a form of transparency, a consequentialist argument for transparency is that it is a precondition to incorporate all interests for maximizing benefits.

4.2 Integrity

If a modeler has integrity, his work follows professional and scientific standards. Integrity is most threatened, if one-sided interests influence central modeling decisions. The case described by Fienberg (1994) (example 3), about the 1990's US decennial census is an example of lacking integrity. Here a political decision favored a less accurate model in order to give benefits to a particular group of people. Fienberg states that there were no epistemic reasons for this decision. Rather he claims that profit was the motivating interest. In addition, he argues that for reasons of equality and higher epistemic accuracy the independently developed model should had been employed instead. Professional and scientific standards hamper such deficits. In this case, professional

standards, e.g. objectivity, or scientific standards, e.g. optimized accuracy or independence from interests, would have led to the usage of the right model.

To be explicit, having *integrity* means that a modeler should follow professional standards, even in situations where it is compelling to follow partial interests or to neglect rules. The literature mentions integrity not explicitly, but many claims can be summarized under it. First of all, there are ethical codes of different professional organizations. These codes set explicit professional standards. Gass (2009) gives a good summary of those. Professional standards include norms of how to perform one's work, as e.g. "Military OR professionals must strive to be … [t]ruthful, complete and accurate in what they say and write"; professional standards include norms about social obligations, e.g. ACM members will "contribute to society and human well-being"; and professional standards include ethical guidelines about the personal responsibilities of modelers, e.g. IEEE members agree "to accept responsibility in making engineering decisions consistent with safety, health, and welfare of the public". (Gass, 2009, pp.1046-1047) As repeatedly criticized by Gass, not all operations research organizations have a professional code yet. One attempt to install such a code has been made by Brans (2002), called the Oath of Prometheus. The Oath is addressed to different fields of operations research and includes professional, scientific, and moral norms. Besides Gass and Brans, also Walker is a strong proponent of professional standards (cf. Walker, 1994, 2009). He states that modelers "will be acting in an ethical manner if they apply the general accepted best practice of their profession." (Walker, 2009, p. 1051)

If a modeler follows professional standards, and if the model fulfills scientific standards, the model and its results are legitimated by these standards. On the one hand, standards have an epistemic value. Professional and scientific standards guarantee scientific rigor, and scientifically justified results. On the other hand, standards serve a moral purpose. Professional codes formulate a minimum of moral standards. By following these codes, modelers incorporate these moral standards into their work. Furthermore, professional standards enforce morality in conflict situations. First, they give orientation about what is right or wrong. Second, they provide a well-accepted argument for maintaining minimal moral standards, namely the role-responsibility of a professional.

4.3 Comprehensiveness

Comprehensiveness refers to the incorporation of all moral aspects that contribute to the performance of the model. Even though comprehensiveness is partly an epistemic concern, it has a moral dimension as well. The case of Geo Information Systems (GIS), of Jenkins and McCauley (2006) (example 2), shows the importance of moral comprehensiveness. Wetlands are crucial for environmentally sustainable decisions. They should thus be part of most decision-making in geo-engineering. So far however, the GIS are only constructed for high topographic accuracy. They neglect the environmental value of the landscape. Therefore, from a moral point of view, the GIS are not comprehensive. Jenkins and McCauley suggest solutions of how to include wetlands into the models. For example they suggest adding information about which cells have been elevated, or refining the assumption such that wetlands are not sorted out any more.

The principle of *comprehensiveness* summarizes the obligations of a full-fledged consideration of all moral issues that are connected to the decision alternatives under scrutiny. As the example suggests, such comprehensiveness does not aim on epistemic accuracy, but on an interrogation that covers moral concerns. If a model is not morally comprehensive, some problems might be overlooked, the moral assessment might be flawed, or good solution might be missed.

The idea of a comprehensive analysis can be found in the literature as well. Kunsch et al. (2009) see a moral need for thorough information assessment: "The main issues are not primarily the technicalities or heuristics. The definition of the human context, the identification of stakeholders and their moral values, the systematic analysis of all connections and entanglement with society impose in-depth recurrent analysis." (Kunsch et al., 2009, p.1101) Drake et al. (2011) see assistance in decision-making as the purpose of modeling. As a consequence they claim that every morally appropriate model should consider at least financial, environmental, and social outcomes. (Drake et al., 2011, pp.16-17) Wenstøp (2005) argues that rational decision-making should also take emotions into account. In his discussion, he already assumes that values and the different moral perspectives -- consequentialism, deontological ethics, and virtue ethics – are part of morally sensitive decision-making. In his work, Mason (1994) concludes that modelers make two covenants, one with reality, and one with the client's values; "they promise to secure improvement in the client's world by

adhering to the client's values with as much fidelity as possible." (Mason, 1994, p. 194) The comprehensive inclusion of all interests is a "responsibility principle" for Gallo: "Applied to our field this principle suggests, for example, taking into account not only the point of view of the 'client', … but also the point of view of all 'stakeholders'" (Gallo, 2004, p. 469).

Comprehensiveness is an issue of impartiality. If only few concerns are incorporated, chances are high that important positions of some stakeholders are left out, i.e. one-sided decisions are thus more likely. Accordingly, consequentialism can easily be taken as a motivation for comprehensiveness. Moral theories that see the acknowledgment of all stakeholders as a moral ideal -- such as discourse ethics in Drake et al. (2011) and Kunsch et al. (2009) -- make comprehensiveness even a necessary condition to allow for incorporating all interests of stakeholders. Also from a virtue ethical perspective comprehensiveness is morally good. If we want to understand what virtuous behavior is, we need to know what kind of problems from current and potential behavior can arise.

4.4 Evaluation

Operations research has developed several methodologies that are able to evaluate decision alternatives under scrutiny. Examples are scorecards, cost-benefit analyses, or multi-criteria approaches. Scorecards are tables comprising of options and their expected quantified consequences. They provide a good comparison of the decision alternatives. Cost-benefit analysis is a tool that sums up the utilities of all effects related to a considered option and calculates a final value from them. The option itself or different alternatives in comparison can be evaluated respectively these final values. Multi-criteria approaches compare decision alternatives in qualitative rankings, resulting in a set of rival optimal solutions. For problems with conflicting interests, multi-criteria approaches allow addressing all interests explicitly, which eases the finding of compromises. In summary, all these methods are able to consider sets of alternatives together with their implications, such as value-conflicts, benefits, harm, risk, or environmental impact.

Taebi and Kadak (2010, p. 1358) give a good example for a score card. They compare different kinds of nuclear energy plants. In fact, they consider issues of environmental friendliness and public safety, security concerns, resource durability, economic viability, and technological applicability.

Since all these properties have impacts on qualitative different areas, it is difficult to find an optimal way of using nuclear energy. Rather different political views, environmental characteristics and social expectations influence the decision about which kind of facility would be best. Therefore Taebi and Kadak provide for each kind of nuclear energy production a list of impacts, effects, and consequences such that decision-makers have a concise overview of advantages and disadvantages.

The principle of *evaluation* denotes the detailed explication of consequences or effects concerning each moral issue. An evaluation of results allows understanding the impacts of each moral concern. Often the comparison of alternatives is an essential part of an evaluation. Since the multi-criteria approach can incorporate conflicting objectives Brans (2004) advocates this method for modeling problems that include moral issues. For the example of the Norwegian HIV policy, Wenstøp and Magnus (2001) describe how multi-criteria decision analyses can incorporate moral concerns into rational decision-making, e.g. the extent of future HIV cases, privacy, general and personal anxiety, life quality reduction, and stigmatization of individuals and groups (Wenstøp & Magnus, 2001, p. 61-62). Here multi-criteria decision analyses support coming to a justified decision for such problematic moral concerns. Gallo (2004) discusses, what he calls, the sharing and cooperation principle, meaning that modelers should share openly the insights they gain. Mason (1994) sees it as one of the obligations of the modeler "to insure that actions the client takes based on the model have the desired effect." (Mason, 1994, p. 184)

The reasons for an evaluation relate to the impact of moral issues. If a moral concern is unlikely to have noteworthy consequences, it might be reasonable to cope with other concerns first. Additionally, some moral concerns exist in theory, but do not become problematic in practice, e.g. because of lucky circumstances. The consequentialist argument for an evaluation is therefore straightforward: An evaluation assesses moral concerns for their benefits and disadvantageous and allows estimating the consequences of an action, and finding an optimal action. From a stakeholder perspective, an evaluation of results is essential when different alternatives should be assessed respectively the effects on stakeholders. Finally, an evaluation is important for virtue ethicists in order to assess the effects of human behavior, see e.g. the meat-consumption forecast (Stehfest et al., 2009).

5. Putting the Principles to Work

Employing moral mid-level principles involves two processes: specification and balancing. The meaning of transparency, integrity, comprehensiveness, and evaluation are clear on a broad scale. However, for every particular case, these principles have to be specified. For example, on a broad scale transparency refers to stating clearly the assumptions and structure of a model. For the models about external energy cost, as discussed by Sundqvist et al. (2004), transparency needs to be specified as the obligations to explain

(a) which methodology is employed to assess external cost;

(b) what is regarded as an external cost;

(c) how the 'fuel cycle' is defined in particular;

(d) which external cost are assumed to be relevant and in which steps these are assessed;

(e) which input parameters are assumed to be relevant;

(f) which public and personal preferences are assumed to be relevant, incl. how the impacts on human health and mortality are measured;

(g) which technologies the model focuses on;

(h) which environmental or site-specific circumstances are assumed;

(i) whether there are implicit trade-offs between different environmental effects;

(j) how complexities and uncertainties are coped with.

(cf. the differences in the assessments of external costs in Sundqvist et al., 2004, pp.234-235, pp.238-243)

In some cases principles might conflict with each other. In such cases the principles should be re-specified, weighed, and balanced. Taket (1994) describes a case where transparency is in conflict with integrity:

"Out of a desire to make my report for [the clients] as accessible as possible, I did not include a discussion of the methods used (cognitive mapping). This decision was based on my judgment (grounded in my own experience) that understanding cognitive mapping requires some exposure to

the use of the technique, beyond a simple written explanation. … however, as I have not produced any other [technical] report, I have not followed the ORSA guidelines …" (Taket, 1994, p. 124)

Taket aims on actual transparency for her clients. She believes that a thorough description would nothing but confuse them, and eventually reduce transparency. Therefore she leaves out some information in order to make the model more transparent. Additionally she remarks to be able to provide the report on request. However, she is still in conflict with the ORSA guidelines that demand a full-fledged technical description of the model. In this case she weighs the ORSA guidelines as less important and decides to inflict them.

Although balancing is necessary for applying moral mid-level principles, it poses a threat: It is imaginable that modelers wrongly balance according to intuitions rather than moral norms and principles. Therefore Beauchamp and Childress state six criteria balancing should follow in almost all cases (Beauchamp & Childress, 2008, p.23)[12]:

1. "Good reasons can be offered to act on the overriding norm rather than on the infringed norm."
2. "The moral objective justifying the infringement has a realistic prospect of achievement."
3. "No morally preferable alternative actions are available."
4. "The lowest level of infringement, commensurate with achieving the primary goal of the action, has been selected."
5. "Any negative effects of the infringement have been minimized."
6. "All affected parties have been treated impartially."

In order to put the mid-level principles to work, let us imagine a specific case, namely the optimization of a surgery schedule in a European hospital. In this example, I specify the moral mid-level principles to the actual circumstances. The details are strongly simplified to ease an understanding of the procedure. Furthermore, the purpose of the example is to show the use of moral

[12] I leave it to the reader to show why Taket's decision fulfills these criteria.

mid-level principles. Since I take a pluralistic approach, I do not assume that there is one right unique specification of moral mid-level principles.

Every hospital has several surgery rooms, e.g. 4 such rooms. On the one hand, there are surgeries that can be scheduled, e.g. fixing a broken arm. On the other hand, there are emergencies that need immediate treatment, such as severe injured victims of accidents. Due to (politically willed) increasing cost pressure, hospitals need to use these rooms in an optimal way. Hence, the aim is finding a cost-optimal scheduling scheme. Besides this aim, three secondary aims are pursued:

(1) The rooms should be used continuously throughout the working day.

(2) The rooms should not be used longer than the regular working day lasts.

(3) No surgeries should be canceled.

Intuitively we feel that modeling surgery schedules has a moral dimension. The tools developed in the first part of the article can help to recognize why: The consequentialist perspective reveals that there are moral issues at stake. It might happen that a particular schedule might lead to disadvantages for some patients. As a worst consequence of a bad schedule, it seems imaginable that patients die because of blocked rooms. The other criteria however are morally neutral. The purpose of the model is epistemic; the model is used to interrogate the optimal scheduling scheme. The target system is also morally neutral, namely the schedule of rooms. Now that we know about potential moral issues, I show how the principles can be specified. (Specifications are written in italics.)

Comprehensiveness: The modeler should acknowledge all stakeholders and assess potential moral issues in detail.

The stakeholders of the surgery schedule are nurses, physicians, patients, and society. On the one hand, physicians and nurses benefit from higher salaries during overtime. On the other hand, there is probably a higher benefit from a schedule that reliably stays within the usual working hours. Society is a stakeholder, because European health costs are refinanced via social contracts[13] such that every

[13] In many countries, health insurance functions via insurance companies. Nevertheless I speak of a social contract for Europe, because even in those cases the decision for being insured is hardly a personal matter. Furthermore, there are strong political influences on the insurance companies and their policies.

person in a European society has a health insurance. The first interest on the societal level is cost effectiveness, because all saved money can be used for other purposes. Patients are the last group of stakeholders. We can distinguish between two groups: Patients with scheduled surgeries and emergency patients with unscheduled surgeries. Since one of the secondary aims of the surgery schedule is to decrease surgery cancelations, patients with scheduled surgeries can expect a benefit. On the other hand, those patients could also face longer waiting times, due to higher time variability in the new schedule. Emergency patients can be affected positively or negatively. Dependent on the changes in the scheduling policy it is imaginable that surgery rooms could be easier available; but it is also imaginable that surgery rooms could be blocked for longer periods of time.

Evaluation: Nurses, physicians, and society at large can expect benefits from an optimal scheduling system. *The modeler should assess how big these benefits are dependent on the final scheduling scheme.* Scheduled patients can expect less cancelation, but might face longer waiting time within a day. *The modeler should analyze the actual advantage or disadvantage.* For emergency patients the evaluation is more complex. It is imaginable that due to reasons of efficiency surgeries with high time-variability could be scheduled at the same time (e.g. in the morning). If emergency patients would arrive during these time periods, they might have to wait for longer periods (because surgeries can usually not be stopped) and might, as a consequence, face chronic damages or death. Therefore, *the modeler should explicitly evaluate the problems of emerging patients.*

Integrity: The modeler should accurately incorporate all these moral evaluations into the model and set them in relation to the cost results.

Transparency: The modeler should make the moral problems explicit to the model-user, in this case the hospital management. *The modeler should highlight the origins of the problems and that the goal-criteria set in the beginning neglect these problems.*

Integrity: The modeler and the management should look for a solution according to the management's wishes and according the moral problems of scheduling of surgeries.

Let us assume in this example that the new scheduling scheme jointly blocks all rooms for longer periods of time. We can distinguish between two cases: Scenario A -- the hospital lies in an urban area; and scenario B -- the hospital lies in a rural area.

Scenario A: During the conversation with the management it becomes clear that the potential problem does fortunately not exist. In emergencies, the ambulances check which hospitals have free capacities. Due to the comparably high number of hospitals in an urban area the waiting times for emergencies are always comparably low.

Scenario B: During the conversation, the management understands the problem for emergency patients. Since further hospitals are far away, a solution needs to be found.

Transparency: The modeler should educate the management about the possible solutions to the problem.

Evaluation: The modeler should suggest incorporating the consequences of the trade-off between risk for waiting emergency patients and cost-efficiency.

Integrity: The modeler should present the refined results to the management and discuss the trade-off and its consequences in detail.

Based on the refined results, it could be decided to restrict the time of jointly blocked rooms up to 30 minutes.

6. Final Notes

Moral mid-level principles are loosely connected to moral theories. They do not claim that moral theories in general are wrong or misleading; rather they explicate the local moral obligations of a specific field, in this case modeling. Moral mid-level principles summarize accepted norms into a concise framework, backed up by moral theories. Nevertheless, due to changes within the field of operations research and changes in morality, also the moral mid-level principles might change over time.

What my discussion has shown is that scientific objectivity and moral sensitivity are not opposed. The debate whether moral issues should be dealt with "outside" or "within models" (Le Menestrel & van Wassenhove, 2004) can fortunately be resolved. Modelers can cope with moral issues while maintaining scientific objectivity. Ethics stays outside of models when demanding for integrity and an objective evaluation of results. Nevertheless, it is within modeling via the comprehensive and transparent character that models should have.

The article proposes a method of identifying and solving practical moral issues. The three moral perspectives and the four moral mid-level principles together form this method. In addition the article sets a starting point for a more theory based discussion of practical moral problems.

Acknowledgments

Several people have given me substantial feedback for this article. Among them are my colleagues from the philosophy department at the University of Technology Eindhoven, colleagues from OZSE conference in 2010, participants of the conference Ethics, Energy and the Future in Delft, 2010, and commentators on earlier versions of this paper. Special thanks go to Martin Peterson, Sjoerd Zwart, Kees van Overveld, Valentina Moskalenko, and Petr Kulikov for reviewing and helping to improve the claims in this article.

References

Aristotle, S. (1980). *The Nicomachean Ethics* (W.D. Ross, Trans.). Oxford: Oxford University Press (Original work published 350 B.C.).

Beauchamp, T.L., & Childress, J.F. (2008). *Principles of Biomedical Ethics*. (6th ed.). Oxford Uinversity Press: New York.

Blair, M., Obenski, S., & Bridickas, P. (1992). GAO/IMTEC-92-26 Patriot Missile Software Problem. Retrieved from http://www.fas.org/spp/starwars/gao/im92026.htm, last accessed 30th Sep 2011.

Brans, J. P. (2002). OR, Ethics and Decisions: the OATH of PROMETHEUS. *European Journal of Operational Research, 140*, 191-196.

Brans, J. P. (2004). The management of the future. Ethics in OR: Respect, multicriteria management, happiness. *European Journal of Operational Research, 153*, 466-467.

Budd, L., Griggs, S., Howarth, D., & Ison, S. (2011). A Fiasco of Volcanic Proportions? Eyjafjallajökull and the Closure of European Airspace. *Mobilities, 6(1)*, 31-40.

Drake, M. J., Gerde, V. W., & Wasieleski, D. M. (2011). Socially responsible modeling: a stakeholder approach to the implementation of ethical modeling in operations research. *OR Spectrum, 33(1)*, 1-26.

Fienberg, S. E. (1994). Ethical Modeling Considerations in Correcting the Results of the 1990 Decennial Census. In W. A. Wallace (ed.), *Ethics in Modeling* (pp. 103-144). Oxford: Elsevier/Pergamon.

Gallo, G. (2004). Operations research and ethics: Responsibility, sharing and cooperation. *European Journal of Operational Research, 153*, 468-476.

Gass, S. I. (2009). Ethical guidelines and codes in operations research. *Omega, 37*, 1044-1050.

Jenkins, D.G., & McCauley, L.A. (2006). GIS, SINKS, FILL, and Disappearing Wetlands: Unintended Consequences in Algorithm Development and Use. In H. M. Haddad (ed.), *Proceedings of the 2006 ACM symposium on applied computing* (pp. 277-282). New York: ACM.

Kant, E. (1998). *Groundwork of the Metaphysics of Morals* (M. Gregor, Trans.). New York: Cambridge University Press (Original work published 1785).

Kleijnen, J. P. C. (forthcoming). Ethical Issues in Engineering Models: An Operations Researcher's Reflection. *Science and Engineering Ethics*, DOI 10.1007/s11948-010-9215-5.

Kunsch, P.I., Kavathatzopoulos, I., & Rauschmayer F. (2009). Modelling complex ethical decision problems with operations research. *Omega, 37*, 1100-1108.

Mason, R. O. (1994). Morality and Models. In W. A. Wallace (ed.), *Ethics in Modeling* (pp. 183-194). Oxford: Elsevier/Pergamon.

Le Menestrel, M., & Van Wassenhove, L. (2004). Ethics outside, within, or beyond OR models?. *European Journal of Operational Research, 153*, 477-484.

Le Menestrel, M., & Van Wassenhove, L. (2009). Ethics in Operations Research and Management Sciences: A never-ending effort to combine rigor and passion. *Omega, 37*, 1039-1043.

Mingers, J. (2011). Ethics and OR: Operationalising discourse ethics. *European Journal of Operational Research, 210*, 114-124.

Morgan, M.S., & Morrison, M. (1999). *Models as mediators: Perspectives on natural and social science*. Cambridge University Press: Cambridge.

Van De Poel, I. (2001). Investigating Ethical Issues in Engineering Design. *Science and Engineering Ethics, 7*, 429-446.

Rabins, M.J., & Harris Jr., C.E. (1997). The Ethics of Modeling. *Control Engineering Practice, 5(4)*, 519-526.

Ross, N. P., & Harris, S. (1994). From Model Building to Risk Management: Evolving Standards of Professional Responsibility. In W. A. Wallace (ed.), *Ethics in Modeling* (pp. 161-166). Oxford: Elsevier/Pergamon.

Stehfest, E., Bouwman, L., Van Vuuren, D. P., Den Elzen, M. G. J., Eickhout, B., & Kabat, P. (2009). Climate Benefits of Changing Diet. *Climatic Change, 95*, 83--102.

Sundqvist, T., Söderholm, P., & Stirling, A. (2004). Electric Power Generation: Valuation of Environmental Costs. In C.J. Cleveland (ed.), *Encyclopedia of Energy* (pp. 229-243). San Diego: Academic Press.

Tacket, A. (1994). Undercover Agency? -- Ethics, Responsibility and the Practice of OR. *The Journal of the Operational Research Society, 45(2)*, 123-132.

Taebi, B., & Kadak, A.C. (2010). Intergenerational considerations affecting the future of nuclear power: Equity as a framework for assessing fuel cycles. *Risk Analysis, 30(9)*, 1341-1362.

Walker, W. E. (1994). Responsible Policy Making. In W.A. Wallace (ed.), *Ethics in Modeling* (pp. 226-241). Oxford: Elsevier/Pergamon.

Walker, W. E.(2009). Does the best practice or rational-style model-based policy analysis already include ethical considerations?. *Omega, 37*, 1051-1062.

Wallace, W. A. (1994). *Ethics in Modeling*, Oxford: Elsevier/Pergamon.

Wenstøp, F., & Magnus, P. (2001). Value focused rationality in AIDS policy. *Health Policy, 57*, 57-72.

Wenstøp, F. (2005). Mindsets Rationality and Emotion in Multi-criteria Decision Analysis. *Journal of Muli-Criteria Decision Analysis, 13*, 161-172.

Wenstøp, F. (2010). Operations research and ethics: development trends 1966—2009. *International Transactions in Operational Research, 17*, 413-426.