Proceedings of the 2018 Winter Simulation Conference
M. Rabe, A.A. Juan, N. Mustafee, A. Skoogh, S. Jain, and B. Johansson, eds.

THE ETHICS OF COMPUTER MODELING AND SIMULATION

F. LeRon Shults

University of Agder and NORCE
Center for Modeling Social Systems
Universitetsveien 19, SKP Building
Kristiansand, 4630, NORWAY

Wesley J. Wildman

Boston University and the
Center for Mind and Culture
566 Commonwealth Avenue, Suite M-2
Boston, MA 02215, USA

Virginia Dignum
Delft Institute Design for Values
Delft University of Technology
Jaffalaan 5
Delft, 2628 BX, THE NETHERLANDS

ABSTRACT

This paper describes a framework for ethical analysis of the practice of computer Modeling & Simulation (M&S). Each of the authors presents a computational model as a case study and offers an ethical analysis by applying the philosophical, scientific, and practical components of the framework. Each author also provides a constructive response to the other case studies. The paper concludes with a summary of guidelines for using this ethical framework when preparing, executing, and analyzing M&S activities. Our hope is that this collaborative engagement will encourage others to join a rich and ongoing conversation about the ethics of M&S.

1. INTRODUCTION

Scholars and practitioners of Modeling and Simulation (M&S) have sometimes explored the ethical assumptions and implications of their work, but these discussions have typically occurred at the edges of the discipline. Although calls for a series of ethical guidelines for the field began to emerge over 30 years ago, it was not until 2002 that a code of ethics was produced and widely accepted (McLeod 1986; Oren et al. 2002). In recent years, it has become increasingly clear that the analytic and forecasting power of computational methodologies warrants even more explicit attention to the ethics of M&S. Insofar as models are always shaped by the intention of the modelers, ethical questions hover around (or hide behind) their construction. This raises questions such as: Is the purpose of a given model "good," and if so, for whom? By whose standards?

2. A META-ETHICAL FRAMEWORK

Elsewhere the first two authors of this paper set out a framework for the conversation about ethics in M&S (Shults and Wildman in press). Rather than a guide for resolving specific ethical dilemmas, that framework is meant to provide a *way of thinking* about the ethics of simulation. In the current paper, we begin to do the actual thinking in the context of specific models. We put the meta-ethical framework to the test, exploring the ethics of simulation in the context of three different case studies, briefly outlining and then commenting on models in whose construction and application we have played a role. Each of us also responds to the ethical reflections of the others, in an attempt to illustrate the value of this level of ethical

discussion among colleagues. We conclude with a brief summary and a call for further engagement. In the remainder of this introduction, we outline the ethical framework whose usefulness we hope to demonstrate.

This framework has three components: an outline of the general principles of ethical theory-construction, an emphasis on the relevance of evolutionary perspectives on moral reasoning and behavior, and an attempt to show how these first two can help mitigate moral confusion and moral evasion, respectively. The first component engages what we call *philosophical meta-ethics*. This involves a distinction between "the good" (what is valuable?) and "the right" (what should individuals or institutions do?). Any ethical theory worth its salt will need to link these in a coherent way. As John Rawls points out, ethical theories of the *teleological* type (from the Greek *telos*, meaning end or purpose) tend to define the good independently from the right, and then define the right "as that which maximizes the good. More precisely, those institutions and acts are right which… produce the most good, or at least as much good as any of the other institutions and acts open as real possibilities" (Rawls 2005, p. 24). Such theories are sometimes called "consequentialist" because they focus on the consequences of (im)moral actions, and evaluate their rightness on the basis of their connection to the chosen standard of goodness. That standard varies widely among consequentialists; for example, hedonists choose pleasure, perfectionists choose human excellence, and eudaimonists choose happiness. We use "consequentialist" rather than "teleological" in what follows for the sake of consistency.

The other major type of ethical theory is *deontological* (from the Greek *deon*, meaning that which is binding, needful, or right). In this case, to quote Rawls again, the theorist "does not specify the good independently from the right, or does not interpret the right as maximizing the good" (Rawls 2005, p. 30). Actions are wrong not because of their consequences, but because the sort of act that they *are* is not right. One popular example of this approach is "divine command" theory, which subsumes the good and the right in the ideal of obedience to the will of a god. Immanuel Kant's categorical imperative is another example: act only according to that maxim whereby you can at the same time will that it should become a universal law. John Stuart Mill's utilitarianism, on the other hand, is an example of ethical theories of the consequentialist type; for him, the highest value (the good) is the promotion of the well-being of the greatest number of people. Insofar as "virtue theories" of ethics focus on virtue itself as the good, and promote moral behaviors that engender its formation, they can also be considered consequentialist in the broadest sense. The first component of our meta-ethical framework for M&S encourages professionals in the field of M&S to get clear on how they think about the good, and the criteria by which they evaluate the rightness of their professional activities.

The second component of our framework is *scientific meta-ethics*. Most ethical theory construction in western philosophy has occurred within what we now call the humanities. Humanities scholars have produced enormously rich literature on the criteria and conditions for human moral behavior. In the last few decades, however, scientific disciplines have increasingly taken up interest in ethics. The evolutionary sciences in particular have contributed insights into the emergence of human morality, demonstrating the extent to which our capacity for cooperation and (at least parochial) altruism has analogues in other species. Even more significantly, research in these fields has shown that the eusociality of our species is a result of natural selection, having provided our ancestors with new capacities that increased their survival advantage. Although there are a wide variety of competing theories about the precise evolutionary mechanisms by which this occurred, and the value of those mechanisms in our contemporary environment, there is a broad consensus that human morality can be explained in evolutionary terms. Our main point here is that meta-ethics can no longer ignore the empirical findings and theoretical developments in these fields. The plausibility of philosophical meta-ethical claims is now constrained by their coherence with evolutionary insights into the evolution of human morality.

The *practical import of meta-ethics* is the third component of our framework. What does all of this mean for practitioners of M&S? We suggest that consequentialist approaches to ethics appeal somewhat naturally to most computer scientists and engineers because these fields invite utilitarian analysis and focus on concrete consequences. But this leaves open the key ethical question: what are the criteria for assessing

the value (goodness) of consequences? Even for those who might prefer a deontological approach, we argue that it is important for M&S professionals to be more explicit about values that guide their ethical judgments. We also suggest that it is easier to accept moral responsibility for our judgments and choices – even as scientists and scholars – if we acknowledge the nature and power of the moral equipment that is part of our phylogenetic inheritance. Computer modelers are human too, and the more we all understand the mechanisms that shape our ethically laden decisions the easier it will be to accept our limitations and to develop more plausible models and more useful simulations to guide our discussions about human behavior. Let's consider some case studies.

3. CASE STUDY: THE ETHICS OF SIMULATING SECULARITY (SHULTS)

The case I have chosen for meta-ethical analysis and reflection is a computational model and set of simulation experiments reported in "Forecasting Changes in Religiosity and Existential Security with an Agent-Based Model," an article our research team recently published in the *Journal of Artificial Societies and Social Simulation* (Gore et al. 2018). Our goal was to identify some of the conditions under which – and mechanisms by which – religious beliefs and behaviors decrease (or increase) in a population. The model can be construed as an attempt to simulate "secularity" as well as "religiosity." I begin with a description of the case study to set the stage. I then offer a series of scientific and philosophical meta-ethical reflections in the context of a discussion of the tension between "religious" and "secular" approaches to organizing society and enforcing moral norms. Empirical findings and theoretical developments in the relevant disciplines provide warrant for the philosophical claim that if our intention is to enhance the well-being of future human generations, we ought to promote secularism (and naturalism) in contemporary populations, a task that can be aided by social simulation.

Within the author team that wrote the article described in this case, we referred to the computational architecture we developed as the Non-Religiosity Model – or NoRM. It is well-known in the scientific study of *religion* that the latter is negatively correlated with individual variables such as higher education and with contextual variables such as higher existential security (e.g., Norris and Inglehart 2015; Hungerman 2014). We developed an agent-based model in order to further specify these claims about the hypothesized causal relationships among religion, education, and existential security, and to construct an artificial society in which the relevant dynamics could be explored. Our agent architectures were based on structural equation models drawn from new statistical and factor analyses of the International Social Survey Programme (ISSP). The agents in NoRM are connected within social networks, and interact with one another based on each agent's education level, religious practices, and existential security, within their natural and social environments. Each agent also has variables related to belief in God, supernatural beliefs, and religious formation. See Figure 1 for a graphical representation of the variable dependencies within NoRM that enabled us to study changes in religiosity and existential security in the artificial society.

We initialized the existential security level of the environment at the beginning of each run using data from the Human Development Report (HDR) on selected countries for particular years. The HDR provides an annual multi-facetted measurement of country-level well-being based on variables such as longevity, health, and standard of living (Sen 2003). The agents were initialized by randomly sampling ISSP respondents from the specified country and year. The characteristics of each simulated agent were parameterized based on the data from one of the selected respondents. Our goal was to simulate the emergence of macro-level shifts in religiosity and existential security within a real-world population (observed in the ISSP and HDR data sets) from the micro-level agent interactions in the model (guided by the scientific literature and our statistical and factor analyses). The model was calibrated by testing its capacity to predict the (real-world) shifts in the relevant variables that occurred during a 10-year period (1990-2000) within 11 countries (selected because they had sufficient ISSP and HDR data).

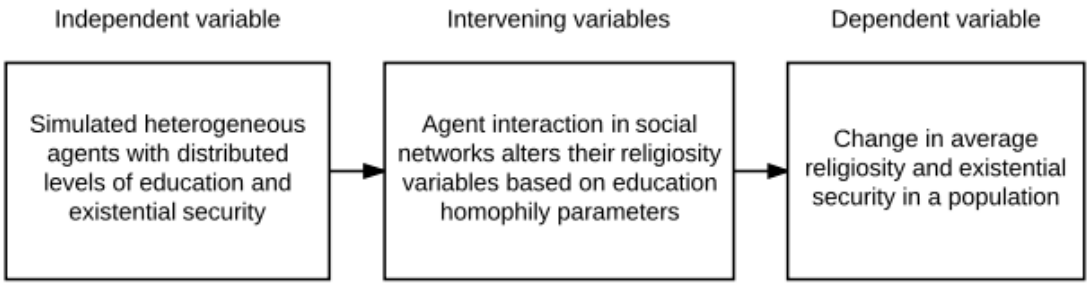

Figure 1: Variable dependencies within NoRM.

We then used the calibrated model to predict shifts in the relevant variables for 22 countries (including 11 other countries) during a different 10-year period (2000-2010). The extent to which all three (alterable) religiosity variables changed in the model was strongly influenced by the educational level of the social networks within which the simulated agents were embedded. As agents interacted with more educated agents in the model, religious practice, belief in supernatural agents, and belief in God all decreased (across levels of existential security), thus providing an initial validation of the micro-level behavioral architecture. A further and stronger validation came from a second set of experimental results in which NoRM was able to simulate the shift in the relationships among religiosity and existential security variables in all 22 countries over the relevant decade. In fact, the predictions of our ABM were up to three times more accurate than its closest competitor, linear regression analysis. For details, see (Gore et al. 2018) and the supplementary materials online at https://github.com/rossgore/JASSS-Special-Issue (accessed August 18th, 2018).

NoRM provides ample fodder for scientific and philosophical meta-ethical reflection. What sort of ethical issues are raised by this model of some of the conditions that decrement religiosity (or increment secularity) in contemporary populations? On the one hand, it raises the same general concerns about "social engineering" raised by any model that purports to forecast human futures based on the identification of mechanisms at work in complex adaptive social systems. I do not have adequate space to address those concerns here. On the other hand, this model also raises specific concerns about the relative value of religious and secular approaches to justifying and enforcing moral norms that hold societies together. Those concerns will be my focus in the space that remains. What consequences do "religiosity" and "secularity" have on the well-being of human societies? A great deal depends on the way in which we define these scare-quoted terms – and on which societies we have in mind. NoRM uses *religiosity* in the same way our team has operationalized it in our work on other models: "shared imaginative engagement with axiologically relevant supernatural agents" (Shults et al. 2017).

Religiosity (as defined in NoRM) leads individuals to try to make sense of ambiguous or frightening phenomena in nature and to act sensibly in society by appealing to the putative revelations of empirically unfalsifiable, human-like, coalition-favoring supernatural agents to which only the members (or ritual elite) of their in-group have ritual access. For those who share in the relevant rituals, these imagined forces are "axiologically relevant" insofar as they are interpreted as having some bearing on the normative judgments of their coalition. Secularity, in the sense I am using the term here is: not that. Secularity (as defined in NoRM) inspires individuals to try to act sensibly in pluralist societies without appealing to the alleged supernatural authorities preferred by any particular religious group. This tendency can be reinforced by naturalism, which inspires individuals to try to make sense of confusing or worrisome phenomena with hypotheses that only include empirically falsifiable claims about natural causes. Secularism and naturalism are reciprocally reinforcing. Scientists and other highly educated people in secular societies are typically taught to be methodological naturalists, and today the most existentially secure societies are typically those with high levels of secularity, which promotes open scientific inquiry.

Insights from a wide variety of disciplines, including evolutionary biology, moral psychology, cognitive science, cultural anthropology, and archaeology, converge in support of the claim that religion (in the sense stipulated within NoRM) emerged in early ancestral societies as a result of attributive guesses reinforced by affiliative pressures (see Shults 2014, 2018). It makes sense that both the tendency to infer "hidden agent" when confused or afraid and the tendency to prefer "tribal norms" when confronted by cultural others would have been naturally selected in the late Holocene. Today, however, most of us do not live in homogenous small-scale groups, hunting and gathering across the African savannah, or in mid-sized sedentary collectives sharing in the domestication of plants and animals in relatively isolated plains in the Levant. The majority of people on the planet now live in pluralistic, densely-packed, large-scale societies rapidly running out of agricultural and other resources to support our expansive sedentation.

Unfortunately, as far as the viability of the human species in our current habitat is concerned, the same cognitive and coalitional mechanisms that engender religiosity in contemporary minds and cultures also appear to foster superstitious beliefs about the causes and consequences of extreme climate change and amplify segregative behaviors that lead to excessive consumer capitalism and escalating cultural conflict (Shults 2015). NoRM has been introduced into a scholarly environment in which researchers have argued that religiosity is correlated with heightened hostility between groups and with a host of other sociocultural dysfunctions in contemporary societies (Zussman 2014; Du and Chi 2016); that "in almost all regards the highly secular democracies consistently enjoy low rates of societal dysfunction" (Paul 2005, p. 1); and that, when it comes to measuring factors such as happiness, valuing motherhood, promoting peace, and murder rates, the least theistic states come out far better than the most theistic states (Zuckerman 2014). Thus, NoRM is immediately implicated in profound ethical debates about human well-being and species survival.

The research informing NoRM suggests that morality need not be dependent on shared imaginative engagement with axiologically relevant supernatural agents. Religiosity (as operationally defined) helped our ancestors survive by reinforcing anxiety about invisible forces that could punish cheaters and freeloaders, and promoting xenophobic tendencies that both fostered parochial altruism and justified killing out-group members. Likewise, the same research implies that, in our current globally connected, ecologically fragile environment, supernaturalist beliefs and behaviors have become maladaptive – at least if one's adaptive goal has to do with the well-being of all human populations and not merely with the dominance of one's own religious in-group – and that religious prosociality is "less effective at improving societal conditions than are secular government programs" (Paul 2009, p. 398; cf. Zuckerman et al. 2016). NoRM speaks directly to these concerns and to the research that informs them, and thus is a morally charged contribution to the debate over the future of the human species. To the extent that an important aspect of "the good" is the improvement of societal conditions for as many human coalitions as possible, utilizing computer models such as NoRM to discover the mechanisms that can promote secularism and naturalism may be "the right" thing to do.

4. CASE STUDY: EXTREMIST VIOLENCE (WILDMAN)

The case I have selected is a computational model on extremist violence published as "Stability of Groups with Costly Beliefs and Practices" (Wildman & Sosis 2011). Costly displays have always provoked the question of evolutionary viability. How does penis laceration (a favorite example of cultural anthropologists, doubtless intended to unnerve readers) or other painful initiation rituals help someone to reproduce successfully? It is not difficult to understand such practices surviving in tiny subcultures for a season, driven by enthusiasm and social pressure, and parasitic on the more adaptive behavior of a larger cultural group. But how does an entire culture survive when that culture is characterized by permanent, extremely costly displays?

The starting point for this computational model was an evolutionary (not a computational) model produced by Joseph Henrich to show that there can be a stable evolutionary equilibrium for an entire population committed to costly displays (Henrich 2009). There is always a no-cost equilibrium for societies

but, under certain circumstances, an entire culture can migrate toward a high-cost equilibrium over a number of generations. This result helped to make sense of costly signals but it was limited by the focus on an entire culture. We wanted to extend the result to the more realistic situation of a population peppered with subgroups committed to high-cost beliefs and practices. We see this, for example, in extremist violence perpetrated by so-called terrorist groups. They exist within majority cultures as tiny subcultures insisting on extremely costly displays. Are such groups here to stay? What is the secret of their longevity?

To investigate this, we constructed an agent-based model having agents equipped with the cognitive capacities presupposed in Henrich's population-level cultural evolutionary model. These agents use success-weighting calculations to determine whether to join or leave a high-cost group. We performed a sweep using Latin hypercube sampling on key parameters and then analyzed results. We discovered that high-cost groups achieve long-term stability within a larger population under a wide range of circumstances, a finding that extends Henrich's result in a more realistic direction. The most important emergent pathway to costly group stability is the simultaneous presence of high charisma and consistency of the group leader and high cost of the group. These findings have strategic implications both for leading groups committed to costly beliefs and practices and for controlling their size and influence within wider cultural settings.

To analyze this model ethically using the framework sketched in the Introduction, we begin with philosophical meta-ethics. I am committed to a consequentialist, not deontological, ethical framework. I have philosophical reasons for this, as well as for the particular kind of consequentialist ethical framework I favor, but that is beside the point right now. The purpose of this first part of the ethical framework is to *make explicit* the relevant criteria for ethical evaluation, rather than surreptitiously importing and imposing criteria outside the reach of proper critical scrutiny. In my case, the criteria for ethical evaluation of moral decisions and actions are their *consequences*. This will prove to be important, as we shall see.

The second part of the ethical framework is scientific meta-ethics. For my current purposes, there are two ways to apply considerations pertaining to scientific insights into human moral behavior and reasoning. On the one hand, we can focus on the subject matter, which is small groups committed to costly displays, including deadly violence and radical self-sacrifice to advance group aims. In this case, the science would shed light on motivations and strategies, supporting a moral evaluation of this kind of group and the associated types of violence. That in turn may motivate our actions as modelers, leading us to focus on this particular phenomenon out of curiosity or perhaps from moral determination to improve the world. On the other hand, we can focus directly on our work as modelers. In this case, we use science to look in the mirror at our own moral decisions and actions in the actual process of constructing and exploring computational models. Science can tell us a lot about our motivations, our moral impulses, and tendencies to cognitive error, and our over-reliance on socially stabilized moral intuitions that perhaps should be evaluated anew.

Notice that Shults focuses on the first application of scientific meta-ethics – applying it to the subject matter, and thus to motivations for choices about what to model and the impact of a model on the research fields into which it is introduced. To balance accounts, I'll briefly dwell on the second application – applying scientific considerations to the ethics of the modeling process itself.

The team behind the model under discussion had blended motivations. A mixture of curiosity and horrified fascination drew our attention to high-cost extremist groups perpetrating violence in the first place, followed by a morally infused determination to understand motivations and ultimately to alleviate the problems associated with the existence of violent extremist groups. Science tells us a lot about how our attention is caught and shaped, and not all of that information is flattering to us as modelers. I suppose this might be the modeling equivalent of rubbernecking at a freeway collision, slowing down traffic due the effect of shock and awe. *Most people leave things there but policy experts, warriors, and modelers can go further, applying their tradecraft to the phenomenon that grabbed their attention.*

We approach such professional tradecraft activities with extraordinarily rich moral intuitions, most of which we take for granted even though we'd be wise to stay aware of them and to evaluate them. Cognitive science has helped us understand the blindness to our assumptions that can easily arise in such situations: we minimize awareness of our complex moral assumptions in order to simply our perception of behavioral

options and to unburden our minds in a way that helps us settle rapidly on action plans. This is as true in the decision to apply modeling to a social problem as it is in any other part of life. But we can become aware of that tendency, slow down the process of generating effective action plans, and interpose rational questions about our morally laden motivations for the choices we make as modelers. In the case of this modeling team, we need to take responsibility for the fact that we are not impartial, that we think violent extremism is bad for human life, and that we believe there are better and even non-violent ways to express dissent that may in fact be more effective. I doubt that we would have pursued this line of research if we had not felt this way about violent extremism.

The third part of the ethical framework is its practical import, which in this case means implementing the operative consequentialist ethics to interpret – in terms of its practical consequences – the model we built and the way we choose to describe it in publications and elsewhere. There are a hundred considerations here so, to maintain focus, I'll focus on three.

First, we defined the groups to which people can choose to belong or leave as led by the most charismatic person in the group, whose behavior is closely scrutinized by group members and outsiders alike. This is an approximation to a social situation that, in reality, is a lot messier. No surprise there: modelers are constantly making intelligent simplifications in a quest to isolate the most relevant dynamics. But, in retrospect, that approximation also strikes me as morally laden in that one person is taken to be especially representative of an entire group and its disruptive visionary worldview. Now, such groups are hierarchical and do have visionary leaders, and those leaders are scrutinized closely and held to a higher behavioral standard, so we are not making a horribly distorted simplification. Nevertheless, we are focusing attention on a single person (not a small cabal or a network of leaders) as a proxy for an entire group and its actions, and nobody should pretend that such representational moves are morally innocent or neutral.

Second, we adopted a pattern of interaction from the literature in the epidemiology of representations, with learner agents interacting with exemplar agents. It is by means of these interactions that learner agents change their worldviews and make decisions about whether to belong to a group, which group to join, and whether to leave their group. Obviously, this is another simplification, hopefully an intelligent one. But our widening ethical sensitivities as modelers urge us to consider whether this modeling assumption has moral consequences for which we need to take responsibility. I think this simplifying assumption does indeed have moral import. In practice, people often belong to multiple groups, which the model does not envisage. These group affiliations function to moderate extremity in one another, produce a practical kind of wisdom that gives people some degree of moral resistance to the revolutionary visions of groups advocating extreme violence. Moreover, the model takes no account of personal trauma, which often lies at the root of the resentment leading people to so identify with a group that they could envision dying for the group as a form of personal self-defense and vengeance, simultaneously. The way we implemented the epidemiology of representations is certainly consistent with Henrich's original cultural evolutionary model but our approach tends to emphasize group influence while minimizing the unpredictable elements associated with personal decisions to commit to high-cost extremist groups.

Third, though we had not anticipated this, we discovered that this model generates a criterion for when to assassinate the leader of a group advocating extremist violence – namely, remove leaders who are the most charismatic and the most consistent (in the sense of practicing what they preach), and whose groups impose the highest costs on members. Groups without these features will tend not to last long anyway, but most of the long-lived groups have precisely these characteristics. This is a morally laden consequence of modeling efforts that nobody could possibly miss. If we were employing a deontological framework in which killing human beings is always wrong, implementing such a criterion would be morally out of bounds. But in a consequentialist ethics, killing for the greater good may be deemed morally preferable to doing nothing. No doubt drone-killings to assassinate leaders of violent extremist groups are conducted with a clearly stated set of criteria that are carefully evaluated ethically before being implemented. We are certainly correct to be concerned about the ethical consequences of such model-driven conclusions.

There are important interactions among these three concerns. The fact that this model is so suggestive of a leader-assassination criterion reflects the first consideration in that it is focused on individual leader influence, and the second consideration in that it neglects unpredictable elements in favor of group influence on individual member recruitment and mind-shaping. It might have been possible to miss these connections, or indeed to fail to notice the ethical connections of model design assumptions altogether, were it not for the ethical framework that forces us to contemplate the model in three clearly marked phases.

5. CASE STUDY: SMOKING IN PUBLIC SPACES (DIGNUM)

The case I've selected is an agent-based model that attempts to understand the relation between personal and social values and the compliance, or not, with formal regulations. The study is published in (Dechesne et al. 2013) using as scenario the introduction of smoking prohibition in public spaces. The paper further attempts to explain the differences in implementation and uptake across several countries of the European regulations on smoking in public spaces by relating these to differences in cultural dimensions across European countries (Hofstede 2001). In this article, we explored how distinguishing different types of norms can provide a way to reconstruct agents' acting according to, or in violation of, norms. In particular, we distinguished three types of norms: legal, social and private norms, and proposed the six possible orderings of those types to characterize different agents.

Legal norms refer to those rules of conduct that are explicitly formulated and imposed on the community by a central entity, and which usually also have an explicit sanction for violation. The laws of a community are typical examples of legal norms. Legal norms make explicit for the entire community how to behave in order to support some underlying value. Acceptance of a legal norm may depend on the extent to which the underlying value is supported (cf. table 1). *Social* norms are more implicit and more flexible: they only cover a subgroup of the community, their boundaries are hardly defined, and an agent can decide to (temporarily) leave a certain subgroup on the basis of lack of support for the social norms in that group. Typically, an agent will comply with a social norm or rule if (a) a sufficiently large number of peers conform to the rule, and (b) a sufficiently large number of others expects her to conform to the rule, and may sanction behavior. *Private* norms refer to the personal normative beliefs a person has developed over his or her life. We abstract from the way they came to be the personal norms of an agent (partly derived from social norms, partly from legal norms, in the different societies an agent has been part of), and assumed them to be fixed for each agent by their personal history. They are the standards of behavior a person holds for him- or herself. Table 1 summarizes the main differences between these norm types.

The culture of a society (with respect to agents' attitudes towards norms) can then be seen in terms of how it is composed of such agent types. In order to demonstrate how norm types can be used to study the global behavior of societies, we developed a simulation of the introduction of a law, prohibiting smoking in cafes, in an agent society. It showed that different compositions of the society in terms of agents' norm type preferences will respond differently to the introduction of the law.

Table 1: Differences between legal, social and private norm types.

	Legal	**Social**	**Private**
Value	Compliance	Conformity	Integrity
Origin	Imposed by institution	Emergent / dynamic	Fixed (by history)
Description	Explicit	Implicit	Implicit
Monitoring	Enforced	Power of numbers	Conscience
Sanction	Formal	Exclusion	Lower self-esteem

Formal smoking prohibitions for restaurants and cafes have been introduced in several European countries over the past years, with Ireland being one of the first (2004), and The Netherlands relatively late (2008). While most people in current society support the underlying value of the introduced law, viz. that

smoking is unhealthy for the smoker and its environment, the introduction of the prohibiting law provoked considerable resistance and --at least in some countries-- vast violation. This example is also interesting from an ethical perspective given that different values are at stake: e.g. health, care for others (with respect to their health), but also: joy or pleasure at the individual level, but also economic interests, and at a higher level issues of authority and freedom of choice.

In order to test how tendency towards a norm type influences behavior we developed an artificial society, simulating a community of people that frequent a café. Based on demographic information, we assumed that smokers are more likely to frequent the cafe than non-smokers. We designed the agents in this simulation to have a private attitude towards smoking and a preference order on the three types of norms. In each round, each agent would decide to go to the café, or if already there, to stay or to leave. Each simulation took 200 rounds and midway the simulation the law banning smoking in public spaces was introduced. The decision rules implementing the social norm type were based on majority rule (i.e. a social agent would accept the stance – smoking or not – followed by the majority of agents in the cafe at the time).

The simulation illustrated how different preferences over the three norm types can result in different behavior changes after the introduction of the anti-smoking laws. Even though the simulation was very simplistic (Dechesne et al. 2014), results show that, as can be expected, highly normative societies (where the percentage of agents with preference for legal norm type is above 50%) react positively to the introduction of the smoking ban. This can be explained by the fact that non-smokers will be more inclined to go to the cafe, as they can be sure that the place will be smoke free. On the other hand, when smokers were in majority prior to introduction of the law, in configurations where social agents are in the majority, the number of clients typically diminishes after the introduction of the law. Non-smokers and lawful agents will not stay in the cafe: they don't feel comfortable, because of the smoke or because the law is not being upheld respectively.

The paper then attempted to analyze a link between norm type preference and cultural differences according to the well-known cultural dimensions model of Hofstede [11], which distinguishes 4 cultural dimensions: Power Distance Index (PDI), Individualism (IDV), Masculinity Index (MAS), Uncertainty Avoidance Index (UAI). Note that we used the original model, and therefore did not consider the dimensions Long-Term Orientation (LTO) and Monumentalism, which were added later. Initial reflection indicates association between legal norms with high PDI (legal norms come from an authority) and/or high UAI (through their explicit formulation, legal norms create clarity and inter-subjectivity); Social norms with low PDI (equality), low AS (caring for others), and low IDV (the importance of belonging to the group); and Private norms with high IDV (the private context is taken as guiding), high MAS (assertiveness), low PDI. These links need to be further explored in other simulations and case studies, given that existing data were not sufficient to verify these assumptions.

From a philosophical meta-ethics perspective, the paper takes a consequentialist approach, in the sense that it looks at the consequences of one's behavior as measure of the success of a policy or rule, in this case the introduction of the non-smoking laws. However, the model is constructed from a deontological perspective, linking the agents' actions to some (pre-defined) internal deontic rules. This is, in my opinion, an important issue in ABMS, which from a scientific meta-ethics perspective is often left implicit in design choices and assumptions. In fact, this study raises several ethical considerations, in particular where it concerns the application of broader cultural characteristics to inform the behavior of individual agents. Individual reasoning and motivation has many different bases; reducing it to its outcomes, and more specifically to the outcomes expected from membership to a given group or society, is indirectly taking a consequentialist view on behavior, in which results are leading over motives or individual norms. This is an issue often experienced in agent-based modeling where descriptive results from social sciences are used as prescriptive models for the specification of the behavior of the agents in the simulation. Even though this can be an acceptable approach, the discipline of simulation can benefit from a deeper reflection on this choice. Finally, one should reflect on the practical importance of this work. This study can better be seen

as explorative in nature, where the focus is not so much on the issue of the smoking ban, but on gaining a better understanding on the effect and worth of non-consequentialist models of behavior.

Given that the use of the agent paradigm to understand and design complex systems occupies an important and growing role in different areas of social and natural sciences and technology, it is important to consider the ethical aspects of this practice. In particular, ethical concerns related to the assumption of rationality underlying many ABM simulations (Dignum 2017). Agent rationality can be summarized as follows: (a) Agents hold consistent beliefs; (b) Agents have preferences, or priorities, on outcomes of actions; and (c) Agents optimize actions based on those preferences and beliefs. This view on rationally entails that agents are expected, and designed, to act rationally in the sense that they choose the best means available to achieve a given end, and maintain consistency between what is wanted and what is chosen. This view fits well with a consequentialist approach but unfortunately, from a modeling perspective, real human behavior is neither simple nor rational, but derives from a complex mix of mental, physical, emotional, and social aspects. Realistic applications must consider situations in which not all alternatives, consequences, and event probabilities can be foreseen. This leads to questions on the use of rational choice approaches unable to accurately model and predict a wide range of human behaviors. And, as a corollary, to the realization that non-consequentialist approaches may have an important role to play in agent-based modeling and simulation. In particular, in our work, we specify agent architectures grounded on sociality principles, which enable agents to (a) to fulfill several roles, and pursue seemingly incompatible goals concurrently, for example, simultaneously aiming for comfort and environmental friendliness, or for riches and philanthropy; (b) to hold and deal with inconsistent beliefs for the sake of coherence with identity and cultural background; and (c) to act based on altruism, fairness, imitation or even laziness. These architectures are based on satisficing rather than maximization as means for decision-making and action.

6. CONVERSATION: RESPONDING TO ONE ANOTHER

6.1 Shults's Response to Wildman and Dignum

I appreciated how explicitly Wildman discussed the motivations that led him to choose to study the stability of groups with costly beliefs and practices. Like so many (myself included) he is concerned about the way in which such coalitions can so easily engender a willingness among their members to kill or die for their group. Wildman also expresses a morally charged determination to promote a more universal prosociality that attends to the well-being of all humans regardless of their group status. I share this concern as well. As the reader can probably guess from my case study presentation, my moral preferences would also lead me to want to model more explicitly the role that religious superstition and segregation can play in amplifying the cognitive tendencies and producing the social conditions that can lead to violent extremism. As I noted above, our team has already developed computational models that study the escalation of religiously motivated intergroup violence. We have also developed other models that more explicitly attend to the way in which shared imaginative engagement with supernatural agents can foster radicalization (Shults and Gore 2018) and hinder immigrant integration (Shults et al. 2018). However, reading Wildman's case study also helped me to remember (some of) my own biases. Because of my worry about the deleterious consequences of religious belief and behavior on human life, it is all too easy for me to forget or ignore the way in which they can also help some individuals avoid radicalization and enable some groups to integrate peacefully within new cultures. Luckily, other team members have been there to be sure we include these dynamics in our models as well. I still think that religion is now maladaptive in most of the contexts in which contemporary humans live, and that we no longer need to appeal to supernatural agents and authorities to make sense of the world and act sensibly in society. That is my moral stance. In my view, one of the most valuable aspects of computer modeling and social simulation is that it forces me to reflect critically on this stance and to defend (or alter) it in open ethical dialogue.

I liked the concreteness of Dignum's case, which provided a nice balance to the somewhat more abstract models discussed by me and Wildman. I found myself imagining the busy café, and the variety of external behaviors driven by the inner mental states and norms of its simulated customers responding within its more or less smoky atmosphere. This concreteness also made it easier to identify the practical import of Dignum's meta-ethical reflections. Focusing on the philosophical component, one can see how the distinction between legal, social, and private norms shaping each agent highlights the complexity of answering questions about "the good" and "the right." Which values orient what sorts of behaviors can vary even within the same individual, depending on contextual cues that trigger dispositions toward following legal, social and/or private norms. Focusing on the scientific component, one can interpret her warning against simplistic notions of rationality and her emphasis on the complexity of mental, physical, emotional and social factors that impact human reasoning as a nod toward the importance of insights provided by evolutionary cognitive science and related fields, which have challenged the hegemony of "rational choice" models. Human reasoning, including ethical reasoning, is "bounded" – shaped by moral equipment that evolved in ancestral environments where relatively automatic, emotionally grounded and biased reactions often provided more survival advantage than careful reflection on rules or calm calculation of costs and benefits. Computer scientists may be able to help clarify the adaptive challenges facing humanity in our current ecological environment by developing models and simulations that incorporate these complexities within their causal architectures and experimental designs.

6.2 Wildman's Response to Shults and Dignum

Shults notes that the NoRM model lands smack in the middle of a furious debate in which the model prefers those on the side of secularism and naturalism. These scholars argue that religion is mostly dangerous in our time, supernatural agents are fictions, we don't need gods to be moral, and we are far better off with secular institutions and forms of social organization focusing on the equality and dignity of all people. Scholars on the other side, not as well represented in the design assumptions of NoRM, often argue that religion needs tweaking not abandonment, not least because supernatural agents are real and our moral deliberations have to take account of their moral interests. This is a good reminder of how modeling and simulation can be bent to social-engineering purposes. But Shults's focus on moral motivations for pursuing a modeling project doesn't come to grips with the less obvious but perhaps more important ethical considerations associated with building computational models: the assumptions baked into design decisions. Such decisions can seem obvious at the time when a team makes them, especially in the sense that the team is collectively unaware of plausible alternatives or the question of alternative designs doesn't arise. In retrospect, however, with enough critical distance, such design decisions can be unmasked as morally freighted and expressive of a thoughtless adoption of one design idea rather than alternatives, which though once invisible are suddenly painfully obvious. I would love to discuss with Shults the moral freighting not of the motivations for starting a modeling project but of the design decisions made along the way. The case study I presented mentions the ethics of motivation for initiating a modeling project but then stresses ethical evaluation of design decisions. Both types of ethical evaluation are essential.

Dignum's case study has the special virtue of directly modeling an explicitly ethics-related process – specifically, the effect of introducing a new anti-smoking law in light of existing norms at the legal, social, and private levels. The other two case studies are ethics-relevant at the levels of motivation and implications but do not set out explicitly to model an ethics-related process. This serves as a valuable reminder that computational modelers can make themselves useful by applying their skills to the analysis of ethical processes, both in general senses (from the formation of norms to conditions for norms changing and spreading) and in concrete senses (from ethical controversies to specific policy changes). Ethicists and policy professionals bring a lot of experience and insight to their work but I feel sure such insights could benefit from being tested in the distinctive way permitted by computational models of the sort that Dignum illustrates. Like Shults, Dignum doesn't explicitly consider the ethical import of design decisions made

during the modeling process. She does note the dangers associated with presuming rational actors, which is a welcome surfacing of the risks of design decisions, but this is not explicitly an ethical concern. Dignum's presentation of the tripartite classification of norms reads like the description of a neutral fact about human life. I would expect such high-level classifications to be more ethically loaded than this neutral presentation suggests. For instance, there is a profound entanglement between social and private norms to the point that the range of private norms typically available for adoption and cultivation critically depends on a socially constructed environment, and social institutions critically depend on private norms that those institutions can't themselves produce. Simplifying this complex entanglement is likely to prevent associated models from relevantly representing the ethical dimensions of the relationship between individuals and their enfolding societies. Correspondingly, modelers might regard as an ethical issue the very design decision about the classification of norms.

6.3 Dignum's Response to Shults and Wildman

I very much appreciated the approaches described by Shults and Wildman. It is very interesting to see the application of computational methods of simulation to the study of religion. My overall impression is that in a sense, their focus is on the practical implications of the works presented. That is about the understanding of religious practice, or lack of it, and the formation of extremist groups. Whereas, in my case, the main issue is understanding the implications of taking a certain scientific approach to modelling and simulation. I would like to further discuss this with both, in order to understand whether they also see this difference in the same way.

The case presented by Shults is grounded on real data and aims to understand the "evolution" of religiousness in the last decades and by doing so, raises the very important issue of how to understand and analyze simulation outcomes. In particular the issue of correlation versus causality. The simulation was calibrated and tested using existing data on religious practice as reported in the International Social Survey Programme (ISSP). Such approaches are able to identify correlations between variables or variable sets (e.g. education level and religious practice) but are not sufficient to identify causality. However, often results are interpreted as describing a causality (e.g. higher levels of education lead to decreasing religious practice). Ensuring the correct interpretation of results and removing any pretense of causality is an important ethical obligation of researchers and modelers. This falls under the principle of practical meta-ethics discussed in the Introduction of this article, where we argue that it is important for M&S professionals to be more explicit about values that guide their ethical judgments. The case presented by Shults demonstrates that these obligations should be extended to include explicit descriptions of the processes by which ethical judgements are made based on simulations.

Wildman's case illustrates his commitment to a consequentalistic, not deontological, approach. Modeling agents based on the perceived actions taking place in a given group or society is consistent with such meta-ethical view. That is, the focus of analysis is the consequences. I particularly appreciated how Wildman reflects on the ethics of the modelers themselves from the perspective of the scientific meta-ethics. Wildman does not provide enough details on the architecture of the agents in his model, but from the description I infer that the agents follow a rational model of utility-optimization, in this case of their 'wellbeing' either in or outside the group. Depending on how the model implements the influence of the group leader, an 'assassination' strategy as the one that emerged, is not too surprising. Where it concerns the practical meta-ethics component, I found it illuminating to see that Wildman relates it strictly to philosophical ethical theories. In particular, the remark concerning the unacceptability of assassination in a deontological stance, made me realize that in my own work, when I refer to a deontological stance, I take a more abstract approach, in the sense that I would consider an agent to be of a deontological nature if it acts according to *any* (pre-defined) set of normative rules (not only in relation to rules such as those prescribed by e.g. the categorical imperative or another philosophical theory of right and wrong). Such deontic agents are different in nature from optimization-based agents, leading to different simulation results.

Reading Wildman's description made me realize that from my perspective, I do not relate the practical meta-ethics implemented in my work to the philosophical and scientific meta-ethics, but take them to describe the choice between what I describe above as rational versus deontic agent models.

7. CONCLUSION

There is a great variety of ethical issues connected with M&S, and not all have been touched on here. For instance, models based on big data analytics, particularly when they are used to derive strategies for communication and advertising, are ethically controverted due to legitimate privacy concerns. We have not dealt with responsibilities related to the presentation of results, accountability to stakeholders, and open access to data and statistical analyses. Even in cases we haven't considered, however, we expect the ethical framework exemplified in this paper to apply. Table 2 provides an overview of some guidelines for applying the meta-ethical framework we outlined in the Introduction.

Table 2: Guidelines for addressing the components of the meta-ethical framework.

Component	Guidelines
Philosophical meta-ethics	• Specify assumptions about 'the good' and the criteria by which one evaluates the 'rightness' of a modeling activity • Identify the overarching ethical theory (e.g. consequentialism, deontology, virtue ethics) that guides one's professional work
Scientific meta-ethics	• Allow one's philosophical meta-reflections to be informed and constrained by insights from the evolutionary sciences • Incorporate those insights into models of human behavior (whenever relevant and possible)
Practical import	• Be explicit about the ethical values guiding the construction of models and experimental designs • Acknowledge the nature and power of the moral equipment that is part of our phylogenetic inheritance

This framework for ethical analysis focused on meta-ethical concerns because they are all too easily ignored when our attention is absorbed in the concrete problem-solving tasks of M&S. In the bulk of this paper each author tried to apply the framework to a specific case study, making explicit our own ethical reflection on the process of designing and executing a computational model. We also responded to one another in an attempt to initiate a broader conversation on the importance of meta-ethical analysis in the construction and evaluation of computer models and simulations. The philosophical, scientific and practical components of the framework are relevant for all sorts of models, but they become crucially important when it comes to modeling human behaviors within artificial societies. We hope this panel paper will spark a broader conversation about these meta-ethical issues, and wider use of the guidelines for applying them.

ACKNOWLEDGEMENTS

F. LeRon Shults and Wesley J. Wildman acknowledge support from The Research Council of Norway for the Modeling Religion in Norway (MODRN) project (grant #250449).

REFERENCES

Dechesne, F., G. D. Tosto, V. Dignum, and F. Dignum. 2014. "No Smoking Here: Values, Norms and Culture in Multi-Agent Systems." *Artificial Intelligence and Law* 21(1):79–107.

Dignum, V. 2017. "Social Agents: Bridging Simulation and Engineering." *Communications of the ACM* 60(11):32-34.

Du, H. and P. Chi. 2016. "War, Worries, and Religiousness." *Social Psychological and Personality Science* 7(5):444–451.

Gore, R., C. Lemos, F. L. Shults, and W. J. Wildman. 2018. "Forecasting Changes in Religiosity and Existential Security with an Agent-Based Model." *Journal of Artificial Societies and Social Simulation* 21(1):1–31.

Hofstede, G. 2001. *Culture's Consequences, Comparing Values, Behaviors, Institutions, and Organizations Across Nations*. London: Sage Publications.

Hungerman, D. M. 2014. "The Effect of Education on Religion: Evidence from Compulsory Schooling Laws." *Journal Of Economic Behavior & Organization* 104(2): 52–63.

McLeod, J. 1986. "But, Mr. President - Is It Ethical?" In *Proceedings of the 1986 Winter Simulation Conference*, edited by J. Wilson, J. Henriksen, and S. Roberts, 1–3. San Diego, California: Society for Computer Simulation International.

Norris, P. and R. Inglehart. 2015. "Are High Levels of Existential Security Conducive to Secularization? A Response to Our Critics." *The Changing World Religion Map*, edited by S.D. Brunn. Dordrecht: Springer.

Oren, T. I., M. S. Elzas, I. Smit, and L. G. Birta. 2002. "Code of Professional Ethics for Simulationists." In *Summer Computer Simulation Conference*, edited by Oren, T. I., M. S. Elzas, I. Smit, and L. G. Birta, 434–435. San Diego, California: Society for Computer Simulation International.

Paul, G. 2009. "The Chronic Dependence of Popular Religiosity upon Dysfunctional Psychosociological Conditions." *Evolutionary Psychology* 7(3):398–441.

Paul, G. S. 2005. "Cross-National Correlations of Quantifiable Societal Health with Popular Religiosity and Secularism in the Prosperous Democracies." *Journal of Religion & Society* 7(1):1–27.

Rawls, J. 2005. *A Theory of Justice: Original Edition*. Reissue edition. Cambridge, Massachusetts : Belknap Press.

Sen, A. 2003. "Human Development Index: Methodology and Measurement." *https://ora.ox.ac.uk/objects/uuid:01108a71-3112-4a84-a19e-1ebaa9a9f127*, accessed August 18th, 2018.

Shults, F. L. 2014. *Theology after the Birth of God : Atheist Conceptions in Cognition and Culture*. New York: Palgrave Macmillan.

Shults, F. L. 2015. "How to Survive the Anthropocene: Adaptive Atheism and the Evolution of Homo Deiparensis." *Religions* 6(1):1–18.

Shults, F. L. 2018. *Practicing Safe Sects: Religious Reproduction in Scientific and Philosophical Perspective*. Leiden: Brill Academic.

Shults, F. L and Ross Gore. 2018. "Modeling Radicalization and Violent Extremism." In *Proceedings of the 2018 Social Simulation Conference*, edited by H. Verhagen, M. Borit, G. Bravo, and N. Wijermans, 1–4. Stockholm, Sweden.

Shults, F. L., R. Gore, W. J. Wildman, J. E. Lane, C. Lynch, and M. Toft. 2017. "Mutually Escalating Religious Violence: A Generative Computational Model." In *Proceedings of the 2017 Social Simulation Conference*, edited by D. Payne, 1-12. Dublin, Ireland.

Shults, F. L. and W. J. Wildman. In press. "Ethics, Computer Simulation, and the Future of Humanity." In *Human Simulation*, edited by S. Diallo, W. Wildman, F. L. Shults, and A. Tolk. Berlin: Springer.

Shults, F. L, W. J. Wildman, S. Diallo, I. Puga-Gonzalez, and D. Voas. 2018. "The Virtual Society Analytics Platform." In *Proceedings of the 2018 Social Simulation Conference*, edited by H. Verhagen, M. Borit, G. Bravo, and N. Wijermans, 1–12. Stockholm, Sweden.

Zuckerman, P., L. W. Galen, and F. L. Pasquale. 2016. *The Nonreligious: Understanding Secular People and Societies*. New York: Oxford University Press.

Zussman, A. 2014. "The Effect of Political Violence on Religiosity." *Journal of Economic Behavior and Organization* 104(3): 64–83.

AUTHOR BIOGRAPHIES

F. LERON SHULTS is Professor at the Institute for Religion, Philosophy, and History at the University of Agder in Kristiansand, Norway, and Director of the Center for Modeling Social Systems at NORCE. Author or editor of 17 books and 80 articles and book chapters, Shults is Principal Investigator on the Modeling Religion in Norway project, funded by the Research Council of Norway. His email address is leron.shults@uia.no.

WESLEY J. WILDMAN is Professor of Philosophy, Theology, and Ethics at Boston University and Executive Director of the Center for Mind and Culture in Boston, USA. A philosopher of religion specializing in the scientific study of religion, Wildman is author or editor of 17 books and over 120 articles and book chapters, and has been working in social and human simulation for the last ten years. His email address is wwildman@bu.edu.

VIRGINIA DIGNUM is Associate Professor at the Faculty of Technology, Policy and Management, Delft University of Technology. She is an expert in the ethical and societal impact of artificial intelligent systems and value sensitive software development. Dignum has published dozens of publications in these areas and led (or participated in) many different collaborative research projects. Her email address is m.v.dignum@tudelft.nl.