ETHICS AND AGENT-BASED COMPUTATIONAL SOCIAL MODELS: DO IMPLEMENTATIONS OF SOCIAL SIMULATIONS HAVE MORAL SIGNIFICANCE?

Nuno David

Iscte – Instituto Universitário de Lisboa (Portugal)

Nuno.David@iscte-iul.pt

EXTENDED ABSTRACT

Social simulations and implementations

The focus of social simulation on representing the social world suggests an investigation of whether its implementations are inherently value-laden. The purpose of algorithms is both functional and representational. Some algorithms may be said to have functional purposes, such as deleting an email or ordering a list. The purpose in some simulations may be said to be essentially representational – to represent other things in the world, such as a model to investigate the problem of ethnical segregation, as in Schelling's (1971). Algorithms may also be purposely designed for certain values, according to certain requirements, such as protecting or undermining privacy. Despite articulating values, algorithms need not be essentially value-laden, that is, they need not comprise essential value judgements. In general, it can be said that an algorithm comprises an essential value judgement if, for the same purpose and everything else being equal, designers who accept different value judgements would have rational reasons for designing different algorithms (Kraemer et al., 2011).

A comprehensive way to approach the purpose of a computational artefact is to take as a starting point that computerization involves an implementation process, of which the algorithm is only an ingredient. I consider the extent to which implementations in social simulation have moral significance on the account of essential value judgements. I use the term social simulations to refer to computational models of the social world in the field of agent-based computational social science. Modelling the social world involves more than representing facts of the physical world, including values and institutional facts whose meanings themselves in the social world depend on human agreement. Insofar as we accept that simulations model social reality, two intrincate questions are raised. At stake is, on the one hand, what implementation is and how it participates in the constitution of a social simulation. And, on the other hand, whether representing the social world by simulation defines its implementations as inherently comprising essential value judgements.

Essential value judgments are closely related to the flexible way in which the implementation of simulations is interpreted. Insofar as computers are semantically interpreted, intended to represent other things in the world, computers can be said to have interpretive flexibility. This is similar to Moor's concept of logical malleability (1985), the fact that the states of computers may be used to stand for anything representable in terms of inputs, outputs and logical operators. If simulations are meant to represent cultural, social and political issues, it is worth addressing the extent to which their designs have moral significance. Not just for the sake of epistemic issues of truth with respect to what they model but for the fact that simulations are interpreted by their users.

The argument I discuss regarding implementation and its moral significance is outlined in three parts, which I briefly introduce below.

Interpretative flexibility

The finding that the meaning of a technology is not homogeneous among its users is corroborated in different scientific and philosophic domains. Studies in social studies of technology, and philosophy of technology argue that the early stages of design have interpretative flexibility, meaning that design responds to the requirements of different social groups and values, which eventually define the technical functions that the artefact will provide. I advocate that social simulations have high interpretative flexibility, which remains beyond design stages. The intention to represent institutions in social simulation makes implementations susceptible to essential value-judgments, deeming simulations interpretively flexible. Consider Schelling's model and the controversy over the meaning of segregation. That is, whether the resulting patterns result from the agents' individual preferences or from illusory mathematical artifacts. The last claim is placed to the extent that the model does not take proper account of the institutional factors of social reality (see e.g. Forsé and Parodi, 2010). Interpretative flexibility is found at two levels. Designers who accept different value judgements may have rational reasons for proposing different implementations. And for a given implementation, individuals who accept differing value judgements may have rational reasons for interpreting the implementation in different ways.

Representation gaps

The purpose of representing the social world through computers implies distinguishing between malfunction and representation gaps. Representation gaps in social simulation go hand in hand with interpretative flexibility. Representations underscore the role of meaning in the purpose of social simulations and whether such meaning is coherently maintained among the conceptual and computational models that make up the implementation process. Representations in computers depend on the existence of interpretations, which require programmers and/or users. Once a set of stipulative definitions is read as a model and the latter implemented on computers, the new model becomes a specification-of-a-model: the definitions are intended to be implemented for the purpose of representation. The model is given a normative mood towards implementing the representations that the model ought to specify. Semantic gaps between the specification and the implementation give rise to representation gaps: When representations expressed through an implementation do not live up to the representations as intended by the specification, that counts as a representation gap.

Hence, if the computer hardware and its resident software are assumed to work properly, at least two reasons exist for an implementation not doing as intended. The first is miscomputing (Fresco and Primiero, 2013). Miscomputations are objective, non-conformities between an implementation and its functional specification, which refer to verification stages in software development. Another reason concerns representation gaps between an implementation and any forerunner specification, in that the implementation may not represent what was specified. This falls within the scope of validation. It is not always clear whether representation gaps result from poor implementation or from the epistemic difficulty of representing social reality in computational models. Several reasons may be advanced: such technical reasons as the use of floating-point variables, which may lack enough precision to represent intended aspects of the model, or epistemic reasons, given the low syntactic complexity of formal and programming languages and the resulting difficulties in expressing social reality.

Techno-symbolic ends

Another way of looking at interpretive flexibility is to observe that implementations have – cumulatively with their technical purposes – symbolic ends. The purpose of a social simulation is determined not only by its computational structure (the computer function) but by its intended meaning in the social world. Meaning, and not just technical function, makes implementations in social simulation morally assessable. Implementation has techno-symbolic ends, which hinge on both technical function and meaning in terms of the institutional world. The computational artefact is interpretatively flexible. If it is morally assessed, this requires looking into the artefact's implementation.

High interpretive flexibility reflects the fact that implementations of social simulations are hardly neutral with respect to the purpose of representing institutions, which depend on different value judgments. For instance, in Schelling's model a token is said to be tolerant if it does not want a majority of coloured-like tokens in his/her neighbourhood. Regardless of how the relation between tolerance and segregation is mathematically interpreted, this implementation carries an essential value judgement. Whether the threshold is above or below the majority, there is no objective fact involved in specifying that a certain threshold represents tolerant, somewhat tolerant or intolerant agents. To do so, eventually amounts to take a moral stand through the artefact. Different designers may have different value judgements, possibly conflicting ones, despite being rationally justified.

Many social simulations are presented as purposely 'abstract', of which Schelling's is a canonical example. Some authors present them as 'metaphorical'. Indeed, 'abstract', in this sense, seems not to convey value neutrality, but interpretative flexibility. Flexibility is the condition for the purpose of the simulation as a metaphorical vehicle of representation of the social world, capable of coping with different value judgements of the institutional world that the computational artefact is specified to represent. From empirical and moral perspectives, high interpretive flexibility warns us that social simulations should always be implemented with open-source code. Greater digital skills will be required among those individuals and groups affected by the resulting policies of simulations, in order to be able to critically interpret computer simulations of the social world.

KEYWORDS: social simulation, ethics, interpretative flexibility, implementation.

REFERENCES

Forsé, M., & Parodi, M. (2010). "Low levels of ethnic intolerance do not create large ghettos: A discussion about an interpretation of Schelling's model". *L'Année sociologique, 60*(2), pp. 445-473.

Fresco, N., & Primiero, G. (2013). "Miscomputation". *Philosophy and Technology, 26,* pp. 253-272.

Kraemer, F., van Overveld, K., & Peterson, M. (2011). "Is there an ethics of algorithms?". *Ethics and Information Technology, 13*, p. 251.

Moor, J. (1985). "What is computer ethics?". *Metaphilosophy, 16*, pp. 266-275.

Schelling, T.C. (1971). "Dynamic models of segregation". *Journal of Mathematical Sociology, 1*, pp. 143-186.