

# Simulation and support in ethical decision making

Iordanis Kavathatzopoulos, Mikael Laaksoharju and Christian Rick

Uppsala University

Dept. of Information Technology, Human-Computer Interaction

[iordanis@it.uu.se](mailto:iordanis@it.uu.se)

## Abstract

The paper presents the structure and function of an ethical micro world simulation and of a support system in ethical problem-solving and decision-making. The ethical micro world simulation models realistic scenarios with interacting independent stakeholders. Users of the simulation are triggered to make autonomous decisions in dilemmas arising in the interaction between stakeholders. The goals are to investigate the possible approaches to implement the psychological approach to ethical problem-solving and decision-making, and to stimulate higher ethical competence. The ethical support system is based on the theory of autonomy. By using it thinking is guided away from heteronomy and toward autonomy. Its basic features are: 1) not allow the user to use the system as a moral authority, 2) not present a ready made set of moral principles and values, 3) help the user to be unconstrained by moral fixations and authorities, 4) help the user to organize and analyze the facts, 5) help the user to weight the relevant values and principles against each other, 6) help the user to solve the moral problem at hand systematically, 7) force the user to motivate his/her decisions in regard to the relevant interests and values.

## Introduction

Information technology has many advantages that can be used for the promotion of ethical competence. It saves time and space, it has an enormous memory storage capacity, it can process and reorganize information fast and reliably, etc. Recent technical developments in particular, which give us the possibility to construct advanced games and simulate the complexity of reality in micro worlds, may further broaden the spectrum of opportunities and possibilities for support in ethical problem solving and decision making. In this paper our efforts to construct an ethical problem-solving support system and an ethical micro world are presented.

There are, however, certain important issues to consider before building such systems. The confounding of moral values with psychological processes can create many problems and sometimes makes it impossible (Blasi, 1980; Greene et al., 2004; Haidt, 2001; Jackson, 1994; Jaffee & Hyde, 2000). Accordingly, our theoretical basis is that successful information technology tools in ethics are those that are adapted exclusively on psychological problem-solving and decision-making processes.

When we are planning to use information technology tools to support ethical decision making we usually run the risk of disregarding the psychological skill aspects of ethical competence. The classical approach focuses normally on informing about moral philosophy, presenting lists of

principles and stakeholder interests, expressing general or contextual principles, or simply producing moral solutions based on predefined normative values (Collins & Miller, 1992; Gotterbarn & Rogerson, 2002; Pfeiffer, 1999; Barth et al., 2007). Creating and using information technology tools based primarily on this classical approach certainly has its strengths, but it also has many weaknesses (Winograd, 1995; Friedman, Kahn & Borning, 2006). Ethical competence can be defined as based on the psychological ability described as autonomy. However, this skill is not so easy to use in real situations. Psychological research has shown that plenty of time and certain conditions are demanded before people can acquire and use the ethical ability of autonomy (Piaget, 1932; Kohlberg, 1985; Schwartz, 2000; Sunstein, 2005). When people face a moral problem they have great difficulties not confusing moral goals, values, feelings and emotions with the decision-making and problem-solving processes and the methods adopted for the solution of the problem. Usually, they do not clearly see the context of the problem nor do they analyze it in the same way they often do with problems of nature. In psychological theory this is described as the moral phase of heteronomy, which in contrast to autonomy, means that the individual does not use functional problem-solving strategies, that is, critical and holistic thinking. Autonomous and critical moral thinking is difficult, more difficult than autonomous technical thinking. In the searching to promote ethical competence we need to be assured that the autonomous ethical thinking is indeed stimulated by the support tools we use. Using information technology to support the acquisition and use of ethical autonomy is due to the special qualities and possibilities of this technology.

The use of real life simulations by decision makers may help them to learn easier how to handle morally complex and controversial situations satisfactorily. One way to do this is by connecting the progress of the simulation to the concrete way users treat moral problems rather than to general normative aspects of given solutions. For example, this can be done by incorporating in the simulation the interests, values, feelings, etc, of stakeholders whose reaction may influence the development of the simulation process.

Furthermore information technology tools have great advantages according to the hypothesis of autonomy. Their memory storage capacity is enormous. They are excellent in doing systematic work and analysis of data. Just by using them as a data base or an expert system in the effort to solve a concrete moral problem, the user can get information about certain values and interests, as well as about alternative ways of action, that otherwise might be overlooked. Reminiscence of the diversity, variety and complexity of the actual moral problem could effectively block decision makers' natural tendency toward heteronomy, and stimulate autonomy.

## **An ethical micro world**

Ethick is a flexible and dynamic micro world simulator and research environment with the main purpose of studying ethical problem solving skills. The goals are to investigate the possible approaches to implement the psychological approach to ethical problem solving and decision making, and to stimulate higher ethical competence (Kavathatzopoulos, 2005). Within the context of the Ethick micro world the researcher can define virtual stakeholders and assign to them interests, drives as well as principles to obey. The dynamics in Ethick are natively derived from these interests and drives and how well they are satisfied in the current state of the micro

world. Each virtual stakeholder acts according to its own best interests which, if the simulated environment is designed purposefully, will cause conflicts with other stakeholders. Colliding interests will cause stress to the stakeholders and they might start acting less rationally and possibly also against principles they obey in order to regain a general state of satisfaction. They might even, if their trust in justice has been compromised, start to act against laws of the simulation. The test subject, i.e. the user of the simulator, is encouraged to identify arising conflicts and try to legislate in order to maintain the desired harmony in the simulation. Laws are created with a self explanatory step-by-step wizard and this is the tool the user has obtainable to control the course of events. The main interest for the research is not to judge which laws and what harmony the user chooses to define. The simulator does not know whether a created law is morally defensible according to any belief or moral philosophy, but will merely accommodate the dynamics to the new conditions. The focus of interest lies instead in *how* the user came up to a certain law. Which information he or she gathered before deciding and what alternatives there were considered. If a law is well founded to create stability in the micro world, the user will maintain better control over the proceedings. Since the stakeholders in the community are ignorant of moral incentives other than the principles of the simulation, they act only within these boundaries to maximise their own satisfaction. There is neither natural remorse nor revengefulness. However, there always is the possibility to mimic such less rational behaviour as values reacting on violations against principles very much at heart. Stress is an important factor of the dynamics since it will cause less predictability from the stakeholders and therefore serve as a penalty for neglecting to take in account important aspects of the conflict handling.

## **Functionality**

Some important and perhaps controversial simplifications are made to scale down the delicate problem of imitating and technically sustaining realistic dynamics in the simulation. Principles held by the stakeholders, like the principle of equality, religious incentives and principle of life etc. are more or less put on the level with legislated laws. Drives and interests are considered as variables whose weighted and summed up maximisation is the sole purpose of each stakeholder. The weights, which should be interpreted as a measure of urgency, are dynamically adjusted according to how important the satisfaction of that interest is for the stakeholder at each time instance. This is correlated with how well the state of the micro world is consistent with the principles of the stakeholder.

The motor behind each simulation is as earlier stated an unsupervised system of deterministic robots reacting to the state of their perception of the world. There is no script for how the simulations will proceed, which makes setting up a realistic simulation the most challenging part of each study. In order to measure true ethical skills and not only theoretical ability to adapt to foreign environments, the user has to be able to accept and regard the proceedings in the micro world as if they were episodes from real life. This is not an easy task for the researcher to define. Apart from the obvious visual simplifications, the main constraint is the interface of communication with the test subject. The stakeholders in the simulation cannot speak their minds and will not give a reason for their actions. The user has to find out the ruling drives by investigating satisfactions with the current state. Naturally this limits the number of variables to a minimal set of relevant values and principles since too much information will make the simulation appear completely impenetrable and random. For example, a simulation around religious conflicts is probably not helped by embroiling values and principles around both

starvation and greed. Even though reasons behind real life conflicts are often very complex and veiled the simulation as an unfamiliar media implies enough mystery and unpredictability even with only a handful of variables.

## **Metrics**

Ethick is no doubt a controversial tool. Disregarding the ethical issues in measuring ethics, the possibility of establishing credible metrics are highly questionable as there will be no obvious pattern of repetition of studies and therefore no possibility to compare results linearly. Even an attempt to replicate a simulation will most likely produce a slightly different outcome since the user's interactions as well as the stakeholders' drives are subject to timing. The boundless divergence of simulation patterns causes problems with comparing results from different test runs. How can we compare two test results when the shape of the metrics is profoundly different? Can we at all compare two test subjects if one for instance brought control over the simulation with founding one law and the other achieved this with ten laws for different possibilities? Neither of them can without further investigation be accused for lacking the sought-for competence. Also we need to keep in mind that the achievement of a good final result is not necessarily what is interesting, although it surely serves the purpose of giving confidence back to the user. We need to accept that the obvious and unbiased metrics are still disguised but even without yet defining the absolute measure we have to determine what we are interested in to look for. Basically we want to bring into light the tendency of uncertainty that is a symptom of autonomous thinking. Suggesting a number of alternative laws for each conflict is an indicator of the test subject's irresolution but it can also be interpreted as reluctance to make a final decision. Depending on how the test subject comes to the conclusion we hope to be able to determine an indication towards autonomy.

For each conflict situation we propose to count:

The number of alternative laws that the test subject formulates.

For each of these alternative laws we count:

The number of investigations of interests and relevant situations.

For each of these interests and relevant situations, count:

The number of investigated possibilities and risks.

Some of these investigations are easy to keep track of simply by supplying the information in a way so that the test subject has to ask for it. In that case we only measure the number of interactive questions asked. Other estimations are less obvious to measure. How do we get a concept of the number of possibilities and risks the test subject is taking into account with each suggestion of a law? One approach is to let the stakeholders in the simulation question the law and ask for explanations why it is valuable to them. In that case we need to keep in mind that there is no script for the simulations and we cannot predict the exact situations so every interaction between the test subject and the simulation has to be handled in a general shape.

It is important to remember that this far no single real case study has been performed. The simulator is still under construction and solely based on theoretical assumptions and it is yet to find out whether the interface will actually be practicable and whether it will be possible to create simulations that proceed in a fashion resembling real life situations. If the perception of the mechanisms in the simulation would be too farfetched no actual benefit will be made from

learning how to solve moral dilemmas in such an environment. However, the most interesting part of this new approach is not the underlying simulation, but the proposition of dynamic metrics for ethical problem solving study.

## **A support system in ethical problem solving and decision making**

It is the purpose of this paper to examine the validity of a computerized tool for supporting professionals in moral decision making. Three main approaches of designing a computerized tool for supporting moral decision making are explored. The first being a computerization of the autonomy ethical analysis constructed and used previously as a paper and pen method (Kavathatzopoulos, 2003, 2004, 2005; Kavathatzopoulos, Persson & Åborg, 2004; Erlandsson & Kavathatzopoulos, 2005). The autonomy method consists of structuring the moral problem with relevant values and interests as rows in a matrix. The columns of the matrix consist of the considered actions. For each value-action pair the impact, both negative and positive, of the action on the value is noted. The hypothetical computerization of the autonomy method consists of an application containing a table. Functions for managing values and actions are given. For each field in the table the user is able to enter the effect the actions have on the relevant values. The second approach uses the computer's ability to hide irrelevant information. The same structure is used but instead of having the user fill in the impact of each action on each value in a table the user has a dialog with the computer by entering relevant values and actions. When the user is ready the application iterates over the gathered information prompting the user to enter the impact of each action on each value one by one. Both these two approaches structure the moral problem in two dimensions: Actions or solutions against values or interests. Neither of these two approaches uses the full power a computerized tool can offer. The third approach uses the computer's ability to manage advanced data structures by letting the user structure the moral problem in multiple dimensions. Using a graphical interface the user can draw the moral problem by specifying stakeholders and the relationship between these. What moral obligations they have, what rules they must follow and what interests or values they possess. Note that the application does not specify what these relations are; it only helps the user to visualize them. When the moral problem has been explored the application lets the user specify possible actions or solutions and then requires the user to specify the impact these have on the stakeholders and their interests, values and obligations.

While none of these applications exists prototypes have been developed and tested against certain criteria for validating that they fulfil the principle of autonomy. Each prototype has also been tested to ensure that the computerized tool provides a benefit over corresponding pen and paper methods if available.

To evaluate the prototypes a set of six criteria was developed based on the description of the ethical skill of autonomy as defined as to be unconstrained by moral fixations, authorities, uncontrolled reactions and to systematically and critically analyse the moral problem at hand in regard to the different interests and values involved (Kavathatzopoulos, 2005). In addition, in order to be able to motivate and argue for a moral decision there must be a logically sound and consistent argument for why the decision was the best possible among many alternative solutions (McFarland, 1991). Combining these two statements the set of six criteria for ensuring autonomy can be seen. These criteria for ensuring autonomy are as stated.

1. not present a ready made set of moral principles and values,
2. force the user to motivate his or her decisions in regard to the relevant interests and values,
3. help the user to systematically solve the moral problem at hand,
4. help the user to be unconstrained by moral fixations and authorities,
5. help the user to organize and analyse the facts,
6. help the user to weight the relevant values and principles against each other.

## **Method**

For each approach an appropriate prototype has been developed. For the first two approaches computerized mock-ups was used and for the final approach a pen and paper mock-up was developed.

The first prototype presented the user with an application consisting of a tabbed window with each tab representing a grouping of values. Each grouping of values may represent a stakeholder or any other constellation of values that belong together. Each tab contains a table with the values and interests as column headers and the actions, aspects and solutions as row headers.

The user can enter the relevant effects of each value-action pair in the relevant table cells. The application provides the user with means to manage the groupings of values, add values to a group and add actions. The actions are global to the whole problem and stay the same for all groups.

Two prototype versions of the second approach were developed. The first uses a shorter dialog sequence to prompt the user for all the relevant values and all the relevant solutions. It then lets the user analyze the problem by specifying the impact of each value-action pair. The second version uses an extension of the previous dialog sequence. The user can now freely enter values and actions in any order and can switch between entering values and entering actions at any time. When finished the user analyses the data as above.

The validation of the prototypes was made by comparing the abilities of the prototype with the specified criteria. For each criterion a reason for why or why not the prototype fulfilled that criterion is specified.

## **Results**

For each criterion a reason for why or why not the prototype fulfilled that criterion was specified. The results of the prototype evaluations are summarized in the table below. The first prototype only conforms to the first and fifth criteria. It fulfils the first criterion by not presenting a ready made set of moral principles and values and it fulfils the fifth criterion by helping the user to organize and analyse the facts.

Criteria	Prototype 1	Prototype 2	Prototype 3
not present a ready made set of moral principles and values	yes	yes	yes
force the user to motivate his or hers decisions in regard to the relevant interests and values	no	yes	yes
help the user to systematically solve the moral problem at hand	no	Yes	yes
help the user to be unconstrained by moral fixations and authorities	no	no	no
help the user to organize and analyse the facts	yes	yes	yes
help the user to weight the relevant values and principles against each other	no	no	no

Table 1: Results of the evaluations.

The second prototype comes out stronger than the first in that it conforms to four of the criteria. It does not help the user to be unconstrained by moral fixations and authorities and neither does it help the user to weight the relevant values and principles against each other.

The third prototype conforms to the exact same criteria as the second even though the approach taken differs greatly. As with the second prototype it does not help the user to be unconstrained by moral fixations and authorities and neither does it help the user to weight the relevant values and principles against each other.

## Discussion

The first prototype conforms to only two of the criteria but given proper training in moral decision making two of the shortcomings will not be as prominent as otherwise. Neither does the computerization provide any obvious benefits compared to the pen and paper method that it is modelled after. Just computerizing the tools of a method does not constitute a method of its own. To be more successful the prototype should have taken in consideration the entire method, not only the form of the method.

The second approach gives a better result in regard to conformity with the criteria. An interesting point is that both the first and second prototypes miss the fourth and sixth criteria. This shows a failure in capturing the inner workings of the method they are based on.

The third approach conforms to the exact same criteria as the second even if the approach taken differs greatly from that of the second prototype. It seems like the two approaches, using a two

dimensional data space and using a multidimensional data space, captures the same abilities of organizing, structuring and solving moral problems.

What all three prototype approaches fail to do is to help the user to be unconstrained by moral fixations and authorities. They also fail in helping the user to weight the relevant values and principles against each other. Using heuristics a computerized support system can learn to recognize heteronomous reasoning from the user based on previous interaction. An extension of the analysis not only takes into consideration the relations between the stakeholders in regard to the different possible solutions of the problem but also why a certain solution was preferred instead of the other solutions. Thus, forcing the user to explain why certain interests and obligations take precedence before other interests and obligations.

Another important aspect of computerized support systems is to deem if they give any improvement upon the non computerized methods they build on. All three approaches investigated in this paper have an analogous pen and paper method. The first two approaches are conceptualizations of the autonomy method capturing the same data and structuring the data in the same manner for analysis. The third approach expands upon the two dimensional relations of the autonomy method utilizing multidimensional relations. The second and third approach utilizes the abilities offered by computerization and thus helps the user to perform a better analysis than otherwise. In the first case by letting the user focus on specific relations and information. The latter by letting the user specify the relations between the stakeholders in a manner not possible using a matrix. The first approach fails to offer any benefits over the non-computerized method it conceptualizes.

None of these results have been proved statistically or by any other means other than reasoning about whether the prototypes conform to the criteria developed. It would be interesting to further investigate the validity of the usage of a computerized aid for moral decision making but such an endeavour is beyond this paper and the master's thesis it is a part of. Since the positive effect of education in moral decision making on an autonomy test has been proved (Kavathatzopoulos, 2004) further investigation in comparing the differences in autonomy test results of those with education using a computerized support system, without education using a computerized support system and those with only education would be of interest. Should the usage of both education and a computerized support system prove to be more effective than education alone, the worth of using computerized support systems for ethical decision making would be evident.

## References

Barth, A., Datta, A., Mitchell, J. C. & Nissenbaum, H. (2007), Privacy and contextual integrity: framework and applications, online at <http://www.adambarth.org/papers/barth-datta-mitchell-nissenbaum-2006.pdf> accessed 20.01.2007.

Blasi, A. (1980), Bridging moral cognition and moral action: A critical review of the literature, *Psychological Bulletin*, 88, 1-45.

Collins, W. R. & Miller, K. (1992), A paramedic method for computing professionals, *Journal of Systems and Software*, 13, 47-84.

Friedman, B., Kahn, P. H., Jr., & Borning, A (2006), Value sensitive design and information systems, online at <http://www.ischool.washington.edu/vsd/vsd-and-information-systems.pdf> accessed 20.01.2007.

Gotterbarn, D. & Rogerson, S. (2002), *Project Planning Software* [Computer software], East Tennessee State University.

Green, J.D., Nystrom, L.E., Engell, A.D., Darey, J.M., Cohen, J.D. (2004), The neural bases of cognitive conflict and control in moral judgment, *Neuron*, 44, 389-400.

Haidt, J. (2001), The emotional dog and its rational tail: A social intuitionist approach to moral judgment, *Psychological Review*, 108, 814-834.

Jackson, J. (1994), Coping with scepticism: About the philosopher's role in teaching ethical business, *Business Ethics: A European Review*, 3, 171-173.

Jaffee, S. & Hyde, J. S. (2000), Gender differences in moral orientation: A meta-analysis, *Psychological Bulletin*, 126, 703-726.

Erlandsson, M. & Kavathatzopoulos, I. (2005), Autonomy method – acquiring skills for ethical analysis of computerization in car driving. In Göran Collste, Sven Ove Hansson, Simon Rogerson, and Terell Ward Bynum (eds), *ETHICOMP 2005: Looking back to the future* [CD-ROM], Linköping University, Linköping.

Kavathatzopoulos, I. (2003), The use of information and communication technology in the training for ethical competence in business, *Journal of Business Ethics*, 48, 43-51.

Kavathatzopoulos, I. (2004), Making ethical decisions in professional life. In Henry Montgomery, Raan Lipshitz, and Berndt Brehmer (eds), *How professionals make decisions* (pp. 277-288), Lawrence Erlbaum Associates Publishers, Mahwah, NJ.

Kavathatzopoulos, I. (2005), Computers for ethical competence. In Göran Collste, Sven Ove Hansson, Simon Rogerson, and Terell Ward Bynum (eds), *ETHICOMP 2005: Looking back to the future* [CD-ROM], Linköping University, Linköping.

Kavathatzopoulos, I., Persson, J. & Åborg, C. (2004), Individual learning and organizational change for ethical competence in the use of information technology tools. *The ETHICOMP Journal*, 1, 1, online journal at <http://www.ccsr.cse.dmu.ac.uk/journal/home.html>.

Kohlberg, L. (1985), *The Just Community: Approach to moral education in theory and practice*. In M. Berkowitz and F. Oser (eds), *Moral education: Theory and application* (pp. 27-87), Lawrence Erlbaum Associates, Hillsdale, NJ.

McFarland, M. C. (1990), Urgency of ethical standards intensifies in computer community, *IEEE Computer*, 23(3), 77-81.

Pfeiffer, R. S. (1999), *Ethics on the job: Cases and strategies*, Wadsworth, Belmont, CA.

Piaget, J. (1932), *The moral judgment of the child*, Routledge & Kegan Paul, London.

Schwartz, B. (2000), Self-determination: The tyranny of freedom, *American Psychologist*, 55, 79-88.

Sunstein, C. R. (2005), Moral heuristics, *Behavioral and Brain Sciences*, 28, 531-573.

Winograd, T. (1995), Computers, ethics and social responsibility. In D. G. Johnson and H. Nissebaum (eds), *Computers, Ethics and Social Values* (pp. 25-39), Prentice Hall, Upper Saddle River, NJ.