Review of Economics & Finance
Submitted on 27/09/2016
Article ID: 1923-7529-2017-02-17-14   Vladimir A. Masch

“Shifting the Paradigm” in Superintelligence

Dr. ***Vladimir A. Masch***

Risk Evaluation and Management, Inc.
94 Old Smalleytown Rd., Warren, NJ 07059, U.S.A.
Tel: +1-908-263-7497   E-mail: skipandscan@optonline.net

Abstract: Sharply increased uncertainty and possibility of catastrophes warrant a new approach to decision-making. To survive Superintelligence, mankind should downgrade its role - from “an agent” that has a will and a preservation goal of its own, to just a tool that yields the power of making decisions to humans – possibly *Risk-Constrained Optimization* (RCO).

RCO is a fundamentally novel system dealing with decision-making under radical uncertainty. Instead of “the best strategy” RCO constructs a “strategy, most acceptable to decision-makers.” RCO develops a number of candidate strategies, filters them and presents to the decision-makers a few reasonably good and safe candidates, easily adaptable to a broad range of future scenarios - likely, “black swan,” and even improbable. The final selection of the strategy to be implemented is performed judgmentally by decision-makers.

RCO overturns upside down Economics, Operations Research/Management Science, Decision Analysis, Scenario Planning, and Risk Management. The new paradigm of Superintelligence becomes *preservation of mankind*.

RCO is just a toolkit. It can be used in any system. But, as far as this author knows, RCO is presently unique in its capability to deal with radical uncertainty – moreover, by simple operations. It is therefore irreplaceable for Superintelligence.

Keywords: Artificial superintelligence; Confidence; Decision analysis; Risk-constrained optimization; Scenario planning; Multiscenario Multicriteria model; Strategic frontier

JEL Classifications: B41, C44, C61, C63

Abbreviations: ASI - Artificial Super Intelligence; DA - Decision Analysis; MMS - Multiscenario Multicriterial Stochastic; RCO - Risk-Constrained Optimization; RLC - Risk-Limiting Constraints

1. Introduction

The current wonderful achievements of science and technology will be good for mankind only if they are integrated with laws of nature. Currently they contradict one of these laws. Need for self-preservation should be considered a law of nature, so that both technical means and scientific methodologies of decision-making that are developed by mankind should be consistent with the concept of its self-preservation. That also refers to the AI as a whole, as well as to one of the most difficult parts of ASI - *decision-making under radical uncertainty* [Barrat (2013): 32-33; Lewis (1992)].

The 21st century combines progressively worsening global perils with radical uncertainty. We expect that catastrophes will happen, but we do not know when, where, their exact type and consequences, or on what scale. The current main goal of mankind is *raising prosperity for a finite* *period*. It is myopic and is largely responsible for the present state of the planet. It absolutely lacks safeguards. It must be replaced by the goal of *attaining unlimited-term sustainable survival of mankind in an acceptable state.* Correspondingly, the old "dodo paradigm" of mankind must be replaced by the new paradigm, which is *self-preservation* [Kuhn (1962)]. (The dodo birds lived in safety on an isolated uninhabited island and had no self-preservation instinct. They were quickly wiped out in the 16th century, when Europeans and rats arrived to the island.)

Under radical uncertainty, neither of the data available is reliable and, most importantly, we do not know the probabilities of future scenarios. That means we have to implement strategies that are flexible and that easily adapt, without catastrophic consequences, to any future scenario – likely, "black swan," or even scenarios with "unknown unknowns" that presently are considered impossible. The mandatory feature of a strategy becomes its *good adaptability with acceptable consequences for the underlying system.*

Moreover, we could have no *definite confidence* in our decisions, and decision-making should take this into account.

Enormous difficulties arise in decision-making, when we know practically nothing about the future, on the one hand, and on our capability to confront the potential dangers (including the danger of ASI [Barrat (2013): 8-18; Bostrom (2014); Chace (2015); Yampolsky (2015)]), on the other hand. As far as this author knows, no previous work has been performed to deal with these difficulties.

But that does not mean that mankind has to create all-powerful ASI agent systems, capable of making and implementing decisions. Sooner or later, such systems will have as a goal its own preservation, rather then the mankind's. They will fight with humans for resources and win. ASI should be limited in their capabilities and their functions. It should be no more than a tool. But an ASI tool can still be dangerous, if it is not so limited and simple that it does not have any risky components, such as capability to learn, and far-reaching sets of data.

Risk-Constrained Optimization (RCO), outlined in this paper, is just such a tool. RCO is a system dealing with one of the most difficult parts of ASI - decision-making under radical uncertainty. RCO is fundamentally novel, both conceptually and technically. It has two stages. At the first stage, enhanced *multiscenario multicriterial stochastic (MMS)* optimization models develop highly adaptable strategies and screens them to prevent unacceptable consequences in their *contingency plans* over a large range of scenarios – likely, "black swan," and even improbable. MMS become *optimizing filters* of strategies. MMS either modify and truncate the strategies or forbid them. At the end of the stage, RCO provides a set of *acceptable candidate strategies.*

At the second stage, screening of strategies continues by an ensemble of several "*synthetic*" *decision-making criteria* in the framework of "*strategic frontier*". The latter by-passes the difficulties of *selecting the level of confidence* by considering simultaneously *all* levels of confidence, from complete confidence to complete non-confidence. The second stage results in a small set of reasonably good and reasonably safe candidate strategies. The final selection of a strategy to be implemented is performed judgmentally by decision-makers.

2. Literature Review

2.1 Dangers of ASI

AI and robotics are transforming the world. Robots are likely to be performing 45% of manufacturing tasks by 2025 (vs. 10% today). The 2020 market is supposed to be $153B ($83B for robots and $70B for AI-based analytics), with 10-year disruptive effect of $14 to $33 trillions, including $8T to $9T in labor savings, with 47% of the US jobs having the potential of being automated and with increased inequality [Bank of America (2015)].

No doubt, the expected results are extremely beneficial for the creators of this revolution. But are they equally favorable for the mankind as a whole, even if they establish a safety net for the displaced people? Chace calls "economic singularity" a time, when "a majority of jobs can be performed more effectively, efficiently or economically by an AI than they can be done by a human" and everybody has a sufficient Universal Basic Income [Chace (2015): 55-56]. Would not it be great? Not necessarily [Dalai Lama and Brooks (2016)].

Stop calling it "technical progress," if it does not lead to a social progress and is dangerous for the society. Tech leaders begin to understand that. But they are concerned not about the people displaced – just about the possibility of strong populist reaction [Brockman (2017)].

The 21st century is full of perils and uncertainty. We are threatened by possible extinction of mankind, with most risk coming from human activities [Bostrom (2013); Cookson (2015)]. In these articles, the greatest threat comes from robots! But AI and especially ASI are much more hazardous. A large part of [Bostrom (2014)] is devoted to that danger. In particular, Chapter 10 of [Ibid.: 145-158] describes various types of AI and ASI systems, from simple question-answering oracles to powerful sovereigns, and analyzes the potential of their escape and the ensuing battle with mankind for resources, such as energy for supercomputers. The forecast is depressing.

The safest type, the tool, is software that simply "does what it is programmed to do" [Ibid.: 151]. It "does not raise any safety concerns even remotely analogous to the challenges discussed in this book" [Ibid.: 151]. But even tools can become dangerous. Failures of programming may lead to outcomes not *intended* by the programmer. We have to beware of combining such failures with ASI and supercomputers.

In "Afterword," added in the 2016 paperback reprint of [Ibid.], Bostrom communicates that, in spite of greater attention to issue of AI and ASI safety, its funding "is still two to three orders of magnitude less than is going into simply making machines smarter." [Bostrom (2016): 323]. The more so we need to be cautious.

Similar concerns are voiced in a number of other books, such as [Barrat (2013): 8-18; Chace (2015); Yampolsky (2015)]. All of them admit that "penalty for failure (which could be the result of a single false step) may be catastrophe" [Chace (2015): 180].

A recent meeting of technological executives pays attention to the existential threats of AI, but clearly not sufficient for effective control [Waters (2016)].

2.2 Uncertainty and optimization in ASI

Unfortunately, neither Uncertainty (especially radical) nor Optimization is in ASI yet. In a number of books on AI those subjects are absent even in their Index sections. ([Bostrom (2014)] mentions optimization many times, but not in the sense of mathematical programming that is the main focus of Operations Research.)

Bostrom deals in [Ibid.] with uncertainty in a very insufficient way. First, he writes about the "possible worlds" (presumably, scenarios). Are these worlds "prior" or "posterior" to our action? What was the action? How did we make the decision to act this way? Nobody knows, including Bostrom himself. *Because nobody knows how to make decision under radical uncertainty*. Of course, AI is not an exception. Ignorance about that subject is universal.

Then, somehow deriving utility measures for "possible worlds" and assigning to these worlds subjective probabilities (and changing the probabilities in Bayesian manner), Bostrom assesses the state of the multiscenario world by calculating mathematical expectation of utility [Ibid: 323].

Contrary to [Arrow and Hurwicz (1972); Luce and Raiffa (1957)], which require multiple criteria, he uses a single criterion.

True, Domingoes (2015) classifies and generalizes a number of ways to deal with uncertainty of such parameters as scenario probabilities. (No decision-making though.) Then, however, he admits that "… first we have to gather a very important, still-missing piece of the puzzle: how to learn from a very little data." [Ibid.: 175]. His solution is to generate missing data from anything that might resemble it – say, averages. Such an approach does not apply to decision-making under radical uncertainty in serious (complex and long-range) problems.

Neither are uncertainty and optimization treated well in AI planning [Pomerol (1997); Ransbotham (2016); Pollock (2004)]. In 1997, Pomerol remarked that "… AI has not paid enough attention to look-ahead reasoning, whose main components are uncertainty and preferences." [Pomerol (1997)] 19 years later, Ransbotham agrees: "Decisions that executives face don't necessarily fit into defined problems well suited for automation. At least for time being, countless decisions still require human engagement." [Ransbotham (2016)] Of course, we cannot even dream about optimization: "… there is no way to define optimality for plans that makes the finding of optimal plans the desideratum of rational decision-making" [Pollock (2004)].

2.3 We are more stupid than Amoebas (Not to Speak of Neanderthals)

How did organisms survive, starting from amoeba, with no or primitive brains, with abundance of dangers and under huge uncertainty, from the times immemorial to the present? Did they use the "natural" way of scenario-and-contingency planning? If they did, what was it? Let us consider the example provided in [Williamson (2010)]; it may lead us to the answer. A Neanderthal man sees a cave. He has to develop his strategy (to enter the cave or not) for two scenarios - either the cave is empty or there is a dangerous beast in it. How does he make a decision?

In terms of this paper, he approaches this conundrum with two dominant simple notions, inherited and instinctive. First, his sole criterion is catastrophe avoidance, catastrophe is being eaten; he does not (and cannot) calculate and maximize utility. Second, he considers it as a single problem with two scenarios and a single common strategy to pursue, rather than two problems with one scenario each and with two mutually independent plans that have to be somehow reconciled. How could one reconcile entering and not entering the cave?

His decision probably will be to avoid risk. Better safe than sorry. If he is sufficiently careful, he survives. Amoebas behave the same way.

Now we have computers that can do quintillions operations per second. But we also have economists and scenario planners whose survival does not depend on their decisions [Baker (2016)]. They do not have to choose between fight and flight. Therefore they obstruct the "natural" approach.

2.4 Scenario planning

The system closest to RCO still remains its predecessor, "the Zone of Uncertainty" (ZU) approach [Makarov and Melentyev (1973)]. In the 1960s, a substantial advance was made by a group of scientists from the USSR energy industry, which, under the influence of [Luce and Raiffa (1957)], used the concept of multiple Decision Analysis criteria to develop a scenario planning system called "the Zone of Uncertainty." The "Zone" was a set of candidate strategies considered the best under different criteria. To evaluate a strategy, the cost behavior of these candidates had to be evaluated on a large range of scenarios.

To derive the candidate strategies, the ZU system constructed and solved a number of single-scenario "What if" linear programming models. (Thus ZU does not pursue the two notions of the

Neanderthal. But as can be seen in Section 2.5, neither have they been followed in the present Western scenario planning. Only RCO returns to simple and “natural.”)

The ZU approach proved enormously progressive for its time. (But, as far as the author knows, the methodology is not widely used in Russia now, if it is used at all.)

The ZU approach introduced three important ideas:

a) Splitting variables into “strategic” and “operational” groups and thus clearly delineating each strategy;
b) Constructing contingency plans and evaluating candidate strategies by the totality of their “post-contingency-plan” outcomes for the whole range of scenarios;
c) Using several DA criteria for finding the best candidates (follow-up of [Luce and Raiffa (1957)]).

2.5 Current scenario planning not up to the ZU standard of quality

In 2006, a group of RAND scientists declared that “no systematic, general approach exists for finding robust strategies using the broad range of models and data often available to decision makers” [Lempert, *et al*. (2006): 514]. That evaluation must have included stochastic programming and scenario planning. The RAND methodology claimed to be such a general approach, but it included neither the three extremely important ideas of ZU listed above. Contrary to [Arrow and Hurwicz (1972); Luce and Raiffa (1957)]), it used, for instance, just a single criterion of minimax regret. In a private communication, it was also revealed that the authors were not aware of the RCO description already published in 2004 [Masch (2004)].

Work on RCO started in the 1960s. In 1967, this author already headed the 15-scenario location study for the Fiat automobile plant to be built in the USSR. (The study recommended the city of Tolyatti, and this recommendation was accepted.) The embryonic version of RCO, called “Robust Optimization” and expanding ideas of ZU and [Luce and Raiffa (1957)], was granted an USA patent in 1999 [Masch (1999)]. That version of RCO was successfully applied at one large corporation in 1992–1993 [Lindner, Jordan, Karwan (1994)]. However, RCO was drastically improved after 2000. It replaced single-scenario models by multiscenario multicriterial ones, with all the cardinal improvements stemming from it. It became a new system. Nine major improvements completely transformed the RCO methodology from ZU [Masch (2010): 426]. In essence, only the above three ideas are the heritage of the ZU approach in RCO.

This author cannot guarantee the verity of the strong statement above of [Lempert, *et al*. (2006)], which would, in turn, imply that all existing systems do not compare to the quality grade of ZU, achieved in the 1960s. It is always possible to miss something. However, the author relies on [Ibid.] in their survey of literature and is not repeating it. The author has no doubt, however, that both “the shift of paradigm” [Masch (2013)] and rest of the ensemble of models and computational methods used in RCO are currently unique. Even ZU methods of the 1960s are unknown and not used in the West, not to speak of their cardinal improvement by RCO.

The business world is already unhappy with the current methodology of scenario planning. According to a recent survey by Bain & Company, satisfaction with it dropped to less than 20 percent [Bain (2015)]. MIT professors agree: “… effect of scenario planning on executive judgment is almost nonexistent.” [Phadnis (2016)]. Perhaps because it was concentrated on likely scenarios and thus did not protect users from the financial crisis of 2007. However, if scenarios are of “black swan” type, some isolated successes do happen [Phadnis (2016)].

3. Risk-Constrained Optimization

The socio-economic problems mankind is facing in the 21st century are characterized by an overwhelming increase in complexity and uncertainty regarding the types of changes we can expect, as well as their scale, speed, and timing. Traditionally, almost nothing was known about future events, even in the short-term. Now “almost” has disappeared and we can only make guesses as to how the future will unfold. This is the state referred to as “radical uncertainty”. Radical uncertainty in the world as a whole, however, is only one of the factors contributing to uncertainty in any system, global or local. Let us not forget about complexity! In complex systems, even a full awareness of individual components and the laws governing their interaction is insufficient to infer the properties and behavior of the system as a whole [Johnson (2012); Lai and Han (2014)]. The system imposes additional (“systemic”) conditions and constraints that are hidden from the observer.

The challenge of this century is sustainability. We must find the ways to navigate skillfully and cautiously between wide ranges of potential dangers, both known and unknown, to ensure that the decisions we make are beneficial for the long-term survival of mankind. Even more difficult is to make mankind follow the right ways. Knowing human nature, the author is pessimistic, but we should at least know these ways. And that is the main purpose of the present paper.

But how to achieve sustainability if we do not know the potential risks? That cardinally changes the process of decision-making. Instead of *maximizing utility* in one form or other, now we should strive for *adaptability* of the strategy and *robustness* of the resulting system.

Adaptability means here the capability of the strategy to absorb the external shock of encountering any scenario – likely, unlikely, and even improbable – without generating excessively risky contingency plans. Accordingly, the *riskiness* of a scenario is important while how *realistic* it is – is not. (Section 3.1 outlines how the RCO decision-making process deals with scenario realism.)

Robustness or *sturdiness* means the capability of the system to withstand potential risks without creating an outcome that the decision-makers consider catastrophic.

The role of the decision-makers grows enormously. At any level, local or global, it is they that determine the survival of the system. Everything depends on their feeling of responsibility, their attitude to each of multiple potential risks, known and unknown, and their expertise (to a much lesser degree). They must participate in the strategy development process practically from the very beginning. By selecting risks to fight and the necessary degree of limiting those risks they in essence define the expected likelihood of the scenarios where those risk manifest.

That means that the essence of the proposed part of ASI completely changes. It does not imply anymore capability of computers and computational systems to exceed many times the capability of human brain [Barrat (2013): 8-18; Bostrom (2014)]. Its role becomes more modest and realistic – just to prepare for human decision-makers reasonably good and reasonably safe candidate strategies. Even that task still remains difficult to the max.

How the present study is associated with ASI? We can safely assume that ASI should provide good decisions. Optimization comes to mind. Optimization is not riskless under uncertainty, however – even with the present methods of scenario planning that do not protect from shocks of unexpected scenarios. RCO thoroughly screens and modifies the candidate strategies, trying to protect the underlying systems from catastrophes. As far as possible, RCO adds what ASI needs - *reasonable preservation of the system*.

Operational Research discipline started in England during WWII, when scientists explored holes in the bodies of planes returning from operations. Somebody reflected – these holes did not lead to catastrophes. Scientists should have studied the holes in the planes lost. Of course, in the 1940s that was impossible. Now we can simulate the "black swan" scenarios, and that is what RCO is doing.

3.1 The two-stage ensemble of RCO

Paradigms are likely to be shifted at times of crisis, when mankind has to overcome new difficult problems. To solve them, more powerful toolkits are required. Perhaps there can be no paradigm shift without a revolutionary change of the toolkit. We need both conceptually and technically novel toolkit that carries out self-preservation under conditions of radical uncertainty. That would completely distinguish this author's approach to ASI from any other work in that field. The mathematical (computational) toolkit here is *Risk-Constrained Optimization (RCO)*; its most complete but already partly outdated description is in [Masch (2013)]. The role of RCO in achieving the general goal of mankind (stated in Section 1) is outlined in [Masch (2015)].

Previously decisions were data-driven and, at best, related to a limited (usually small) set of likely future scenarios. Under the present conditions, we know nothing about even the short-term future. We have some ecological, technological, and economic information, but that is just a scintilla of what is needed, and is completely unreliable. So we have to make decisions predominantly on the basis of our emotions, such as confidence and "animal spirits." Therefore RCO rejects the very concept of "the correct" or "the best strategy," replacing it with a "strategy most acceptable to decision-makers" (see Section 3.2). To protect the latter from making a serious mistake in selecting a strategy, we have to develop flexible candidate strategies, easily, without substantial risks, *adaptable* to any potential future scenarios, likely or not. These candidates must result in a system that is sufficiently *robust* to withstand as many conceivable shocks as possible. We need reasonably safe candidate strategies that are like space suits, protected from both "known" and "unknown" unknowns [Rumsfeld (2011)] – somewhat different suit # 1, # 2, etc. (As examples of "unknown unknowns," Rumsfeld gives 9/11 and Pearl Harbor. He sadly qualifies this category "to be the difficult one".)

RCO does exactly that development in two stages, both involving optimization but using it for analysis, ferreting out risks, and screening out or truncating the bad candidates, rather than for selection of the best. It starts with creating scenarios. As outlined in Section 3, the realism (or likelihood) of scenarios is not important, therefore they can be constructed by a simple combinatorial technique. (That narrows the role of Complexity Theory, which basically has been created to construct realistic scenarios. Now we need just simple and approximate values of outcomes of extreme behavior of the complex system.) RCO is not concerned about on what scenario basis it develops the strategy. All features of basis except its riskiness become almost irrelevant.

The first stage performs *strong screening.* Its main tool is enhanced *multiscenario multicriterial stochastic* (MMS) model. The model contains a set of scenario submodels. It has two sets of criteria. The first set fights risks by applying a *catastrophe-avoidance algorithm* that imposes *risk-limiting constraints* (RLC); the second maximizes *utility*, in one form or another. ("Enhanced" denotes here models with imposed RLC.) The objective function of MMS contains mostly the utility criteria.

The process is iterative. It is started by solving the model without any RLC. Each scenario submodel generates a different *contingency plan.* If, in some plans, some outcomes present excessive risks, we add RLC as the upper limits on these risks and optimize again. Thus the obtained extreme solutions are sharply curtailed by sets of RLC. Each candidate strategy is found as a result of solving a new MMS model, with a different set of RLC. So the problem really becomes - what set of RLC has to be imposed on a reasonably good utility strategy to make it a reasonably safe space suit? Not “what to do,” but rather “what not to do.” (The values of strategic variables may be defined by these constraints.) The MMS models function as *optimizing filters.* Modifying the sets of RLC constraints, RCO generates a sufficient number of safe candidate strategies for further analysis.

The MMS models with RLC completely change the role and capability of Operations Research. In spite of their virtuosity, all optimization models and algorithms, created since the 1940s, are capable of finding no more than *extreme*, rather than *optimal*, solutions of deterministic problems and problems under moderate uncertainty. In the real world, no serious problem is of that type. In contrast, RCO goes as far as possible in addressing radical uncertainty. *Currently available OR models and algorithms are great for the abstract world. MMS makes OR useful for the real world.*

The two main advantages of MMS models are as follows. First, these models are the only tool that can find a flexible and robust strategy against the background of many risks and a large range of scenarios. (Some strategies may even be compromises that will never be obtained from applying any number of single-scenario models.) Second, by changing the set of RLC, or parametrically changing their right-hand sides, we can measure the impact of different risks and different constraints on the evolving strategies.

It should be particularly emphasized that imposition of RLC means that we find the relevant corresponding scenarios to be somewhat plausible. The “guess probability” that we originally assigned to that scenario should be thus somewhat raised. That is especially important for scenarios with initial zero probability. Each candidate strategy emerging from the first stage has therefore a specific set of scenario probabilities different from the original set.

Because of lack of space, I refer the readers to [Masch (2013)], which provides more detailed description of the process of strong screening.

When enough satisfactory candidate strategies are generated, they are subjected, at the second stage, to *weak screening* which replaces the current Decision Analysis (DA), that is reductionist and absolutely inadequate. Again, the goal is filtering rather than selection. In the current DA, all but one major criterion are “single” selection criteria (see Section 3.2). The decision-maker is assumed to know exactly, even under radical uncertainty, his level of confidence in scenario probabilities. It is either complete confidence, as in the weighted average and maximax criteria, or complete lack of confidence, as in the minimax payoff and minimax regret criteria. True, the “pessimism-optimism index” criterion is “synthetic”; it combines pessimism and optimism, but at some arbitrarily fixed intermediate level of confidence. The assumption of knowing the level of confidence is absurd – that level is a very bad subjective “unknown.” This assumption was made just to meet the mainstay requirement of all reductionist disciplines, to have a *single best solution.*

RCO is against the very idea of the best solution. Therefore it can not only replace all “single” criteria by novel “synthetic” ones, but also introduce a fundamentally new concept of “strategic frontier” (see Section 3.2), where the candidate strategies are compared with each other at *all* levels of confidence at once, from complete confidence to complete lack of it. We cannot determine the proper level of confidence; therefore we look at all levels. The powerful ensemble of strategic frontier and several “synthetic” criteria assures reasonably good final filtering of candidates. A few good and universally safe space suits, that remain after all filtering, are presented to the decision-makers for subjective judgmental selection.

An RCO-type system has four main advantages that will make it irreplaceable for addressing difficult problems presently facing mankind. First, it allows combining different contradictory theories and data in a single model, without pre-judging them, presenting them as different groups of scenarios and playing with their weights. Second, the system allows using very approximate input data, which in turn permits addressing off-the-cuff extremely complex problems that otherwise would require years of preliminary studies or cannot be approached at all. As well, the most unreliable part of input data, scenario probabilities, is changed in the RCO process. Third, special structure of MMS allows clustering large sets of scenario submodels, thus allowing solving optimization models with a very large number of scenarios. Fourth, RCO employs an ensemble of conceptually novel techniques that, in the author's possibly biased opinion, will eventually become an integral part of any reliable approach in long-term and complex planning and research problems. These techniques are:

a) Enhanced multiscenario multicriterial stochastic (MMS) optimization models that include the catastrophe-avoidance algorithm.

b) A complex of several synthetic filtering criteria of decision-making, used jointly in the framework of strategic frontier.

The totality of methods, algorithms, and models contained in the two stages of RCO make an *ensemble*. Senge (1990) suggests that a major breakthrough could result only from the combining of a special ensemble of efficient component technologies that come from diverse fields of science or technology, and only when all necessary components of that ensemble come together. He strongly emphasizes that the power of the ensemble comes mainly not from the individual components, but from their combined impact within the process. In his words, they form an inseparable ensemble "and are critical for each others 'success.' " [Ibid.]

3.2 Criteria of decision-making

Perhaps one of the general laws of nature can be formulated as: "An organism or a system will become extinct if it does not take sufficient care of its own self-preservation." The existing "dodo paradigm" of decision-making violates that law.

In a universally recognized psychological "pyramid of needs and wishes" the first priority belongs to physiological needs, such as breathing [Maslow (1943)]. The next are needs of safety of the individual and his "community." He may consider as such whatever he likes, from his family to whole mankind. Only after his needs are met, and he controls various kinds of risk in accordance with his attitudes about those risks, he can initiate satisfaction of his discretional wishes.

Conversely, economics begins immediately with the maximum satisfaction of wishes, thus making a hidden assumption about a guaranteed satisfaction of physiological needs and safety requirements. Such an assumption never has been true. In the 21st century, that assumption is no more valid. In the proposed paradigm "maximization of utility" (or something similar) becomes secondary to "self-preservation."

In the seminal DA book of [Luce and Raiffa (1957)] candidate strategies are selected under radical uncertainty on the basis of their payoffs and regrets under different scenarios, where payoffs are, say, profits of the post-contingency plans [Ibid.]. ("Regret" is the measure of opportunity lost.) The book considered five "single" and one "synthetic" criteria of comparison. RCO replaces them by six "synthetic" criteria. Again, because of lack of space, for detailed analysis of both old and new criteria the reader is referred to [Masch (2013)].

As mentioned in Section 3.1, in all old DA criteria the analysis assumes just one level of confidence in probabilities of scenario payoffs or regrets. This assumption evidently is too simplistic and too unworkable. So, in addition to (first) screening strategies instead of selecting and (second) to introducing new "synthetic" criteria, RCO (third, and very important) embeds these criteria into a framework of a fundamentally novel *strategic frontier*. Instead of comparing candidate strategies at some arbitrarily fixed *single* level of confidence, strategic frontier simultaneously compares them at *all* levels of confidence, from zero to 1.0. Strategic frontier from [Masch (2010): 456] for "index of pessimism-optimism" payoff criterion is demonstrated in Fig. 1.

The strategic frontier provides the following valuable information about the relative merits and faults of any strategy:

a) The composition of the subset of strategies that form the frontier.
b) The width of the interval supporting each frontier strategy.
c) The order of the frontier strategies on the optimism-pessimism spectrum.
d) The difference between the frontier strategy and each other strategy, which shows the possible impairment of results in choosing a non-frontier strategy.

The strategic frontier allows us to apply subjective estimates in a more prudent, convenient, and less demanding way. That is, the decision-maker does not need to specify in advance his level of confidence. Specifically, the frontier replaces hard-to-estimate "point" indices by index ranges. For instance, the current user of the "pessimism-optimism index" criterion may ask the question: "Which strategy, 0 or 2, is better if our guesstimated value of the index equals 0.8?" This means that we compare the strategies at precisely 0.8 probability of the "bad" outcome and 0.2 of the "good" outcome. Instead, when we use the strategic frontier, it is sufficient to say that Strategy 0 is preferable if the value of the index is no more than 0.768 and Strategy 2 otherwise.

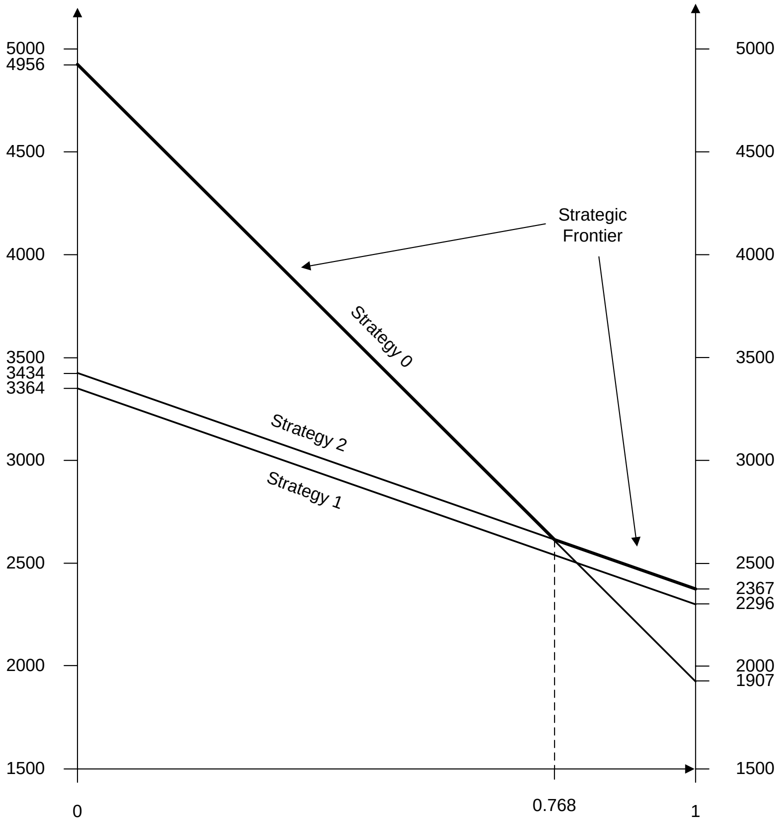

Figure 1. Strategic frontier for "Pessimism-Optimism Index" criterion

Again, the synthetic criteria and strategic frontiers do not select the strategy. They just shrink the list of the "finalist candidates" to a few reasonably most acceptable and safest candidates, leaving the final selection to the decision-makers. Similar to the risk-limiting constraints, *they do not find the best - they eliminate the worst, which is easier.* This stage is not connected with the optimization MMS models used to generate the candidate strategies. Therefore they help to compensate, to some extent, for possible flaws in these models. They are "from another field of research," as requested by Senge [Senge (1990)].

Throughout history, the general assumption was that there exists the correct and best decision. In particular, that is assumed in reductionist disciplines - economics and adjacent disciplines, such as DA, OR, and risk management. These disciplines are full of dangerous oversimplifications: "Its (i.e., reductionism's - Author) leading article of faith is that to every scientific problem is one and only one solution." [Ravetz (2009)]

In the maximization paradigm that assumption is further narrowed down: a strategy is the best if it leads to the greatest value of some quantity. But even without maximization, the very assumption of the existence of the best immediately restricts and damages the process of analysis and selection of a strategy. If there exists "the correct" decision, there also should exist "the correct" method of identifying that decision. This must be the one and the only one method. Different methods or different levels of confidence may give different answers. Therefore, both additional criteria and additional values of levels of confidence must be prohibited. The very concept of the existence of the correct decision thus ties the hands of decision-maker. Perilous strategies may not be detected and screened out – the very concept is dangerous. President Truman, who looked for a one-handed economist with one best solution, would be angry about RCO.

3.3 Stop the paradox – We'll get off

Applying strategic frontier allows a novel approach to such well-known paradoxes as "prisoner dilemma": they stop being paradoxical. If a prisoner is optimistic down to, say, 0.67 in the [0, 1] range of confidence levels, believing that his partner-in-crime will stand firm, he should remain silent, too. Otherwise, he should testify as asked.

Similarly, strategic frontier can provide quantitative evaluations for necessary confidence level in making financial decisions, thus complementing and refining observational results of behavioral economics.

3.4 Summary of RCO

Let us summarize the main principles of RCO as following:

1) Predictions do not matter – decisions (strategies) matter.
2) Utility of a strategy does not matter much – adaptability and robustness matter.
3) Likelihood of a scenario does not matter much – its riskiness matters.
4) Level of confidence at the strategy comparison stage does not matter – we can look simultaneously at all its levels, 0 to 1.
5) There is no correct strategy (or a correct method for finding it) – there is only the strategy most acceptable to decision-makers.
6) The main decision criterion is catastrophe avoidance, all others criteria are secondary.
7) There is a non-Bayesian (RCO) approach to constructing the posterior probabilities of scenarios.
8) Tasks should be simple, as well as the operations to perform them.
9) No extensive learning is necessary.
10) The operations should form a powerful interconnected ensemble.

As far as the author knows, RCO is unique to embody this new general philosophy and its tools. RCO goes as far as possible in dealing with decision-making under radical uncertainty. No other system seems to be even remotely close to RCO in that respect.

However imperfect, RCO-type systems still may be our best, if not the only possible, approaches to addressing complex societal and environmental problems that presently appear on the agenda. That also remains true for lower-level problems under radical uncertainty, including any problems of long-range planning or business research.

As far as the author knows, *for the first time since the creation in the 1940s of both computers and optimization models*, *RCO legitimizes the possibility of combined use at a high analytical level* of these two wonders of modern science and technology.

4. Concluding Remarks

The study underlying the RCO part of this paper is completely original. Since there were no previous studies, this paper does neither support nor contradict them. Time has come for a shift of paradigm; further delay is dangerous.

The proposed ASI and its RCO toolkit are full to the brim of fundamentally novel, both conceptually and technically, ideas, methods, models, and algorithms. Most of them are outlined in the cited author's publications. The cardinal distinction of the proposed mankind preservation paradigm from other theories is that it is supported by a toolkit that comprises several major innovations.

The significance of the paradigm shift to preservation of mankind is that it refutes the very foundation of economics and changes the role and capability of such adjacent disciplines as Operations Research and Decision Analysis, eliminating their reductionism and making them usable in real world. The paradigm also narrows the role of Complexity Theory.

ASI must be used just as a tool, rather as an agent. It also must have a limited goal and contain only simple operations. RCO meets these requirements. It performs simple operations that do not need ASI to have too much power, knowledge, and data. It provides a win of brain over brawn, avoiding the need in super-duper computers. Thus it eliminates or minimizes the enormous danger of ASI to mankind.

Thank God, this author is not an economist. He is an engineer-economist who builds tools. RCO is just a tool. But good tools create enormous opportunities.

As all new research, this study should be expanded in several directions, such as criteria and methods of collective decision-making. As everything else, this study is not perfect. In [Keynes (1921)] Keynes remarks: "There is much here, therefore, which is novel, and, being novel, unsifted, inaccurate, or deficient." No revolution is tidy. So let us be vigilant – and tolerant.

Shifting to the mankind preservation paradigm would require fundamental unpopular changes. They might be implemented only if we are frightened by a serious shock. Let us pray that the shock would not be too serious. Amen.

Acknowledgements: I would like to acknowledge extremely constructive suggestions made by editorial reviewers and by Professor Shih-Kung Lai. Dr. Yuriy V. Dashkov and Dr. Victor S. Smirnov provided useful comments. Needless to say, the author is fully responsible for the remaining faults.

References

[1] Arrow K.J., Hurwicz, L. (1972). "An Optimality Criterion for Decision Making under Ignorance". In: Carter C.F. and Ford J.L. (Eds.), *Uncertainty and Expectations in Economics*. Oxford, UK: Basil Blackwell.

[2] Bain & Company (2015). "Scenario and Contingency Planning". *Bain & Company Guide*. June 10, 2015.

[3] Baker D. (2016). "Economists keep it wrong because the media cover up their mistakes". *Real-World Economic Review Blog*, September 30, 2016.

[4] Bank of America – Merryll Lynch (2015). *Thematic Investing Robot Revolution – Global Robot & AI Primer*. New York: BofA.

[5] Barrat J. (2013). *Our Final Invention and the End of Human Era.* New York: Thomas Dunne Books.

[6] Bostrom N. (2013). "Existential Risk Prevention as Global Priority". *Global Policy*, 4(1): 15-31.

[7] Bostrom N. (2014). *Superintelligence: Paths, Dangers, Strategies*. Oxford UK: Oxford University Press.

[8] Bostrom N. (2016). *Superintelligence: Paths, Dangers, Strategies.* (Paperback edition). Oxford: Oxford University Press..

[9] Brockman T. (2017). "Tech leaders at Davos fret over effect of AI on jobs". *Financial Times*, January19, 2017.

[10] Chace C. (2015). *Surviving AI. The promises and peril of artificial intelligence.* San Mateo, CA: Three Cs Publishing.

[11] Cookson C. (2015). "Twelve ways the world could end". *FT Magazine*, February13, 2015.

[12] Dalai Lama & Brooks A. (2016). "Dalai Lama: Behind Our Anxiety, the Fear of Being Unneeded". *NY Times*, November 4, 2016.

[13] Domingos P. (2015). *The Master Algorithm. How the Quest for Ultimate Learning Machine Will Remake Old World*. New York: Basil Books.

[14] Johnson N. (2012). *Simply Complexity: A Clear Guide to Complexity Theory*. London, UK: Oneworld Publications.

[15] Keynes J.M. (1921). *A Treatise on Probability.* London UK: V. Macmillan.

[16] Kuhn T.S. (1962). *The structure of scientific revolutions*. Chicago: University of Chicago Press.

[17] Lai ShK, Han H. (2014). *Urban Complexity and Planning: Theories and Computer Simulations*. Ashgate, Surrey: Routledge.

[18] Lempert, R.J., Droves D.G., Popper S.W., and Bankes S.C. (2006). "A General, Analytic Method for Generating Robust Strategies and Narrative Scenarios". *Management Science,* 52 (4): 514-528.

[19] Lewis H.W. (1992). *Technological Risk.* New York: W. W. Norton.

[20] Lindner-Dutton L., Jordan M., Karwan M. (1994). "Beyond mean time to failure: Praxair's novel reliability modeling techniques keep gas flowing to customers". *OR/MS Today*, 21 (2): 30-33.

[21] Luce R.D., Raiffa H. (1957). *Games and Decisions*, pp.278-306. New York: Wiley.

[22] Makarov A.A., Melentyev L.A. (1973). *Methods of Studying and Optimizing the Energy Sector*, pp. 115-287. Nauka, Novosibirsk, USSR. (Макаров АА, Мелентьев ЛА (1973). *Методы исследования и оптимизации энергетического хозяйства*, pp.115-287. Наука, Новосибирск, СССР.)

[23] Masch V.A. (1999). *Computer aided risk management in multiple-parameter physical systems*. US Patent 5930762.

[24] Masch V.A. (2004). "Return to the natural process of DM leads to good strategies". *J. of Evolution Econ*, 14 (4): 431-462.

[25] Masch V.A. (2010). "An application of Risk-Constrained Optimization (RCO) to a problem of international trade". *Int. J. Oper Quant Manag.*, 16 (4): 415-465.

[26] Masch V.A. (2013). "Extensions of stochastic multiscenario models for long-range planning under uncertainty". *Environment, Systems, and Decisions*, 33 (1): 43-59.

[27] Masch V.A. (2015). "Shifting “The Dodo Paradigm”: To Be or Not to Be". *World Journal of Social Sciences*, 5 (3): 123-142.

[28] Maslow A.H. (1943). "A theory of human motivation". *Psychological Review*, 50 (4): 370-396.

[29] Phadnis Sh., Caplice C., and Sheffi Y. (2016). "How Scenario Planning Influences Strategic Decisions". *MIT Sloan Management Review*, May 27, 2016.

[30] Pollock J.L. (2004). "Plans and Decisions". *Theory & Decision*, 57 (2): 79-107.

[31] Pomerol J. Ch. (1997). "Artificial intelligence and human decision making". *European Journal of Operational Research*, 99 (1): 3-25.

[32] Ransbotham S. (2016). "Can Artificial Intelligence Replace Executive Decision Making?". *MIT Sloan Management Review*, June 28, 2016.

[33] Ravetz J. (2009). "Preface". In: Bammer G., Smithson M. (Eds.), *Uncertainty and risk — Multidisciplinary perspectives.* London UK: Earthscan.

[34] Rumsfeld D. (2011). *Known and unknown: a memoir*, xiii-xv. New York: Penguin Books.

[35] Senge P.M. (1990). *The fifth discipline*, 6. New York: Doubleday Currency.

[36] Waters R. (2016). "AI is ‘Next Big Thing’ to worry about". *Financial Times*, September 29, 2016.

[37] Williamson T. (2010). "Reclaiming the Imagination". *International Herald Tribune*, p.8, August 17, 2010.

[38] Yampolsky R.V. (2015). *Artificial Superintelligence: A Futuristic Approach.* Boca Raton, FL: CRC Press.