Nuclear Deterrence and the Metaphysics of Time

Jean-Pierre DUPUY
Professor, Stanford University

I. Metaphysics of the prophecy of doom

1. Bergson and the possible

The next time an atomic bomb will be dropped over a civilian population, breaking what has been called the "nuclear taboo", it is very likely that the event will be interpreted as the bursting forth of the possible into the realm of impossibility, as was the case with the destruction of the twin towers. From now on, one heard it said, even the worst horrors have become possible. Note that if something *becomes* possible, presumably this is because it was not possible before. And yet, common sense objects, if it actually occurs, this must be because it *was* possible all along. Common sense proves here once more to be a detestable guide.

In *The Two Sources of Morality and Religion* (1918), French philosopher Henri Bergson described the sensations he felt on 4 August 1914 on learning of Germany's declaration of war on France:

> Horror-struck though I was, and though I felt a war, even a victorious war, to be a catastrophe, I experienced what William James expresses, a feeling of admiration for the smoothness of the transition from the abstract to the concrete: who would have thought that so terrible an eventuality could make its entrance into reality with so little disturbance? The impression of this facility was predominant above all else.[1]

Yet this disturbing familiarity stood in sharp contrast to Bergson's feelings *before* the catastrophe. The prospect of war appeared to him and his friends "as *at once probable and impossible:* a complex and contradictory idea that lasted right down to the present day".[2]

---

[1] Henri Bergson, *The Two Sources of Morality and Religion,* trans. R. Ashley Audra and Cloudesley Brereton (Garden City, N.Y.: Doubleday, 1935), 159–60.
[2] Ibid., 159.

Some years later, Bergson managed very well to unravel this apparent contradiction in reflecting upon the nature of a work of art in an essay entitled "The Possible and the Real" (1930). "I believe in the end we shall consider it evident", Bergson wrote, "that the artist in executing his work *is creating the possible as well as the real*".[3] Why is it, then, he asked, that one might "hesitate to say the same thing for nature? Is not the world a work of art incomparably richer than that of the greatest artist?"[4] The hesitation to extend this idea to acts of destruction is greater still. And yet who has contemplated the images of 11 September and not been filled with a feeling of exaltation and dread that resembles what one feels in the presence of the sublime, in the sense that Burke and Kant gave to this word? Of the terrorists, who could hardly have failed to have sensations of the same kind, we may also say that they created the possible at the same time as they created the real. This was, as I say, the metaphysical view that most commentators spontaneously adopted.

The explanation of our inaction in the face of many looming disasters is to be found right here: anyone who wishes *to prevent* a catastrophe must believe in its possibility *before* it occurs. The paradox is that if one succeeds in actually preventing it, its non-realization keeps it firmly within the domain of the impossible, and efforts at prevention appear in retrospect to have been useless.[5]

3. Being a compatibilist

My starting point has been the age-old problem of the compatibility between determinism and free will in its modern version fleshed out by such philosophers as David K. Lewis and Robert Stalnaker[6].

---

[3] Henri Bergson, "The Possible and the Real", in *The Creative Mind: An Introduction to Metaphysics,* trans. Mabelle L. Andison (New York: Philosophical Library, 1946), 121.

[4] Ibid.

[5] A semi-comical illustration: the YK2 efforts at preventing a universal computer collapse at the (false) turn of the century, a collapse that didn't take place, were deemed by many afterwards to have been a waste of resources.

[6] David K. Lewis, *On the Plurality of Worlds*, Oxford: Blackwell Publishers, 1986; Robert Stalnaker, *Ifs: Conditionals, Belief, Decision, Chance, and Time*, Dordrecht: D. Reidel, 1981.
  As far as modalities are concerned let me recall that, given an adequate definition of a possible world, possible means that which is true in at least one possible world; necessary that which is true in all possible worlds; impossible that which is untrue in all possible worlds; and contingent that which is possible without being necessary.

Lewis calls "soft determinism" "the doctrine that sometimes one freely does what one is predetermined to do; and that in such a case one is able to act otherwise though past history and the laws of nature determine that one will not act otherwise". He then defines compatibilism as "the doctrine that soft determinism may be true".[7]

Let us call C the state of the world at a time t_1. We have:

A1:   C was the case at t_1

Consider a subject S whose action x at $t_2 > t_1$ is determined by the laws that govern his world according to:

A2:   If C was the case at t_1, then S does x at t_2

From A1 and A2 we derive by modus ponens:

A3:   S does x at t_2

Can x be a free although predetermined act? To defend soft determinism, it is always useful to start from the argument(s) put forward by those who deny it. The so-called "incompatibilist" thesis uses an operator ❑ which, applied to a proposition p, asserts that p is true in all possible worlds: it is necessary. More specific to our problem, we will call ❑^S_t the operator of necessity such that:
❑^S_t (p) means: p is true and S is not free at t to perform an act such that, if he were to perform it, p would be false.

The incompatibilist argument can be written as follows:

N1:   ❑$^S_{t2}$ (C was the case at t_1)
N2:   ❑$^S_{t2}$ (If C was the case at t_1, then S does x at t_2)
Thus, by modus ponens[8],
N3:   ❑$^S_{t2}$ (S does x at t_2)

N1 expresses the principle of the fixity of the past. N2 says that the laws that determine the subject's actions remain the same in all possible worlds. The conclusion N3 states

---

[7] David K. Lewis, "Are We Free to Break the Laws?", *Theoria*, 47 (1981), p. 112.
[8] Whether modus ponens remains valid under the operator of necessity could be questioned.

that S does actually do x at t_2, but he does not act freely since it is not in his power to act otherwise.

Can this argument be refuted? Depending on the nature of the problem, there are two possibilities, neither of which has greater *a priori* legitimacy than the other.

a) We could accept N1, in which case we would have to reject N2. The past is fixed, and the subject, supposedly able to act otherwise, has the power to invalidate the fixity of the temporal chain which links C to x. The nature of this power must be made very clear. As Lewis puts it, we must distinguish between two versions:

Strong version: "I am able to break a law".
Weak version: "I am able to do something such that, if I did it, a law would be broken".[9]

Obviously, there is no way that *in our world* the subject could act so that the link between C and x would be violated: this would be contrary to hypothesis A2, which indeed remains valid. The strong version is eliminated but not the weak one. To paraphrase Lewis, the way in which I was determined not to do anything other than x "was not the sort of way that counts as inability".[10] The power that this sort of ability represents is called "counterfactual".

b) Conversely, we could accept N2, in which case we would have to reject N1. This time the temporal chain A2 is held to be fixed (that is, true in all possible worlds). To maintain that the agent's action, x, is free although determined by the past and the laws that govern the world, we have to grant the agent a power to invalidate the past. This power obviously cannot be causal. Here too we must distinguish between:

a *strong version*: "I am able to change the past," which is "utterly incredible", to use Lewis's terms, and a *weak one*: "I am able to do something such that, if I did it, the past would have been different from what it was in the actual world".

---

[9] "Are We Free to Break the Laws?", loc. cit., p. 113.
[10] Ibid., p. 112.

The Calvinist theologian and analytic metaphysician Alvin Plantinga, who defends the weak version, has logically dubbed "counterfactual power over the past" this kind of ability.[11]

Although, as I said, the two ways of grounding compatibilism have an a priori equal legitimacy, contemporary philosophers such as David K. Lewis or Robert Stalnaker, probably because of their respective stints in the domain of rational choice theory, have focused almost exclusively on the former, which preserves the fixity of the past. I have explored thoroughly the second approach and been able to show that it formalizes elegantly the properties we have discovered as characterizing the prophecy of doom.

The first thing to be noted is that there exist situations in which the counterfactual power an agent possesses over the past causally prohibits him from acting in a certain way.[12] Let's consider a paradigmatic illustration which has been the object of numerous cogitations from Hobbes onward: the promise case.[13] At t_1 Mary asks Peter to lend her $1,000 and she promises to pay off her debt at $t_2 > t_1$. We are in a state of nature à la Hobbes: there are no state institutions, no judicial system, no rule of law. The agents are only guided by their self-interest. If the loan could take place, it would be mutually beneficial.

In the temporality that preserves the fixity of the past, it is immediate that the loan is impossible. Reasoning by backward induction we realize along with Peter that Mary at t_2 will break her promise. Peter would be a fool to lend her anything.

In the temporality that maintains N2, that is a necessary link between past conditions and future action, at the cost of doing away with the fixity of the past, things work very differently. Let's say Peter is an omniscient predictor capable of anticipating Mary's actions in all possible worlds. If Mary held her promise at t_2 Peter would anticipate it and the mutually beneficial loan would take place. On the other hand, if she were to renege on her promise, Peter would anticipate it as well and he wouldn't lend her the money. We see here in action the counterfactual power that Mary has on her past via her action. However, if the loan doesn't exist, Mary is not in a position to renege on her promise to pay off her debt. Hence a contradiction which

---

[11] Alvin Plantinga, "On Ockham's Way Out", *Faith and Philosophy*, 3, 1986.

[12] This paradox is akin to the so-called "grandfather paradox" that appears to be a consequence of the assumption of time travel. If I could travel to the past and kill my grandfather "I" couldn't be. The grandfather paradox relies unnecessarily on causal connections though, which is not the case of the implications of the counterfactual power over the past.

[13] Also known in game theory as the assurance game.

is immediately solved by the conclusion that Mary won't renege on her promise if the loan takes place. The loan will indeed take place to the mutual benefit of Peter and Mary.

This example illustrates that in the temporality we are examining it is not true that any future goes, since "it is not the future if you stop it".[14] The future must be such that the past that it counterfactually determines doesn't causally prevent its occurrence. In other words, the future, far from being the outcome of the laws of nature applied to determinate initial conditions (prediction) or something that we create according to our will ("prospective"[15]), is the solution (one of the solutions) to an equation in which the unknown x—he future action—appears on both sides of the equation in the following form:

$$x = F[x],$$

as if it were determining itself. According to the received terminology, we will say that the future appears as the *fixed point* of a certain operator F. The latter expresses the causal consequences of a past that is itself determined counterfactually by the future x. This can be represented graphically as follows:

Counterfactual

Past

Future

Causal

---
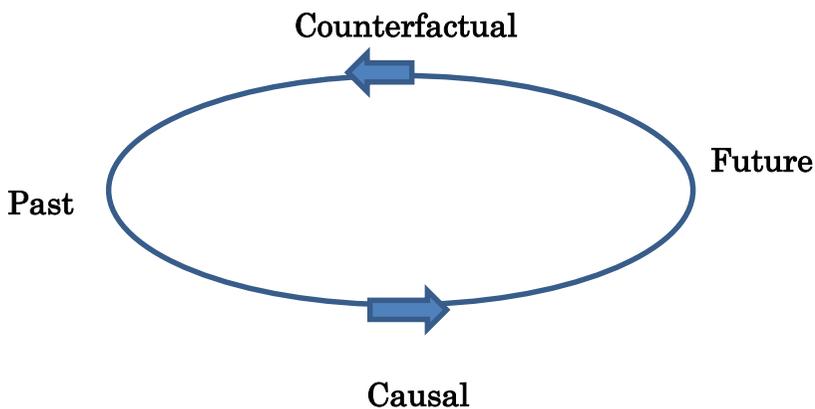
[14] Quote from Philip K. Dick's tale, "Minority report", a beautiful and profound illustration of some of the ideas presented here.

[15] Reference to the method known in France as *Prospective*, elsewhere as the Scenario method, or, more vaguely, "futurology", invented by the French philosophers Gaston Berger and Bertrand de Jouvenel at the end of the 1950's. One of its current proponents, Michel Godet, wrote: "All who claim to foretell or forecast the future are inevitably liars, for the future is not written anywhere—it is still to be built". (Michel Godet, "Creating the future: the use and misuse of scenarios", *Long Range Planning*, 29, 2, 1996.)

In this conception of time, the future is fixed, that is necessary, since it is linked to the past by N2, a proposition that states that this link is true in all possible worlds. However, this is only true once the past is determined, which presupposes that the future itself is determined. In other terms the future is necessary—it has always been necessary—but only once it has become actualized. This is the essential trait we have learned to ascribe to the metaphysics of the prophecy of doom.

4. On the multiplicity of metaphysics and the choice of the most pertinent

The indeterminacy of the past as long as action has not been performed along with the necessity of the future once action is taken serve to define a metaphysics of temporality which I have dubbed "Projected time". In what follows, in order to prepare the ground for my analysis of nuclear deterrence, I will introduce another metaphysics, which I name "Occurring time", the one that supports all strategic reasoning, be it carried out by an economist, a game theorist, a planner, an engineer, a designer or a military strategist. It corresponds to a very distinct conception of free will for which the agent's actions are driven by a set of beliefs and desires rather than "pushed" by a determinism. Named the belief-desire model, its most familiar graphic representation today is the decision tree. At every node of the tree an agent has the choice between several possible future options. When he chooses among them he holds the past as fixed, that is counterfactually independent of his choosing. Fixed past, open future, the metaphysics of occurring time is obviously in sharp contrast with that of projected time.

If metaphysics is the branch of philosophy that explores the fundamental nature of reality, according to a received definition, the question arises, how can we account for the plurality of metaphysics?

In the 4th Century BCE, a member of the Megarian School named Diodorus Kronos proposed an axiomatic, that is a set of propositions held to be self-evidently true, designed to show that the actual is the only possible and that the future is already determined.

The three axioms are:

1) Every true proposition about the past is necessary.
2) The impossible does not logically follow from the possible.
3) There is a possible which neither is presently true nor will be so.

Diodorus demonstrated that they are incompatible. One of them at least must go. Axiom 3 seems self-evident to most people today. However, if they hold like Diodorus that 1 and 2 too are self-evident, then they must deny 3. That is, they must hold that an event that happens neither in the present nor in the future is an impossible event.

One of the greatest French philosophers of the 20th century, Jules Vuillemin has written a history of western metaphysics on the simple basis of which axiom or axioms various philosophers decided to drop. This makes a fascinating story.[16]

The multiplicity of metaphysics finds its origin in Diodorus's theorem of incompatibility. A comparison comes to mind with the history of geometry. Once it was demonstrated fairly late in the history of mathematics that Euclides's fifth axiom, the so-called parallel axiom, couldn't be derived from the first four, it became conceivable to imagine a geometry in which this axiom wouldn't hold. The concept of a Riemannian manifold followed. And it proved extremely *useful*, as is well known, to Albert Einstein who was in the process of elaborating his theory of general relativity. French mathematician Henri Poincaré then asked: "Is Euclidian geometry true? This question is deprived of meaning altogether….A geometry cannot be truer than another; it is enough for it to be more convenient".[17] Likewise, let's not ask whether projected time is truer than occurring time, but if it is or not more useful than the latter. It all depends on the kind of problem we are facing. Let's note first that projected time and occurring time are two ways of skirting Diodorus's aporia. The former denies axioms 1 and 3, the latter endorses them both and therefore denies 2.[18] In my work on catastrophes[19]—including a nuclear conflict—I've shown that projected time avoids many paradoxes which occurring time, i.e., strategic thinking, comes up against when it comes to conceptualizing the temporality that separates us from a looming disaster the date of which is unknown. The second part of this paper will illustrate this point. Projected time defines an attitude that is neither complacency or voluntarism nor fatalism. Complacency stresses that the catastrophe although possible is not inevitable: the future is open. Fatalism makes it inevitable. By granting the agent

---

[16] The English version is more complete and a few errors have been corrected. See Jules Vuillemin, *Necessity or Contingency. The Master Argument*, CSLI Publications, Stanford University, 1996.

[17] Henri Poincaré, *La Science et l'hypothèse*, Paris, Flammarion, 1917.

[18] Hence the paradoxes of backward induction. See J.-P. Dupuy, "Philosophical Foundations of a New Concept of Equilibrium in the Social Sciences: Projected Equilibrium", *Philosophical Studies*, 100, 2000, p. 323–345.

[19] Jean-Pierre Dupuy, *Pour un catastrophisme éclairé*, Paris, Seuil, 2002.

the counterfactual power to act upon the past conditions that determine him, projected time helps him navigate between the devil of catastrophism and the deep blue sea of dumb optimism.

For reasons already mentioned our Zeitgeist leans toward the latter. It is worth then reminding that the experience of projected time has accompanied humankind since time immemorial. It is intimately linked to the religious apprehension of the world. In all traditional societies, there are people called prophets (*nabis* in Ancient Israel) whose function is to interpret and convey the divinity's will. The prophets of the Bible, for instance, were extraordinary men, often great eccentrics and they did not go unnoticed by their neighbors. The influence their prophecies had on the world around them and on the course of events had purely human and social causes; but it was due also to the fact that those who heard them believed that the word of the prophet was the word of the Lord and that this word, which came to the prophet directly, from on high, had the power to bring about the very thing that it announced. We would say today that the word of the prophet had a *performative* power: in saying things, he brought them into being. However, the prophet was well aware of this. One might be tempted to conclude that the prophet had the power to which political revolutionaries aspire: he spoke so that things might change in the direction that he wished to impress upon them. But this would be to overlook the fatalistic aspect of prophecy, which reads out the names of all those things that will come to pass, just as they are written down on the great scroll of history, immutably, ineluctably.

Revolutionary prophecy, particularly in the form it came to acquire in Marxist doctrine, has preserved the highly paradoxical mixture of fatalism and voluntarism that characterizes biblical prophecy. German philosopher Hans Jonas said of dialectical materialism that it was "a most peculiar mixture of colossal responsibility for the future with deterministic release from responsibility".[20]

The metaphysics of projected time enables us to extend the notion of prophecy to our secular age and substitute for the obscure dialectic between voluntarism and fatalism a rigorous and non-paradoxical third way that is neither one nor the other. For the modern prophet, especially the prophet of doom, it is necessary to seek the fixed point of the loop between past and future, at which an expectation (on the part of the past with regard to the future) and a causal production (of the future by the past) coincide. The prophet, knowing that his public announcements are going to have a causal impact on the world, must take account of this fact if he wants the future to

---

[20] Hans Jonas, *The Imperative of Responsibility. In Search of an Ethics for the Technological Age*, University of Chicago Press, 1985, p. 113–14.

confirm what he foretold. The future is an x, that is a solution to an equation which says that the reactions to the past anticipations of x causally bring about x.[21]

In this sense, prophets are legion in our modern democratic societies, founded on science and technology. The experience of projected time is facilitated, encouraged, organized, not to say imposed by numerous features of our institutions. All around us, more or less authoritative voices are heard that proclaim what the more or less near future will be: the next day's traffic on the freeway, the result of the upcoming elections, the rates of inflation and growth for the coming year, the changing levels of greenhouse gases, etc. The *futurists* and sundry other prognosticators know full well, as do we, that this future they announce to us as if it were written in the stars is a future of our own making, even if it is in reaction to these very announcements. We do not rebel in general against what could pass for a metaphysical scandal. We have then the experience of projected time.

II. Metaphysics of Nuclear Deterrence

1. Caveat

I am writing these lines in April 2022, at a time when the prospect of a nuclear war between the US and Russia is deemed by many observers stronger than it has ever been. In a book published in 2015 and titled *My Journey at the Nuclear Brink*,[22] former Secretary of Defense William Perry wrote: "Today, the danger of some sort of a nuclear catastrophe is greater than it was during the Cold War, and most people are blissfully unaware of this danger".

The American film maker Errol Morris, in his movie *The Fog of War*,[23] asks Robert McNamara, the former Secretary of Defense of President Kennedy, what he thinks protected humanity from extinction during the Cold War, when the United States and the Soviet Union permanently threatened each other with mutual annihilation. Deterrence? Not at all, McNamara replies: "We lucked out". Twenty-

---

[21] Not any future goes. The prophet Jonah knew that if he prophesied the fall of Niniveh as God had asked him to do the Ninivites would repent and God would forgive them. He preferred to run away from God's gaze.
[22] Stanford University Press, 2015.
[23] Errol Morris, *The Fog of War. Eleven Lessons from the Life of Robert S. McNamara*, Sony Classics, 2003.

five or thirty times during this period, he notes, mankind came within an inch of apocalypse. I will show that this response is self-contradictory. All those "near-hits" may have been the necessary condition for nuclear deterrence (ND) to work. To the extent that ND can be at times efficient, my objective is to show that everything occurs then *as if* the protagonists had immersed themselves in the peculiar logic of projected time.

Let me hasten to add that this is in no way meant to be a justification of nuclear deterrence in the MAD form. My conviction is that the latter is morally abhorrent. But there is a logic to it that can be discerned quite clearly.

My strategy will be as follows. In a first phase, I will expound the broad lines of the intellectual history of ND, following Steven P. Lee's excellent book, *Morality, Prudence, and Nuclear Weapons*.[24] There is not one argument put forward by the protagonists in that discussion that has not been questioned, disputed, challenged, refuted by some, defended by others, in an unending quest for reason and justice. I won't enter in those controversies and will be content with just reporting what the dominant arguments have been. My critical standpoint resides elsewhere, and I will expound it in a second moment. It consists in showing that confusions spoil the debate, and they stem from the fact that a good number of arguments belong to strategic reasoning and find their place within the metaphysics of occurring time while others, in general more recent, pertain to projected time and presuppose the renunciation of strategy. Two incompatible metaphysics of time clash invisibly.

2. A Brief History of Nuclear Deterrence Theory

For more than four decades during the Cold War, the discussion of "mutual assured destruction" (MAD) assigned a major role to the notion of *deterrent intention*, on both the strategic and the moral level. And yet the language of intention can be shown to constitute the principal obstacle to understanding the logic of deterrence.

2.1. In June 2000, meeting with Vladimir Putin in Moscow, Bill Clinton made an amazing statement that was echoed almost seven years later by Secretary of State Condoleezza Rice, speaking once again to the Russians. "The antiballistic shield that we are going to build in Europe, they explained in substance, is only meant to defend us against attacks from rogue states and terrorist groups. *Therefore be assured:* even

---

[24] Cambridge University Press, 1996.

if we were to take the initiative of attacking you in a first nuclear strike, you could easily get through the shield and annihilate our country, the United States of America".

Plainly the new world order created by the collapse of Soviet power in no way made the logic of deterrence any less insane. This logic requires that each nation exposes its own population to certain destruction by the other's reprisals. Security becomes the daughter of terror. For if either nation were to take steps to protect itself, the other might believe that its adversary considers itself to be invulnerable, and so, in order to prevent a first strike, hastens to launch this strike itself. Before being a doctrine, MAD is a situation, in which nations are at once vulnerable and invulnerable: vulnerable because they can die from attack by another nation; invulnerable because they will not die before having killed their attacker—something they will always be capable of doing, thanks to a second-strike capacity, no matter how powerful the attack that will have brought them to their knees. Clearly the confrontation between the US and North Korea doesn't meet this definition, nor would a face-off between Israel and a nuclearized Iran.

2.2. Throughout the Cold War, two *a priori* arguments were made that seemed to show that nuclear deterrence in the form of MAD could not be effective. The first argument has to do with the non-credible character of the deterrent threat under such circumstances: if the party threatening a simultaneously lethal and suicidal response to aggression that endangers its "vital interests" is assumed to be at least minimally rational, calling its bluff—say, by means of a first strike that destroys a part of its territory—ensures that it will not carry out its threat. The very purpose of this regime, after all, is to issue a guarantee of mutual destruction in the event that either party upsets the balance of terror. What chief of state having in the aftermath of a first strike only a devastated nation to defend would run the risk, by launching a retaliatory strike out of a desire for vengeance, of putting an end to the human race while committing suicide in the process? In a world of sovereign states endowed with the minimal degree of rationality that Hobbes granted to the inhabitants of the state of nature, namely the instinct of self-preservation, the nuclear threat has no credibility whatever.

The credibility question occupies the great majority of the debates about ND. Many experts conclude in particular that it is folly to make extreme threats that one is not sure one will deliver on. If your enemy calls your bluff, either you deliver and you risk what Clausewitz called the escalation to the extreme, that is mutual annihilation, or you cave in and your credibility is down for the future. One of the best ways to keep your credibility intact is to multiply the occasions in which you show the world

that your threats are not empty words: you do deliver and build a reputation of toughness.

2.3. The last remark leads to the second argument present in the literature that likewise points to the incoherence of the MAD strategic doctrine. Its premise is that, in Leon Wieseltier's words, "Nuclear deterrence is the only public arrangement that is a total failure if it is successful only 99.9 percent of the time". To be effective, ND must be absolutely effective. Not even a single failure can be allowed, since the first bomb to be dropped would already be one too many. But in that case never will the adversaries be in a position to test the other's resolve to deliver on its threats. Perfect nuclear deterrence is said to be self-defeating or "self-stultifying"[25] since it undermines the very conditions that would make it efficient.

2.4. Nuclear deterrence doesn't work because the threat to retaliate is not credible. It doesn't work also because if it did, that assumption would lead to a contradiction. Those two reasons add up to the conclusion that the nuclear opponents are unable to deter one another. And yet, the Cold War, also known as Nuclear Peace, seemed to demonstrate the opposite, in spite of a significant number of "near-hits". An explanation had to be found.

Belatedly, it came to be understood that in order for deterrence to have a chance of succeeding, it was absolutely necessary to abandon the notion of deterrent *intention*. In principle, the mere *existence* of two deadly arsenals pointed at each other, without the least threat of their use being made or even implied, is enough to keep the warheads locked away in their silos. As two major philosophers put it, "The existence of a nuclear retaliatory capability suffices for deterrence, regardless of a nation's will, intentions, or pronouncements about nuclear weapons use". [Gregory Kavka[26]]; or:

---

[25] Expression used by Gregory Kavka (*Moral Paradoxes of Nuclear Deterrence*. Cambridge. Cambridge University Press, 1987) apropos of a different but kindred argument, which has for a long time been the ethical justification of the French nuclear doctrine known as deterrence "from the weak against the strong". The claim is that the deterrent intention to inflict "incommensurable" harm to the other party if it attacks you, is not a genuine intention, since your true intention is to not have to carry it out. As the tortuous expression goes, "We form the deterrent intention in order to make it so that the conditions that would lead to its being acted upon are not realized".

[26] Op. cit., p.48.

"It is our military capacities that matter, not our intentions or incentives or declarations". [David K. Lewis[27]].

Initially due to McGeorge Bundy, this doctrine has received the name of existential deterrence. The insistence on the causal power of the mere existence of nuclear weapons is a way to downplay the importance of strategy, intentions, plans, all major constituents of military thinking. If there is no need to threaten anyone it is because the weapons themselves, due to their incommensurate power, speak for us. If rationality plays a role here it is "the kind of rationality in which the agent contemplates the abyss and simply decides never to get too close to the edge".[28]

3. Fate and the Tiger

How exactly does existential deterrence work? Who or what deters whom? It is significant that the explanations provided by the best theoreticians rely on a non-human actor. We will consider two of them.

Let's start with David K. Lewis and the following quote:

> "*You don't tangle with tigers – it's that simple*". [29]

The implication is that the game is no longer played between two adversaries. It takes on an altogether different form. Let's admit we are convinced that neither is in a position to deter the other in a credible way. *However, both want and need to be deterred*. The way out of this impasse is brilliant. It is a matter of creating jointly a fictitious entity that will deter both at the same time. The game is now played between one actor, humankind, whose survival is at stake, and its double, namely its own violence exteriorized in the form of a wild animal. The fictitious and fictional "tiger" we'd better not tangle with is nothing other than the violence that is in us but that we project outside of us. It is as if we were threatened *but also protected* by an exceedingly dangerous entity, external to us, whose intentions toward us are not evil, but whose power of destruction is infinitely superior to all the earthquakes or tsunamis that Nature has in store for us.

---

[27] David K. Lewis, "Finite Counterforce" in Henry Shue (ed.), *Nuclear Deterrence and Moral Restraint*, Cambridge, Cambridge University Press, 1989, p. 67.
[28] Steven P. Lee, *Morality, Prudence, and Nuclear Weapons*, op. cit., p. 248.
[29] David K. Lewis, loc. cit., p. 68.

According to French anthropologist René Girard,[30] the sacred stems from a similar mechanism of self-externalization of human violence. It used to be said of the atomic bomb, especially during the years of the Cold War, that it was our new sacrament. Very few among those who were given to saying this sort of thing saw it as anything more than a vague metaphor. But in fact there is a very precise sense in which the bomb and the sacred can both be said to *contain* violence in the twofold sense of the verb "to contain": to have within oneself and to keep in check. The sacred holds back violence through violent means, the original one being sacrifice. In the same way, throughout the Cold War, it was as though the bomb had protected us from the bomb. The very existence of nuclear weapons, it would appear, had prevented a nuclear holocaust.

One must not come too near to the sacred, for fear of causing violence to be unleashed; nor should one stand too far away from it, however, for it protects us from violence. Likewise, we cannot risk coming too close to the nuclear tiger, lest it should devour us; nor can we risk standing too far away, lest we forget the danger it represents. For deterrence to work it's all about finding the right distance from the big cat.

The second quote is from Bernard Brodie:

It is a curious paradox of our time that one of the foremost factors making deterrence really work and work well is the lurking fear that in some massive confrontation crisis it may fail. Under these circumstances one does not tempt fate.[31]

Fate has replaced the tiger, but both images have in common that they place the deterrent in something else than human agency. We will return in the conclusion to a salient feature of this extraordinary quote, namely that it conjoins contingence (eventuality of failure) and necessity (fate), but we can pause at this stage and consider the following claim: the metaphysics of nuclear deterrence in its existential form is projected time. The renunciation of strategic thinking, the recourse to fate and the minimization of human agency, are all features that point in that direction.

4. Nuclear Deterrence in Projected Time

---

[30] *Violence and the Sacred*, New York: Continuum, 2005; origin. 1972. See also Jean-Pierre Dupuy, « René Girard. Desire, Violence, and Religion », *Inference*, vol. 2, issue 2.
[31] Bernard Brodie, *War and Politics*, New York, Macmillan, 1973, p. 430–31.

Let us admit for the sake of the discussion that the threat that underlies nuclear deterrence in its MAD form is not credible. The reasoning that supports this conclusion is strategic, and it is grounded in the metaphysics of occurring time. We reason by backward induction and we posit that if the bluff of the menacing party is called, the latter will prefer to yield rather than being annihilated. The would-be attacker won't be deterred. The question is, doesn't projected time provide an alternative ground that would account for the efficiency of nuclear deterrence?

Given what we have learned in the first part of this paper, we can easily reach a conclusion, and it is negative. In projected time, nuclear deterrence doesn't fare better than in occurring time, but it is for entirely different reasons. The reasoning goes as follows:

1. If deterrence works the escalation to the extreme, that is, the realization of the MAD threat, doesn't take place.
2. If the escalation to the extreme doesn't take place, then it is impossible. [Negation of Diodorus' 3rd axiom.]
3. If it is impossible, then nuclear deterrence doesn't work.
4. We have shown that if nuclear deterrence works, then it doesn't work.
5. Therefore, nuclear deterrence doesn't work.

The core of this argument is of course proposition 2, which expresses the condition that in projected time the future is necessary: an event that happens neither in the present nor in the future is an impossible event.

This reasoning gives a solid foundation to the second argument put forward by the critics of MAD. The alleged "self-defeatingness of a successful deterrence" appears to be a tortuous way of expressing a straightforward *reductio ad absurdum* (propositions 4 & 5).

The detour via the metaphysics of projected time proves unsuccessful. There is a way however to render it successful and it consists in taking seriously the dialectic between contingency and necessity that is suggested in Bernard Brodie's quote. Meanwhile we are going to realize that projected time is capable of solving the paradoxes of nuclear deterrence much more easily than strategic reasoning.

5. Nuclear Deterrence and the Indeterminacy of the Future

The suggestion that the manipulation of uncertainty can be a strategic tool that helps solve the credibility problem is not new. The conviction that if the agents are minimally rational they won't deliver on their threat to retaliate and launch the escalation to the extreme has led to the idea that it can be rational to pretend that one is irrational. It was first conceptualized by economist and game theorist Thomas Schelling in his landmark *The Strategy of Conflict*[32] but made famous under the moniker "Madman Theory" by Richard Nixon during the Vietnam War. The following quote is eloquent. Nixon to his chief of staff H. R. Haldeman:

> I call it the Madman Theory, Bob. I want the North Vietnamese to believe I've reached the point where I might do anything to stop the war. We'll just slip the word to them that, "for God's sake, you know Nixon is obsessed about communism. We can't restrain him when he's angry—and he has his hand on the nuclear button" and Ho Chi Minh himself will be in Paris in two days begging for peace.[33]

The problem of course remains: what happens if the other side calls your bluff? In the face-off between Donald Trump and Kim Jong Un that scared the world during the summer of 2017, the question was, who is pretending to be mad and who is not pretending, because he is really mad?

However, in Brodie's quote, we are no longer talking strategy. The twofold reference to fate and the eventuality of failure takes us to a completely different world. The notion that it requires an accident for fate to come to pass is as old as the oldest myths of the planet. Think of Oedipus: it was proclaimed by the Oracle that he would commit parricide and incest. What precipitated the realization of this prophecy was a random encounter with a disgruntled old man who was barring his way. The merger of fate and accident is a common theme of many religious traditions. Rome had a goddess who represented at the same time luck (good or bad) and fate—or, to use the language of modalities, contingency and necessity. Her name was *Fortuna*.

Once again, the metaphysics of projected time offers a framework capable of giving a precise and formalized rendering of these intuitions. The key is a concept I haven't yet introduced: the uncertainty of the future in projected time.

---

[32] Harvard University Press, 1960.
[33] Haldeman, H. R., *The Ends of Power*, Times Books, 1978, p. 122.

The uncertainty of the future in *occurring* time is approached with the usual tools. In the Madman theory, the agent confronting some crazy behavior asks himself whether the folly is feigned, in which case the Madman will likely yield if his bluff is called, or whether he is *really* mad, in which case he may launch the escalation to mutual destruction if attacked. The agent ascribes a subjective probability epsilon, hopefully very small, to the latter possibility and the complement to 1 to the former. The way he comes to a decision is left to him—he may deem the Savage criterion of the maximization of expected utility senseless if the magnitudes are extreme: exceedingly large for the consequences, very small for epsilon—, but one thing is assured: the two options make up a partition of the set of possibles, that is a *disjunction* without overlap.

In projected time, uncertainty takes on a radically different form. There are no alternative possible futures, since the future is necessary. What replaces the disjunction is a *superposition* of states. Both the escalation to the extreme and its negation are part of the fixed future. It is because the former figures in the future that deterrence has a chance to work. It is because the latter figures in the future that the adversaries are not bound to destroy each other. Only the future when it comes to pass will tell.

The signature feature that distinguishes the two forms of uncertainty is the following: in occurring time, epsilon, the probability of the catastrophic scenario, can be equated to zero without that leading to a contradiction. If we continue to call epsilon the relative weight that this scenario has in the superposition of states, then it is essential that epsilon remain strictly positive. Were it to become naught, the escalation to the extreme would become impossible, for the reasons already adduced, and deterrence would fail. Superposition of states and strict positivity of epsilon are kindred concepts.

It's time to conclude. The nuclear deterrent that really works has been, and still is potentially, *the indeterminacy of the future in a conception of time that makes the future necessary*. It is indeed possible to provide rational foundations to the efficiency of nuclear deterrence. And that conclusion is horrendous.

References

Bergson, Henri (1935) *The Two Sources of Morality and Religion.* Trans. R. Ashley Audra and Cloudesley Brereton (Garden City, N.Y.: Doubleday).

105

Bergson, Henri (1946) The Possible and the Real, in *The Creative Mind: An Introduction to Metaphysics*. Trans. Mabelle L. Andison (New York: Philosophical Library).

Brodie, Bernard (1973) *War and Politics*. (New York: Macmillan).

Dupuy, Jean-Pierre (2000) Philosophical Foundations of a New Concept of Equilibrium in the Social Sciences: Projected Equilibrium. *Philosophical Studies*, 100, 323–345.

Dupuy, Jean-Pierre (2002) *Pour un catastrophisme éclairé*. (Paris: Seuil)

Dupuy, Jean-Pierre (2016) « René Girard. Desire, Violence, and Religion ». *Inference*, vol. 2, issue 2.

Girard, René (2005; origin. 1972) *Violence and the Sacred*. (New York: Continuum).

Godet, Michel (1996) Creating the future: the use and misuse of scenarios. *Long Range Planning*, 29.2.

Haldeman, Harry Robbins (1978) *The Ends of Power*. (New York: Times Books).

Jonas, Hans (1985) *The Imperative of Responsibility. In Search of an Ethics for the Technological Age*. (Chicago: University of Chicago Press).

Kavka, Gregory (1987) *Moral Paradoxes of Nuclear Deterrence*. (Cambridge: Cambridge University Press).

Lee, Steven P. (1996) *Morality, Prudence, and Nuclear Weapons*. (Cambridge: Cambridge University Press).

Lewis, David K. (1981) Are We Free to Break the Laws? *Theoria*, 47.3, 113–21.

Lewis, David K. (1986) *On the Plurality of Worlds*. (Oxford: Blackwell Publishers).

Lewis, David K.(1989) Finite Counterforce, in Henry Shue (ed.), *Nuclear Deterrence and Moral Restraint*. (Cambridge: Cambridge University Press).

Morris, Errol (2003) *The Fog of War. Eleven Lessons from the Life of Robert S. McNamara*. (Sony Classics).

Perry, William (2015) *My Journey at the Nuclear Brink.* (Redwood City: Stanford University Press).

Plantinga, Alvin (1986) On Ockham's Way Out. *Faith and Philosophy*, 3.3, 235–269.

Poincaré,Henri (1917) *La Science et l'hypothèse*. (Paris: Flammarion).

Schelling, Thomas (1960) *The Strategy of Conflict.* (Cambridge: Harvard University Press).

Stalnaker, Robert (1981) *Ifs: Conditionals, Belief, Decision, Chance, and Time*. (Dordrecht: D. Reidel).

Vuillemin, Jules (1996) *Necessity or Contingency. The Master Argument*. (Chicago: The University of Chicago Press).