Nuclear Deterrence and the Metaphysics of Time

Jean-Pierre Dupuy

“[T]hat […] is how the world moves: Not like an arrow, but a boomerang. […] The end is in the beginning.” Ralph Ellison

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

---

1 Ellison 1952, p. 6. Thanks very much to Armand Braun for suggesting this quote. Armand Braun is writing a biography of French philosopher Gaston Berger (1896-1960), the inventor of the approach to the future called “prospective.” One day Berger said: “From now on I’m going to devote myself fully to the study of time.” He died a few days later in an awful car accident.

2 “We can’t attack [North Korea],” said Robert Kelly, a professor of diplomacy at Pusan University in South Korea, on August 9th 2017: “We’re talking about a million people who are going to be killed. So let’s not do that.” The “can’t” obviously refers not to a physical or technological impossibility, but to a moral necessity.
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), Henri Bergson described the sensations he felt on August 4, 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. (Bergson 1935, p. 149)

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, which lasted right down to the present day” (ibid., p. 148; my emphasis).

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” (Bergson 1946, p. 109; my emphasis). 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?” (Ibid.) The hesitation to extend this idea to acts of destruction is greater still. And yet who has contemplated the images of September 11 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.

In the same text, Bergson reports a delightful conversation with a journalist who had come to interview him, during the Great War, on the subject of the future of literature. “How do you conceive, for example, the great dramatic work of tomorrow?” he was asked. “But,” Bergson objected, “the work of which you speak is not yet possible.” “But it must be, since it is to take place,” retorted the other (ibid., pp. 106-07). To this Bergson replied:

“No, it is not. I grant you, at most, that it will have been possible.” “What do you mean by that?” — “It’s quite simple. Let a man of talent or genius come forth, let him create a work: it will then be real, and by that very fact it becomes retrospectively or retroactively possible. It would not be possible, it would not have been so, if this man had not come upon the scene. That is why I tell you that it will have been possible today, but that it is not yet so.” “You’re not serious! You are surely not going to maintain that the future has an effect upon the present, that the present brings something into the past, that action works back over the course of time and imprints its mark afterwards?” — “That depends. That one can put reality into the past and thus work backwards in time is something I have never claimed. But that one can put the possible there, or rather that the possible may put itself there at any moment, is not to be doubted. As reality is created as something unforeseeable and new, its image is reflected behind it into the indefinite past; thus it finds that it has from all time been possible, but it is at this precise moment that it begins to have been always possible, and that is why I said that its possibility, which does not precede its reality, will have preceded it once the reality has appeared.” (Bergson 1946, p. 107; second emphasis mine)

Before 1907, the year when Picasso painted Les Demoiselles d’Avignon, that painting was not possible. More than a century
later, the fact that it has become real entails that it was indeed possible before 1907. The truth value of the proposition “The painting *Les Demoiselles d’Avignon* was possible in 1900” depends on the time at which it is enunciated. Hence the recourse to a grammatical tense that is not frequently used in English—the future perfect: “it *will have been possible.*”

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 nonrealization keeps it firmly within the domain of the impossible, and efforts at prevention appear in retrospect to have been useless.³

### 2. Günther Anders and the Quandary of the Prophet of Doom

German philosopher Günther Anders (1902-1992) was the most profound and radical thinker to have reflected on the major catastrophes of the twentieth century. He is less well known than two of his companions who studied with him under Heidegger: Hans Jonas, who was his friend, and Hannah Arendt, of whom he was the first husband. This is probably due to his intransigence and the fragmented character of his work. Rather than weighty systematic treatises, Anders preferred shorter pieces on current issues, sometimes written in the form of a parable. More than once, he will have told in his own way the story of the flood.

Noah was tired of playing the prophet of doom and of always foretelling a catastrophe that would not occur and that no one would take seriously. One day,

> he clothed himself in sackcloth and put ashes on his head. This act was only permitted to someone lamenting the loss of his dear child

---

³ 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.
or his wife. Clothed in the habit of truth, acting sorrowful, he went back to the city, intent on using to his advantage the curiosity, malignity and superstition of its people. Within a short time, he had gathered around him a small crowd, and the questions began to surface. He was asked if someone was dead and who the dead person was. Noah answered them that many were dead and, much to the amusement of those who were listening, that they themselves were dead. Asked when this catastrophe had taken place, he answered: tomorrow. Seizing this moment of attention and disarray, Noah stood up to his full height and began to speak: the day after tomorrow, the flood will be something that will have been. And when the flood will have been, all that is will never have existed.

When the flood will have carried away all that is, all that will have been, it will be too late to remember, for there will be no one left. So there will no longer be any difference between the dead and those who weep for them. If I have come before you, it is to reverse time, it is to weep today for tomorrow’s dead. The day after tomorrow, it will be too late. Upon this, he went back home, took his clothes off, removed the ashes covering his face, and went to his workshop. In the evening, a carpenter knocked on his door and said to him: let me help you build an ark, so that this may become false. Later, a roofer joined with them and said: it is raining over the mountains, let me help you, so that this may become false. (Anders 1972; quoted in Simonelli 2004, pp. 84-85; my emphases)

The whole quandary of the prophet of doom, as well as the ingenious way of getting out of his predicament, is inscribed in this magnificent parable. The prophet of doom is not heard because his word, even if it brings information, does not fit with the beliefs of those to whom it is addressed. It is often said that if we fail to act in the face of catastrophe it is because our knowledge is uncertain. Yet, even when we have all the clues at our disposal, we are unable to transform this information into belief.

Noah’s way out is brilliant. It consists in the staging of mourning for deaths that have not yet occurred; in this way, says Anders, time is “reversed,” since the effect (mourning) comes before the cause (the deaths). This is indeed unusual since it could
only make sense if the deaths to be mourned were inscribed in the future at a determinate date. What makes death bearable, for many of us, is precisely that we tend to liken an unknown end to an indeterminate end, and so to the absence of an end. “Whatever certainty there is in death,” 17th-century French moralist La Bruyère remarked, “is mitigated to some extent by that which is uncertain, by an indefiniteness in time that has something of the infinite about it” (Bruyère 1933, p. 398). By contrast Anders’s aim with this parable is to stress that the catastrophe is inevitable, or rather, will have always been inevitable once it occurs.

The prophecy of doom purports to be the antidote of Bergson’s metaphysics. Like any pharmakon, at the same time poison and remedy, the former retains some essential traits from the latter. In the same way that for Bergson the catastrophe is impossible before it occurs and starts having always been possible once it occurs, in the prophecy of doom the catastrophe is not necessary before it occurs and begins to have always been necessary once it occurs. In either case, it is the actualization of the event—the fact that it occurs—which retrospectively substitutes a modality for a previous one: possibility, in one case, necessity (or impossibility) in the other.

The paradox of the prophecy of doom is as follows. Making the perspective of catastrophe credible requires one to increase the ontological force of its inscription in the future. The foretold suffering and deaths will inevitably occur. But if this task is too well carried out, one will have lost sight of its purpose, which is precisely to raise people’s awareness and spur them to action so that the catastrophe may not occur—“let me help you build an ark, so that this may become false.”

What prevents the metaphysics inherent in Anders’s parable from boiling down to an old-fashioned classical fatalism is its future-perfect structure. Sophisticated though the latter may sound it is of common usage among us. Consider the following statement: “If Mr. Trump’s presidency ends in humiliation, future
generations may well conclude that it was bound to fail all along.”

This statement is true if either Mr. Trump’s presidency doesn’t end in humiliation or it was bound to fail all along. The statement doesn’t say that Mr. Trump’s presidency is necessarily doomed; only that if it turns out to be a failure, it will have been destined to be such all along. Necessity here is retrospective.

3. Being a Compatibilist

Although very suggestive, Anders’s parable remains a poetic rendering of abstract ideas, that is, an allegory. In my own work, I’ve tried to give it a formal interpretation in the framework of analytical metaphysics (and analytical theology).

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. 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

---


5 Another significant example is literary, drawn from Henry de Montherlant’s La guerre civile (1965). Montherlant imagines the following conversation between Pompey and his general Cato about Caesar. Cato: “When Caesar crossed the Rubicon, there was not a town that did not welcome him with open arms. Those who come to him increase [in number] each day. They say: ‘Resistance is futile. Caesar is inevitable.’” Pompey: “These are the words of cowards. Once someone stands in his way, Caesar will no longer be inevitable.” Cato: “But no one stands in his way.” Fate, in other words, is the sum of our individual failures to act.

6 See Lewis 1986, and Stalnaker 1981. As far as modalities are concerned let me recall that, given an adequate definition of a possible world, possible is that which is true in at least one possible world; necessary is that which is true in all possible worlds; impossible is that which is untrue in all possible worlds; and contingent is that which is possible without being necessary.
otherwise.” He then defines compatibilism as “the doctrine that soft determinism may be true” (Lewis 1981, p. 113).

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 $\Box$ 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 $\Box^S$ the operator of necessity such that:

$\Box^S (p)$ 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: $\Box^S_{t_2} (C$ was the case at $t_1$)

N2: $\Box^S_{t_2} (If$ C was the case at $t_1$, then S does x at $t_2$)

Thus, by modus ponens,$^7$

N3: $\Box^S_{t_2} (S$ 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 that S does

$^7$ Whether modus ponens remains valid under the operator of necessity could be questioned.
actually do x at t, 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” (Lewis 1981, p. 113).

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” (ibid., p. 112). 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.”

The Calvinist theologian and analytic metaphysician Alvin Plantinga, who defends the weak version, has logically dubbed this
kind of ability as counterfactual “power over the past” (Plantinga 1986, p. 243).

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. Let’s consider a paradigmatic illustration which has been the object of numerous cogitations from Hobbes onward: the promise case. 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.

---

8 My master Maurice Allais, one of the founders of neoclassical economics and for a long time the only French Nobel laureate in that discipline, used to say, “When it comes to rationality, the fundamental maxim is: only the future matters.” Obviously, he did not mean that the past is not important. What Allais’s maxim asserted was that the past will always be what it was and, specifically, our present decision cannot change it.

9 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 with the implications of the counterfactual power over the past.

10 Also known in game theory as the assurance game.
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 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.”\(^\text{11}\) 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”\(^\text{12}\)), is the solution (one of the solutions) to an equation in which the unknown $x$—the future action—appears on both sides of the equation in the following form:

$$x = F \left[ x \right]$$

\(^\text{11}\) Quote from Philip K. Dick’s tale, “Minority Report,” a beautiful and profound illustration of some of the ideas presented here.

\(^\text{12}\) Reference to the method known in France as prospective, and 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 1950s. 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” (Godet and Roubelat 1996, p. 164).
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:

\[ \text{Counterfactual} \]

\[ \text{Past} \quad \rightarrow \quad \text{Future} \]

\[ \text{Causal} \]

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. Called 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, according to the received definition, is the branch of philosophy that explores the fundamental nature of reality, 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
Jean-Pierre Dupuy

on the simple basis of looking at which axiom or axioms various philosophers decided to drop. This makes a fascinating story.\(^\text{13}\)

Probably influenced by Diodorus, one of the first deniers of axiom 3 was Aristotle, as is well known because of his metaphor of the sea battle.\(^\text{14}\)

The multiplicity of metaphysics finds its origin in Diodorus’s theorem of incompatibility. A comparison with the history of geometry comes to mind. Once it was demonstrated fairly late in the history of mathematics that Euclid’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 remarked that “we choose this geometry rather than that geometry, not because it is more true, but because it is the more convenient. / To ask whether the geometry of Euclid is true […], is […] absurd” (Poincaré 1898, p. 42). 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.\(^\text{15}\) In my work on catastrophes (see Dupuy 2002)—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

\(^{13}\) The English version is more complete and a few errors have been corrected. See Vuillemin 1996.

\(^{14}\) In chapter 9 of his De Interpretatione: If a sea-battle will not be fought tomorrow, then it was also true yesterday that it will not be fought. But all past truths are necessary truths. Therefore, it is not possible that the battle will be fought.

\(^{15}\) Hence the paradoxes of backward induction. See Dupuy 2000.
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 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 ourselves 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
Jean-Pierre Dupuy

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” (Jonas 1985, pp. 113-14).

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 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.\(^\text{16}\)

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,\(^\text{17}\) the rates of inflation

\(^{16}\) 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.

\(^{17}\) A poll, by making the state of public opinion known to the public, alters this very state. When a new poll is taken, some respondents, taking note of the prior results, may be inclined to prefer the winner of the earlier poll; others, as Montesquieu long ago suggested, may try to redress the balance by throwing their support behind the runner-up. It is in order to avoid such effects that polling is prohibited in some countries in the days just before an election. In a celebrated paper published in 1954, Herbert Simon, a future Nobel laureate in
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 the winter of 2018, at a time when the prospect of a nuclear war between the US and a number of other nuclear powers is deemed by many observers stronger than it has ever been. In a lecture, former Secretary of Defense William Perry announced: “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” (Perry 2016). That was long before two heads of state whose rationality is in doubt started threatening each other with mutual annihilation. Donald Trump’s USA can indeed physically wipe the tiny North Korea off the map of the world. If it commits this abomination, even if that is in reaction to a nuclear attack launched by North Korea, it will

---

economics and one of the founders of artificial intelligence, showed that public opinion includes a “fixed point,” that is, an opinion that remains stable when informed of its own state. He even showed that cases exist in which there are several fixed points. If a polling organization wishes to give the impression that it is able to predict the future, without thereby affecting the actual course of events, it must compute and announce one of the fixed points so that public knowledge of the predicted state doesn’t change it. The existence of several such fixed points, if it is known, however, threatens the experiencing of projected time as the pollsters appear to manipulate public opinion by choosing one fixed point rather than another. Cf. Simon 1954.
be difficult for young Americans for generations after generations to look at themselves in a mirror. The event will remain forever an inexpungible mark, an indelible stain. Experts have concluded that the only way to curb Kim Jong Un, now that it is clear that all other ways fail, is to contain and deter him. However, if deterring him implies threatening to annihilate his country, isn’t the horror of the act transmitted to the horror of the threat?

In spite of their urgency, I won’t deal with these issues here, for at least two reasons that have to do with the modesty of my goals. Firstly, I will limit my analysis of nuclear deterrence to a special case: the situation called “mutually assured destruction,” known by its opportune acronym MAD. It will be clear that the confrontation between the US and North Korea doesn’t fall under that category. Secondly, I won’t address the ethical implications of MAD, only what can be called, oxymoronically, its rationality.

Here too, what I want to achieve is limited. The American filmmaker Errol Morris, in his movie The Fog of War,\(^{18}\) 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-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”\(^{19}\) 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.

---


\(^{19}\) The usual phrase is “near-misses.” Interestingly enough, it literally says the opposite of the meaning it is supposed to convey.
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. Above I referred to the religious mind as probably the cradle of the experience of projected time. In what follows it will appear that a number of features of MAD keep the mark of the sacred (see Dupuy 2014).

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* (Lee 1996). 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 into those controversies, but 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 move. 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 other arguments, in general more recent, pertain to projected time and presuppose the renunciation of strategy. Two incompatible metaphysics of time clash invisibly.20

2. A Brief History of Nuclear Deterrence Theory

For more than four decades during the Cold War, the discussion of “mutually assured destruction” (MAD) assigned a major role

---

20 To mention a similar case: in the 1970s and 1980s it was hoped by some that game theory could help give new and more solid foundations to economic theory. Efforts in that direction didn’t go very far. A tentative explanation is that game theory and its key concept of Nash equilibrium are the epitome of strategic reasoning and rest on the metaphysics of occurring time, whereas the central model of economics, that is, the Walrasian general equilibrium, exhibits a fixed-point structure that links it with projected time (cf. Dupuy 2014).
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 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 USA.

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 noncredible 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.\footnote{In a private communication, David Berlinski objects to this statement. He writes, very much in tune with philosopher David Gauthier (cf. Gauthier 1990): “No one—I mean no one—doubts that a nuclear power made victim of a first strike would unhesitatingly respond with a second strike, no matter the consequences.” I know at least one exception: in his memoirs, Valéry Giscard d’Estaing confessed that as President of France he would never ever have pressed the nuclear button, even after a first nuclear attack, lest his country should be destroyed. The French armed forces have never forgiven Giscard for having written those lines which in effect reduce the credibility and the deterrent power of the French nuclear forces to almost zero. David Berlinski goes on to say: “If the Koreans destroyed an American city, North Korea would at once be obliterated. You do not doubt that, do you?” Yes, I do: even if a Donald Trump is at the helm, I am far from certain that the US would make good on his promise to annihilate the adverse country. I am not postulating an improbable concern for the rest of the human race, but simply the anticipation of the retrospective judgment of History. Germany has not succeeded in extirpating itself from the abyss into which it fell and may never do.}

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 Carl von Clausewitz called the escalation to the extreme, that is mutual annihilation, or you cave in and your credibility is damaged 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 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 “must be the only public arrangement that is a total failure if it is successful only 99.9 percent of the time” (quoted in Lee 1996, pp. 136-37). 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 the adversaries will never 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” 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

---

22 The expression was used by Gregory Kavka (Kavka 1987, p. 20) 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, because 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. Plenty are the cases in the literature where the theory of nuclear deterrence is said to be self-invalidating.
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” (Kavka 1987, p. 48); or: “It is our military capacities that matter, not our intentions or incentives or declarations” (Lewis 1989, p. 67).

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” (Lee 1996, p. 248).

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.* (Lewis 1989, p. 68; my emphasis)
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, human-kind, 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.

According to French anthropologist René Girard, 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 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;

---

23 See Girard 2005 (orig. 1972); see also Dupuy 2016.
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._ (Brodie 1973, pp. 430-31; my emphasis)

Fate has replaced the tiger, but both images have in common
that they place the deterrent in something other than human
agency. We will return in the conclusion to a salient feature of
this extraordinary quote, namely that it conjoins contingency
(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

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 menac-
ing party is called, the latter will prefer to yield rather than being
annihilated. The would-be attacker won’t be deterred. The ques-
tion 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’s 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 (see Schelling 1960) but made famous under the moniker “Madman Theory” by Richard Nixon during the Vietnam War. The following quote is eloquent. Nixon said 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. (Haldeman 1978, p. 122)

The problem of course remains: What happens if the other side calls your bluff? In the current face-off between Donald Trump and Kim Jong Un, the question is, who is pretending to be mad and who is not pretending, because he really is mad?

However, in Brodie’s quote, we are no longer talking strategy. The twofold reference to fate and the eventuality of failure\(^\text{24}\) 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

\[\text{24} \text{ The English verb “to fail” comes from the French “faillir” whose initial meaning was to make a mistake, to commit an error—we would say today in the nuclear context: “a miscalculation.” Then “faillir” came to mean “almost,” in such expressions as “J’ai failli tomber” = “I almost fell,” that is, in the same context, a near miss/a near hit. Something bad happened, or came very close to happening, and this because of an error. To subsume all these meanings, I will use the word “accident” in its philosophical and etymological sense, that is, that which occurs without being the result of an agent’s will. The natural metaphysical habitat of accident is occurring time. This paper is about what happens to it when it is relocated to projected time.}\]
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.

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 *really* is 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.

I have arrived at the notion of superposition of states via a conceptual itinerary that owes nothing to quantum theory. However, one cannot but notice the resonances. There is probably an affinity, to say the least, between the metaphysics of projected time and some of the basic concepts of quantum theory. I cannot pursue this line of inquiry here (cf. Dupuy 2000). However, I have proposed to name the kind of uncertainty proper to projected time indeterminacy. That is the correct translation of the German word Unbestimmtheit which Heisenberg chose to name his famous relation: Unbestimmtheitsrelation, infelicitously translated as “principle of uncertainty.”

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.25

Bibliography


25 I share this conclusion with one of the greatest specialists of the domain, Stanford professor emeritus Barton Bernstein. Private communication, winter 2018.
Jean-Pierre Dupuy


Nuclear Deterrence and the Metaphysics of Time


