BELGRADE PHILOSOPHICAL ANNUAL/
FILOZOFSKI GODIŠNJAK 29/2016
Institute of Philosophy, Faculty of Philosophy, University of Belgrade
Belgrade, Čika Ljubina 18–20

Belgrade

Year XXIX

YU ISSN 0353–3891

UDK–1

Belgrade Philosophical Annual online at http://www.f.bg.ac.rs/bpa

Printed by
„Službeni Glasnik," Belgrade

This issue is financially supported by
Ministry of Education, Science and Technological Development of the Republic of Serbia

BELGRADE PHILOSOPHICAL ANNUAL 29/2016

PHILOSOPHY OF PHYSICS

ARTICLES

Tim Maudlin
NYU

Original Scientific Paper
UDK 530.145:111

THE METAPHYSICS OF QUANTUM THEORY

Why does philosophy of physics exist? Physics has been historically important to philosophers, from Thales and Democritus and Empedocles through Plato and Aristotle, Descartes and Leibniz and Kant. All of these philosophers engaged in speculation about the physical world, and that speculation informed their philosophical views. But once physics became a specialized subject of its own the relation between physics and philosophy was bound to change. Physicists with no philosophical training or interests could advance physical theory. Philosophers could call on physics if necessary, but would have little to contribute to that field *per se*. In particular, philosophers interested in physical ontology could defer to the most recent physical theories. On this model, the philosopher is the junior partner, learning at the feet of the physicist about the physical world.

This account of the relationship between physics and philosophy might have been accurate. But in order for things to work this way, the physicist would have to supply clear answers to the philosopher's queries. What the philosopher wants to know, above all, is the ontology of the best physical theories, what those theories postulate as existing. And it is precisely here that the physicist sometimes fails to provide answers. Hence the need for the philosopher to clarify the situation.

A crucial example is provided by quantum theory. The physicist has a mathematical apparatus, and knows how to use it to make predictions. But a mathematical apparatus alone does not specify an ontology. It is a further step to explicate exactly how the mathematical objects are being used as representations of physical reality. In quantum theory, there is not even agreement about what it is a representation *of*.

Let's consider how the quantum physicist goes about modeling a concrete situation. There is a target system of interest, which is called a system of N particles. Already we are in contested territory. One should not presume, just because it is called an "N-particle system" that the physics is postulating that N particles, i.e. N small or pointlike items with definite positions at all times, exist. Indeed, most physicists will deny that there really are N particles in an "N-particle system". Why then the terminology?

The talk of N particles arises as part of the construction of the mathematical apparatus. The physical state of a classical N-particle system is represented by a point in an N-particle configuration space. This is an abstract space of 3N dimensions. Each point in the configuration space represents a possible configuration of the system, i.e., represents a set ofN locations in physical space. The configuration space is, in essence, N copies of physical space, one copy for each particle. Classical configuration space does not represent a real physical space but is rather a convenient mathematical representation of a multiplicity of particles in real physical space. It is this 3N-dimensional mathematical space that the quantum physicist wants to employ in representing an "N-particle system".

But the 3N-dimensional space is not used as a configuration space: it is rather the space over which the *wavefunction* of the system is defined. While classical physics represents the state of a system by a point in configuration space, quantum theory uses instead a complex-valued function on the whole space. The wavefunction is supplied with a dynamics, a law of motion. In non-relativistic quantum mechanics this is Schrodinger's equation.

At this point we can raise a central issue: what, if anything, does the wavefunction ascribed to a system represent? Does it represent anything physically real? If so, does it represent everything physically real in the system? If not, what else is there?

And we can be more precise: if the wavefunction represents something physically real, how does it do so? Which mathematical degrees of freedom in the representation correspond to physical degrees of freedom in the system? Or, to put it another way, when do two mathematically distinct wavefunctions represent the same physical situation?

Without answers to these questions the mathematical formalism alone cannot address the philosopher's ques tions.

Let's consider some of the proposed answers to these questions. The most basic question is whether the wavefunction represents any physical feature of a system at all. A negative response is offered by the quantum Bayesians (cf. Fuchs, Mermin and Schack 2014, 749). They hold that the wavefunction represents only the information some agent has about a system, and the Schrodinger equation specifies how that knowledge does—or ought to—evolve in time. So on this view quantum theory provides little or no insight into the objective physical situation.

Quantum Bayesianism disappoints one's hopes for insight into physical reality, but may seem to provide an escape for those disturbed by quantum theory. In particular, the quantum Bayesian seems to have a simple and satisfying understanding of the so-called collapse of the wavefunction: it is nothing but conditionalization on receipt of new information. If the wavefunction represents information then gaining new information would naturally cause it to suddenly change.

Do we have good reason to reject quantum Bayesianism and take the wavefunction more seriously as a representation of some feature of the system itself? A direct reason is provided by the iconic quantum phenomenon: the two-slit experiment. Particles shot one by one through a pair of open slits show interference bands. The experiment demonstrates that the system is physically sensitive to the state of both slits, so *something* must interact with both. The spreading of the wavefunction in space is not merely a matter of our becoming more uncertain where the particle is, but is rather a matter of some real physical item spreading enough to interact with both slits.

What's more, the wavefunction does not have the right form to directly represent a state of knowledge. A state of information would be represented by a probability measure over a set of propositions, but the wavefunction is not a probability measure. It is rather a probability amplitude: one needs to take the absolute square to get a probability measure. It is only because it is an amplitude that it can give rise to interference phenomena such as the two-slit experiment. So the wavefunction is not suited mathematically to represent information, and the attempt to interpret it as epistemic falls afoul of the very phenomena one wants to explain.

We are led to the conclusion that the wavefunction represents *something* physical. This still leaves open both whether it represents *everything* in the system and exactly how it represents the physical situation.

The view that the wavefunction represents everything, together with universal Schrodinger evolution, yields a Many Worlds theory. The Many Worlds character follows from the Schrodinger evolution. In a standard Schrodinger cat situation, for example, there is one experimental condition that will yield a live cat and another that will yield a dead cat. Let's call the resulting wavefunctions in these two cases |alive> and |dead>. *At this point we assume that the physical condition represented by the former contains a live cat and the physical condition represented by the latter a dead cat.* We create a Schrodinger cat situation by. e.g., making the cat's state contingent on whether an electron is deflected up or down for a spin measurement in the z-direction. If we feed an x-spin up electron into the apparatus the linearity of Schrodinger evolution guarantees that the final state will be $1/\sqrt{2}$ |alive> + $1/\sqrt{2}$ |dead>. We need some more interpretive rules to make sense of this state.

The usual story goes as follows. The cat contains many particles, and where a particle ends up if the state is |alive> is in general different from where it ends up in |dead>. That is, the configuration of particles in a live cat is different from that in a dead cat, and differs for most of the particles. Further, the wavefunction only has significant amplitude in the regions of configuration space that correspond to these configurations. Therefore the state $1/\sqrt{2}$ |alive> + $1/\sqrt{2}$ |dead> will have two lumps to it, one in the live cat configuration and one in the dead cat configuration. These lumps will be far from each other in many dimensions of configuration space. It is very, very unlikely that further Schrodinger evolution will ever bring the lumps together again. In technical terms, the state $1/\sqrt{2}$ |alive> + $1/\sqrt{2}$ |dead> decoheres.

Decoherence means that for all practical purposes one can think of there being two states, $1/\sqrt{2}$ |alive> and $1/\sqrt{2}$ |dead>, each evolving independently of the other. But we still need to get from this to the existence of two cats, one alive and the other dead. That is, even granting that the physical situation is the sum of two pieces, we still need an argument the 1/√2 |alive> piece represents a live cat and the 1/√2 |dead> piece a dead cat.

One issue is the factor of 1/√2. Why are we justified in ignoring this factor? The general answer to this is functionalism. However it is that a wavefunction like |alive> represents a physical situation with a live cat, we assume that all that is important are the relations of parts of the physical situation to one another. An overall scaling factor like 1/√2 will not change these relations, so 1/√2 |alive> represents a live cat just as much as |alive> does.

Notice: this answer relies on the premise that |alive> represents a live cat, but as of yet we have no account of how *that* comes about. |alive> is a wavefunction with most of its amplitude concentrated in a region of the 3N-dimensional space over which it is defined, while |dead> is a wavefunction concentrated in another region. Ignoring the labels "alive" and "dead", what makes the one a representation of a live cat and the other a representation of a dead cat?

One answer that is very tempting relies on the notion that the 3N-dimensional space is, indeed a *configuration* space, and that the clumping of the wavefunction in a particular region corresponds to N particles being arranged in the corresponding configuration. But in order to be entitled to this answer a theory must postulate the existence of N particles and a 3-dimensional space they exist in. This is something Many Worlds theorists refuse to do. And even postulating the particles and the space would not solve all the problems: the wavefunction clumps in regions of the 3N-dimensional space, not at a point, so there is no particular configuration that the wavefunction would indicate.

|alive> has a particular mathematical form as a complex function of a 3N- dimensional space. It is treated as a representation of a living cat. But it is not clear exactly how this representation is to be achieved. The mathematical object must be supplemented with a *physical ontology* and *semantic rules* specifying how the physical ontology is represented by the wavefunction. Let's consider three quite different approaches to this problem.

The first approach tries to take the wavefunction as nearly isomorphic to the physical ontology as possible. Since the wavefunction is a complex function on a 3N- dimensional space, one postulates a 3N-dimensional *physical* space and a field, or pair of fields, on it. This position has been advocated by David Albert and has come to be known as *wavefunction realism* (Albert 1996, 277).[1]

Wavefunction realism makes the semantic rule connecting the wavefunction and the physical ontology easy. But the price it pays is making the connection between the physical ontology and the familiar world of experience hard. The world is not presented to us as in 3N-dimensional space but in 3-dimensional space. The familiar features of the cat, the features that distinguish the live cat from a dead cat, concern the three-dimensional structure of the cat and its parts. If the wavefunction represents physical reality this way, it is obscure how to recover that three-dimensional structure.

[1] The terminology has been adopted in Ney and Albert 2013. See Wallace and Christopher Timpson 2010, 697.

Albert's approach to this problem relies on functionalism. He seeks to show that the 3n-dimensional situation is isomorphic to a 3-dimensional one in the right way to produce an emergent 3-dimensional world. This requires an account of emergence, and in particular an emergence rather different from the way fluid behavior emerges from molecular interactions. The latter occurs because the collective action of the atoms produces behavior that is approximately that of a perfect fluid, at least at certain scales. Clearly getting a 3-dimensional space containing objects to emerge from a 3N-dimensional space is a different sort of relation.

Any successful physical theory must forge a connection between the physical world it posits and the familiar world of everyday experience. This approach to the wavefunction stands or falls on this point since the physical ontology does not obviously contain even approximations to the world of experience.

If one wants a 3-dimensional world in the physics the most obvious solution is to postulate it. Our second approach takes this tack, adding a real physical three-dimensional space to the ontology. But even with this in hand it is not immediately obvious how the wavefunction represents any material contents of the space. David Wallace and Chris Timpson have offered a suggestion called Spacetime State Realism.[2] Mathematically, the question is how to define the material contents of space from the wavefunction.

There is a natural mathematical operation that can be applied to the wavefunction of a whole system to define a mathematical representation of a part of the system: taking the partial trace. Wallace and Timpson propose using this operation to assign properties to regions of 3-dimensional space, the properties being represented by the reduced density matrix that results from the operation.

By adding the 3-dimensional space and populating it we have managed to produce a physical ontology that could correspond to the world of experience. If |dead> clumps the wavefunction in a region of configuration space corresponding to a dead cat then the properties will be d istributed in the corresponding way. Our state 1/√2 |alive> + 1/√2 |dead> would have properties distributed in a live cat configuration and properties distributed in a dead cat configuration. And since the 1/√2 |alive> piece of the wavefunction evolves independently from the 1/√2 |dead> piece, the two cats will move independently of one another, passing through each other in the 3-space. We have a Many Worlds picture.

Mathematically the wavefunction in this Many Worlds theory is identical to the wavefunction of wavefunction realism; what has changed is the physical ontology and the semantic rule. We have gotten a 3-dimensional space containing an evolving set of properties because we postulated them in the physical ontology. So our first two approaches to understanding the wavefunction demonstrate that its mathematical form does not establish its ontological significance.

The Many Worlds approach, of course, has its own challenges. We have seen how it seems to get a multiplicity of cats from the Schrodinger evolution of the wavefunction, and how decoherence ensures that the cats will not molest each other. That same decoherence also ensures that the relative squared-amplitudes for the cats will be irrelevant to how they behave. That poses the question of why the relative squared- amplitudes should have any physical significance at all. But the squared-amplitude is taken to be the *probability* of the various outcomes, so if it has no physical significance then the Many Worlds theory will be powerless to explain the way the quantum-mechanical formalism is actually used.

There is another conceptual puzzle about Spacetime State Realism. It postulates a 3-dimensional world containing localized properties, but no particles. If there are no particles there are no configurations, so why is the space over which the wavefunction is defined treated as a configuration space? We have already seen that the mathematics of the wavefunction derived from classical configuration space, but if there are no particles that is a curious choice.

Our third approach starts with this aspect of the wavefunction. This approach postulates not just a 3-dimensional space but also N particles in the space. The wavefunction can now be interpreted as a complex function over the configuration space of the system.

Once a theory postulates particles, the essential question is how the particles move. The motion of the particles constitutes a change in the configuration of the particles, so the evolution of the system corresponds to a trajectory through configuration space. Since the wavefunction can now be interpreted as a complex function on the configuration space of the system, it is natural to ask whether the trajectory could be determined by the wavefunction. Indeed it can, and the theory postulating the simplest rule here is the pilot wave theory or Bohmian Mechanics. The way the wavefunction determines the evolution of the configuration is specified by the *guidance equation*.

Bohmian mechanics has a physical ontology with three parts: a 3-dimensional space, N particles in the space, and a physical item represented by the wavefunction. Let's call the item represented by the wavefunction the *quantum state* of the system. What is the semantic rule connecting the wavefunction and the quantum state?

Mathematically different wavefunctions represent the same quantum state. If the wavefunction is changed by an overall phase the physical condition it represents doesn't change because the trajectories it produces don't change. Rescaling the magnitude uniformly also makes no difference to the motions produced by the guidance equation. In technical terms, the quantum state corresponds not to a vector in the Hilbert space of square-integrable complex functions over configuration space, but rather a ray. What most directly corresponds to the quantum state, in other words, is an element of *projective* Hilbert space.

Notice the dialectic between the mathematical representation and the physical ontology. At times the mathematical representation suggests a physical item, as with Spacetime State Realism, while at other times the physical ontology suggests adjustments to the mathematical apparatus or semantic rule. This is why the purely mathematical aspects of the theory can be uncontroversial but the ontological implications still remain obscure.

In the Bohmian theory the wavefunction represents the quantum state, which plays a purely dynamical role, guiding the particles. In Spacetime State

Realism the wavefunction plays a dual role, representing a quantum state and also a property distribution in 3-dimensional space. In wavefunction realism the wavefunction represents a real physical field on a 3N-dimensional space.

The difference in ontologies necessitates differences in the semantic rules. While a change in the overall phase of the wavefunction does not change the quantum state represented in Bohmian mechanics it does imply a different physical situation in wavefunction realism. The change in phase will not affect the property distribution in Spacetime State Realism, but it is open whether the quantum state would change. Presumably the intention is that the quantum state represented also remains the same.

The wavefunction does double duty in Spacetime State Realism because there is a dual ontology—a quantum state on the one hand and a property distribution in 3- dimensional space on the other—but only a single mathematical object. Bohmian mechanics also has a dual ontology but it employs dual mathematical representations: a point in configuration space to represent the disposition of the particles and a field on configuration space to represent the quantum state. The duality appears also in the dynamics, with the Schrodinger equation determining how the quantum state evolves and the guidance equation how the particles move.

The dualism in Bohmian mechanics has often been used as a point of criticism: the theory is said not to be quantum mechanics because it adds something—the particles—to the ontology. Many Worlds theory, in contrast, has been presented as an ontologically monistic theory. But this is misleading: Spacetime State Realism is a dualistic ontology that happens to use the same mathematical object to represent both parts of the ontology.

Not every advocate of Many Worlds is a Spacetime State Realist. But a monistic Many Worldser only has the quantum state to work with, and it does not exist in 3- dimensional space. The question of how to connect the ontology of the theory to experience therefore becomes acute. Since the evidence for the theory is expressed in terms of happenings in 3-dimensional space there must be something in the ontology of the theory that corresponds to such happenings.

There are, of course, other issues that the Many Worlds theory must face, particularly the understanding of probability in the theory. But those questions cannot even be properly framed before the physical ontology and semantic rule for interpreting the wavefunction are given.

Why hasn't this problem been more widely acknowledged? To a great extent, it has been disguised by a linguistic trick. For example, we discussed the Schrodinger cat problem using the wavefunctions called |alive> and |dead>. It is all too easy to slip into thinking that |alive> represents a physical state containing a live cat and |dead> represents a state containing a dead cat just because of how they have been labeled. But a wavefunction has to earn the right to the label via the physical ontology that it represents. Without representing some matter in a 3-dimensional space it is unclear how a wavefunction could indicate a cat in any state.

How does Bohmian mechanics do on this score? In that theory the wavefunction does not represent any matter in 3-dimensional space. But the dynamics of the theory implies that the actual configuration of the particles will typically inhabit a location in configuration space where the wavefunction has a high magnitude. So a wavefunction like |alive>, which is non-zero only in regions of configuration space that correspond to a configuration of a live cat, implies that the actual configuration is really in one of those states. The wavefunction does not directly represent the matter distribution, but it has implications for the matter distribution.

Let's look at one last issue: the collapse of the wavefunction. Standard presentations of quantum theory postulate that the wavefunction collapses—undergoes a sudden radical change—when a measurement occurs. This formulation runs afoul of the measurement problem since it is not clear exactly what physical conditions are required for a measurement to occur. How do our theories deal with this?

As we mentioned, quantum Bayesianism has a clear account of collapse. Collapses occur when one receives new information about the system, and are to be expected if the wavefunction represents information. The Many Worlds theory rejects the collapse—the wavefunction always obeys Schrodinger's equation—and reaps the consequence that there are many worlds after a measurement-type interaction. But what about Bohmian mechanics?

The wavefunction never collapses in that theory. But this does not yield many cats in a Schrodinger cat situation because the cat is composed of particles and the particles are in a definite configuration at all times. There is an additional fact: the particles are only influenced by the quantum state where the actual configuration is. In a Schrodinger cat situation the quantum state after the experiment will be the uncollapsed state represented by 1/√2 |alive> + 1/√2 |dead>. The actual configuration of the cat will either be in the region of configuration space associated with |alive> or that associated with |dead>. But since those two states will have decohered, one can safely throw away the piece that does not correspond to the actual configuration. In this way there is an effective collapse of the wavefunction even though there is no physical collapse.

The effective collapse again illustrates the dialectic between the physical ontology and the mathematical representation. Because it is part of everyday practice to collapse the wavefunction one might well expect the dynamics of the theory to have a collapse in it. But the Bohmian theory does not postulate any sudden change in the physical dynamics. Rather, collapsing as a mathematical step is justified by the character of the dynamical laws.

There is no royal road from the mathematical formalism to the physical ontology. It should be the part of the physicist to make clear what the physical ontology is and how it is represented. But at least in the case of quantum theory mainstream physics has not articulated a clear theory. Work in the foundations of physics—work done by physicists, mathematicians and philosophers—seeks to fill these gaps.

There is a certain amount of free play between the mathematical representation and the physical ontology. We have seen how this allows for competing physical accounts using the same mathematical formalism. There are questions of which parts of the mathematical formalism represent physical ontology. And there are further questions about the relation between the physical ontology and the world as we experience it. It is here, between the mathematics and the physics and the lived world, that philosophical work can thrive, both clearly articulating the ontologies of existing theories and investigating novel possibilities.

References

Albert, David. 1996. "Elementary Quantum Metaphysics." In *Bohmian Mechnaics and Quantum Theory: An Appraisal*, edited by James Cushing, Arthur Fine and Sheldon Goldstein, 277–284. Dordrecht: Kluwer.

Fuchs, Christopher A., N. David Mermin, and Rudiger Schack. 2014. "An Introduction to QBism with an Application to the Locality of Quantum Mechanics." *American Journal of Physics* 82(8): 749-754. doi: 10.1119/1.4874855

Ney, Alissa, and David Albert, eds. 2013. *The Wave Function*. Oxford: Oxford University Press.

Wallace, David, and Christopher Timpson. 2010. "Quantum Mechanics on Spacetime 1: Spacetime State Realism." *British Journal for the Philosophy of Science* 61(4): 697–727. doi: 10.2307/40981311Dae net es mi, cum earchicabori blabo. Itaquib usandebitia voluptatur magnis nectur simendis deles pratemporem qui ommos autem inctur, quibus nobit occaero restemquid eat.

Zalán Gyenis[*]
Department of Logic,
Eötvös Loránd University,
Miklós Rédei[†]
Department of Philosophy,
Logic and Scientific Method,

Original Scientific Paper
UDK 510.6

COMMON CAUSE COMPLETABILITY OF NON-CLASSICAL PROBABILITY SPACES

Abstract: *We prove that under some technical assumptions on a general, non-classical probability space, the probability space is extendible into a larger probability space that is common cause closed in the sense of containing a common cause of every correlation between elements in the space. It is argued that the philosophical significance of this common cause completability result is that it allows the defence of the Common Cause principle against certain attempts of falsification. Some open problems concerning possible strengthening of the common cause completability result are formulated.*

1 Main result

In this paper we prove a new result on the problem of common cause completability of non-classical probability spaces. A non-classical (also called: general) probability space is a pair $(\mathcal{L}, \phi)$, where $\mathcal{L}$ is an orthocomplemented, orthomodular, non-distributive lattice and $\phi\colon \mathcal{L} \to [0, 1]$ is a countably additive probability measure. Taking $\mathcal{L}$ to be a distributive lattice (Boolean algebra), one recovers classical probability theory; taking $\mathcal{L}$ to be the projection lattice of a von Neumann algebra, one obtains quantum probability theory. A general probability space $(\mathcal{L}, \phi)$ is called common cause completable if it can be embedded into a larger general probability space which is common cause complete (closed) in the sense of containing a common cause of every correlation in it. Our main result (Proposition 4.2.3) states that under some technical conditions on the lattice $\mathcal{L}$ a general probability space $(\mathcal{L}, \phi)$ is common cause completable. This result utilizes earlier results on common cause closedness of general probability theories (Kitajima 2008; Gyenis and Rédei 2014; Kitajima and Rédei 2015) and generalizes earlier results on common cause completability of classical probability spaces (Hofer-Szabó, Rédei, and Szabó 1999; 2000; Wronski 2010; Gyenis and Rédei 2011; Hofer-Szabó, Rédei, and Szabó 2013, Proposition 4.19; Marczyk and Wronski 2015; Wronski 2014).

The main conceptual-philosophical significance of the common cause extendability result proved in this paper is that it allows one to deflect the arrow

[*] Department of Logic, Eötvös Loránd University, Budapest, Hungary, gyz@renyi.hu

[†] Department of Philosophy, Logic and Scientific Method, London School of Economics and Political Science, Houghton Street, London WC2A 2AE, UK, m.redei@lse.ac.uk

of falsification directed against the Common Cause Principle for a very large and abstract class of probability theories. This will be discussed in section 5 in the more general context of how one can assess the status of the Common Cause Principle, viewed as a general metaphysical claim about the causal structure of the world. Further sections of the paper are organized as follows: Section 2 fixes some notation and recalls some facts from lattice theory and general probability spaces needed to formulate the main result. Section 3 defines the notion of common cause and common cause in general probability theories and the notion of common cause completability of such theories. The main result (Proposition 4.2.3) is stated in section 4. The detailed proof of the man proposition is given in the Appendix.

2 General probability spaces

Definition 2.1. A bounded lattice $\mathcal{L}$ with units 0 and 1 is *orthocomplemented* if there is a unary operation $\perp : \mathcal{L} \to \mathcal{L}$ satisfying

1. $a \vee a^{\perp} = 1$ and $a \wedge a^{\perp} = 0$.
2. $a \leq b$ implies $b^{\perp} \leq a^{\perp}$.
3. $(a^{\perp})^{\perp} = a$ for every $a \in \mathcal{L}$.

$\mathcal{L}$ is called *σ-complete* if $\bigvee_{i=0}^{\infty} a_i$t exists in $\mathcal{L}$ for all $a_i \in \mathcal{L}$.

Two elements $a, b \in \mathcal{L}$ are said to be *orthogonal* if $a \leq b^{\perp}$ and this we denote by $a \perp b$. $\mathcal{L}$ is called *orthomodular* if $a \leq b$ implies

$$b = a \vee (b \wedge a^{\perp}).$$

We say a and b *commutes* if

$$a = (a \wedge b) \vee (a \wedge b^{\perp}).$$

Definition 2.2. A map $\phi : \mathcal{L} \to [0, 1]$ on an orthomodular lattice $\mathcal{L}$ is defined to be a *probability measure* if the following two stipulations hold

1. $\phi(0) = 0$ and $\phi(1) = 1$.
2. Whenever $\bigvee_{i=0}^{\infty} a_i$ exists for pairwise orthogonal elements $a_i \in \mathcal{L}$ we have

$$\phi\left(\bigvee_{i=0}^{\infty} a_i\right) = \sum_{i=0}^{\infty} \phi(a_i), \qquad a_i \perp a_j \quad (i \neq j),$$

Definition 2.3.

General Probability Spaces A general probability space is a pair $(\mathcal{L}, \phi)$ where $\mathcal{L}$ is a σ-complete orthomodular lattice and ϕ is a probability measure on it.

Classical Probability Spaces If $(\mathcal{L}, \phi)$ is a general probability space and $\mathcal{L}$ is distributive then $(\mathcal{L}, \phi)$ is called semi-classical. Note that every distributive orthomodular lattice is a Boolean algebra. If $\mathcal{L}$ is isomorphic to a Boolean σ-algebra of subsets of a set X then $\mathcal{L}$ is *set-represented* and we say $(X, \mathcal{L}, \phi)$ is a classical probability space. If X is clear from the context, we omit it.

Quantum Probability Spaces Let $\mathcal{H}$ be a Hilbert space and denote the C^*-algebra of bounded linear operators in $\mathcal{H}$ by $\mathcal{B}(\mathcal{H})$. If $\mathcal{M} \subseteq \mathcal{B}(\mathcal{H})$ is a von Neumann algebra then by $\mathcal{P}(\mathcal{M})$ we understand the set of projections of $\mathcal{M}$. It is known that $\mathcal{P}(\mathcal{M})$ is a complete orthomodular lattice. A *quantum probability space* is a pair $(\mathcal{P}(\mathcal{M}), \phi)$ where $\mathcal{P}(\mathcal{M})$ is the set of projections of a von Neumann algebra $\mathcal{M}$ and ϕ is a probability measure on $\mathcal{P}(\mathcal{M})$. Probability measures arise as restrictions to $\mathcal{P}(\mathcal{M})$ of *normal states* on $\mathcal{M}$.

It is clear from the definition that both classical and quantum probability spaces are special general probability spaces. Note also that classical probability measure spaces are specific cases of quantum probability spaces: If the von Neumann algebra $\mathcal{M}$ is commutative, then the quantum probability space $(\mathcal{P}(\mathcal{M}), \phi)$ is in fact a classical probability measure space (see Rédei and Summers 2007 for a review of quantum probability spaces and their relation to classical probability theory). Thus all the definitions and notions involving general probability spaces have classical and quantum counterparts in a natural way.

Definition 2.4. A general probability space $(\mathcal{L}, \phi)$ is called *dense* if for any $a \in \mathcal{L}$ and $r \in \mathbb{R}$ with $0 \leq r \leq \phi(a)$ there is some $b \leq a$ such that $\phi(b) = r$.

The space $(\mathcal{L}, \phi)$ is called *purely non-atomic* if for all $a \in \mathcal{L}$ with $0 < \phi(a)$ there exists $b < a$ such that $0 < \phi(b) < \phi(a)$.

Remark 2.5. A general probability space $(\mathcal{L}, \phi)$ is dense if and only if it is purely non-atomic.

Definition 2.6. The general probability space $(\mathcal{L}', \phi')$ is called an *extension* of $(\mathcal{L}, \phi)$ if there exists a complete ortholattice embedding h of $\mathcal{L}$ into $\mathcal{L}'$ such that

$$\phi(x) = \phi'(h(x)) \qquad \text{for all } x \in \mathcal{L} \tag{1}$$

A lattice embedding is complete if it preserves all the infinite lattice operations as well.

3 Common cause and common cause completeness in general probability spaces

Definition 3.0.1. In a general probability space $(\mathcal{L}, \phi)$, two *commuting* events a and b are said to be (positively) correlated if

$$Corr_{\phi}(a, b) = \phi(a \wedge b) - \phi(a)\phi(b) > 0 \tag{2}$$

The event $c \in \mathcal{L}$ is a common cause of the correlation (2) if it commutes with both a and b and the following (independent) conditions hold:

$$\begin{aligned}
\phi(a \wedge b|c) &= \phi(a|c)\phi(b|c) && (3)\\
\phi(a \wedge b|c^{\perp}) &= \phi(a|c^{\perp})\phi(b|c^{\perp}) && (4)\\
\phi(a|c) &> \phi(a|c^{\perp}) && (5)\\
\phi(b|c) &> \phi(b|c^{\perp}) && (6)
\end{aligned}$$

where $\phi(x|y) = \phi(x \wedge y)/\phi(y)$ denotes the conditional probability of x on condition y, and it is assumed that none of the probabilities $\phi(x)$, $x = a, b, c, c^{\perp}$ is equal to zero.

Note that taking c to be either a or b, conditions (3)–(6) are satisfied, so, formally, both a and b are common causes of the correlation (2) between a and b; intuitively however such a "common cause" is not a "proper" common cause. A common cause is *proper* if it differs from both a and b by more than measure zero. In what follows, a common cause will *always* mean a proper common cause.

Definition 3.0.2. A general probability space $(\mathcal{L}, \phi)$ is called common cause closed (complete) if $\mathcal{L}$ contains a common cause of every correlation in it, and common cause incomplete otherwise.

Both common cause complete and common cause incomplete probability spaces exist: It was shown in Gyenis and Rédei (2003) (also see Gyenis and Rédei 2011; and Hofer-Szabó, Rédei, and Szabó 2013, Chapter 4) that no classical probability space with a Boolean algebra of finite cardinality can be common cause complete and that a dense classical probability space is common cause closed. The converse is not true, a classical probability space can be not purely non-atomic and still common cause closed; in fact, common cause closedness of classical probability spaces can be characterized completely: a classical probability space is common cause closed if and only if it has at most one measure theoretic atom (Gyenis and Rédei 2011).

A similar characterization of common cause closedness of general probability spaces is not known, only partial results have been obtained: It was proved by Kitayima that if $\mathcal{L}$ is a nonatomic, *complete*, orthomodular lattice and ϕ is a *completely* additive faithful probability measure then $(\mathcal{L}, \phi)$ has the denseness property; which was then shown to entail that every correlation between elements a and b that are logically independent has a common cause. This result was strengthened in (Gyenis and Rédei 2014) by showing that for $(\mathcal{L}, \phi)$ to be dense it is enough that $\mathcal{L}$ is only σ-complete and the faithful ф is σ-additive, and that in this case every correlation has a common cause in $\mathcal{L}$ – not just correlations between logically independent elements – hence such general probability spaces are common cause closed. Just like in the classical case, a general probability space can be common cause closed and not being purely nonatomic: it was shown in (Gyenis and Rédei 2014) that (under an additional technical condition called "Q-decomposability" of a probability measure) a general probability measure space is common cause closed if it has only one measure theoretic atom. The Q-decomposability condition could be shown to be redundant in the particular case when $(\mathcal{L}, \phi)$ is a quantum probability space: it was proved in (Kitajima and Rédei 2015) that if ϕ is a faithful normal state on the von Neumann algebra $\mathcal{N}$, then $(\mathcal{P}(\mathcal{N}), \phi)$ is common cause closed if and only if there is *at most one* ϕ-atom in the projection lattice $\mathcal{P}(\mathcal{N})$. This result entails that the standard quantum probability theory $(\mathcal{P}(\mathcal{H}), \phi)$, which has more than one measure theoretic atom, is not common cause closed. It is however still an open question whether having *at most one* measure theoretic atom is equivalent to common cause closedness of a general probability space.

If a probability space $(\mathcal{L}, \phi)$ is common cause incomplete, then the question arises whether it can be common cause completed. Common cause completion of $(\mathcal{L}, \phi)$ with respect to a set $\{(a_i, b_i) : i \in I\}$ of pairs of correlated elements is meant finding an extension $(\mathcal{L}', \phi')$ of $(\mathcal{L}, \phi)$ (Definition 2.6) such that $(\mathcal{L}', \phi')$ contains a common cause $c_i \in \mathcal{L}'$ for every correlated pair (a_i, b_i), $i \in I$. If a probability space $(\mathcal{L}, \phi)$ is common case completable with respect to *all* the correlations in it, we simply say that $(\mathcal{L}, \phi)$ is common cause completable.

Note that the definition of extension, and in particular condition (1), implies that if $(\mathcal{L}', \phi')$ is an extension of $(\mathcal{L}, \phi)$ with respect to the embedding h, then every single correlation $Corr_p(a, b)$ in $(\mathcal{L}, \phi)$ is carried over intact by h into the correlation $Corr_{\phi'}(h(a), h(b))$ in $(\mathcal{L}', \phi')$ because

$$\begin{aligned} \phi'(h(a) \wedge h(b)) &= \phi'(h(a \wedge b)) \\ &= \phi(a \wedge b) > \phi(a)\phi(b) = \phi'(h(a))\, \phi'(h(b)) \end{aligned}$$

Hence, it does make sense to ask whether a correlation in $(\mathcal{L}, \phi)$ has a Reichenbachian common cause in the extension $(\mathcal{L}', \phi')$.

It was shown in Hofer-Szabó, Rédei, and Szabó (1999) that every common cause incomplete classical probability space is common cause completable with respect to any *finite* set of correlations. This result was strengthened by proving that classical probability spaces are common cause completable with respect to *any* set of correlations (Gyenis and Rédei 2011), i.e. every classical probability space is common cause completable. In fact, it was showed in Wronski (2010) that every common cause incomplete classical probability space has an extension that is common cause closed (also see Gyenis and Rédei 2011; Hofer-Szabó, Rédei, and Szabó 2013, Proposition 4.19). This settled the problem of common cause completability of classical probability spaces.

For non-classical spaces, Hofer-Szabó, Rédei, and Szabó (1999) proved that every quantum probability space $(\mathcal{P}(\mathcal{M}), \phi)$ is common cause completable with respect to the set of pairs of events that are correlated in the state ϕ. But common cause completability of general probability theories has remained open so far. The proposition in the next section states common cause completability of general probability theories under some technical conditions on the lattice $\mathcal{L}$.

4 Proposition on dense extension and common cause completability of certain general probability spaces

The results on common cause closedness of general probability spaces recalled in the previous section make it clear that in order to show that a general probability space $(\mathcal{L}, \phi)$ is common cause com- pletable, it is enough to prove that $(\mathcal{L}, \phi)$ has an extension that is dense. In this section we present a method of extending certain general probability spaces to a dense one. In the case of classical and quantum probability spaces there are standard methods of extension that can be applied to obtain a dense extension of these probability theories. We discuss briefly this method in Subsection 4.1. However, this method does not seem to work in the general case; hence in Subsection 4.2 we introduce another construction of an extension which can be applied to a broader class of probability spaces and which leads to a dense extension. Finally, in Subsection 5.1 we discuss how this new construction can be applied to classical and quantum probability spaces. Here and in Subsection 4.1 we will be sketchy.

4.1 Brief overview of standard product extension of classical and quantum probability spaces into dense spaces

The main reason why classical and quantum probability theories can be extended into a dense one is that one can define the notion of tensor product of these probability spaces. In contrast, it is not that obvious how to define tensor product of general orthomodular lattices; in general such tensor products do not exits.

Classical probability spaces. Let $\mathfrak{L} = (X, \mathcal{L}, \phi)$ and $\mathfrak{X} = (I, \Lambda, \lambda)$ be two classical probability spaces and let $\mathfrak{L} \otimes \mathfrak{X}$ be their usual product. Then the function $h : \mathcal{L} \to \Lambda$ defined as

$$h(A) = I \times A$$

is an ortho-embedding which shows that $\mathfrak{L} \otimes \mathfrak{X}$ is an extension of $\mathfrak{L}$. Taking then $\mathfrak{X}$ to be dense, for example Λ is the set of Borel subsets of the unit interval I with the Lebesgue measure λ, it is not hard to see that $\mathfrak{L} \otimes \mathfrak{X}$ is dense, too. A detailed proof can be found in Wronski (2010) and Hofer-Szabó, Rédei, and Szabó (2013) (proof of Proposition 4.19).

Quantum probability spaces. A similar method works in the quantum case: Let $(\mathcal{P}(\mathcal{M}), \phi)$ and $(\mathcal{P}(\mathcal{N}), \psi)$ be two quantum probability spaces where $\mathcal{N}$ is a type III von Neumann algebra and ψ is a faithful normal state on $\mathcal{N}$. Then by Lemma 4 in Rédei and Summers (2002) the quantum probability space $(\mathcal{P}(\mathcal{N}), \psi)$ is dense. Consider now the tensor product quantum probability space $(\mathcal{P}(\mathcal{M} \otimes \mathcal{N}), \phi \otimes \psi$. This is an extension of $(\mathcal{P}(\mathcal{M}), \phi)$ with the embedding $h(X) = X \otimes I$. Since $\mathcal{M} \otimes \mathcal{N}$ is a type III von Neumann algebra if $\mathcal{N}$ is (Kadison and Ringrose 1986, Chapter 11), again by Lemma 4 in Rédei and Summers (2002) the quantum probability space $(\mathcal{P}(\mathcal{M} \otimes \mathcal{N}), \phi \otimes \psi)$ is dense too.

4.2 Extension of a-continuous general probability spaces

Let $(\mathcal{L}, \phi)$ be a general probability space. To state our main proposition we have to recall some further lattice theoretic notions. Let (a_i) be a sequence of elements of $\mathcal{L}$. Then

$$\liminf(a_i) \doteq \bigvee_{i=0}^{\infty} \bigwedge_{k=i}^{\infty} a_k \tag{7}$$

$$\limsup(a_i) \doteq \bigwedge_{i=0}^{\infty} \bigvee_{k=i}^{\infty} a_k . \tag{8}$$

if the right hand sides exist. We define $\lim(a_i)$ if and only if $\limsup(a_i) = \lim\inf(a_i)$, and in this case

$$\lim(a_i) \doteq \lim\inf(a_i).$$

Clearly

$$\bigvee_{i=0}^{\infty} a_i = \mathsf{lim}(a_0, a_0 \vee a_1, \ldots, \vee_{i=0}^{k} a_i, \ldots) \tag{9}$$

$$\bigwedge_{i=0}^{\infty} a_i = \mathsf{lim}(a_0, a_0 \wedge a_1, \ldots, \wedge_{i=0}^{k} a_i, \ldots), \tag{10}$$

whenever the left hand sides are defined in $\mathcal{L}$ (e.g. $\mathcal{L}$ is σ-complete).

Next we prove that

$$\phi(\lim(a_i)) = \lim_{i\to\infty} \phi(a_i),$$

whenever $\lim(a_i)$ is defined. In order to show this observe first, that by the orthomodular law, since $a \leq a \vee b$ we have

$$a \vee b = a \vee (a^{\perp} \wedge (a \vee b)),$$

where clearly $a \perp a^{\perp} \wedge (a \vee b)$. Hence, by an easy induction argument, it follows that if (a_i) is a sequence of elements such that $\vee_{i=0}^{\infty} a_i$ exists, then there is another sequence (b_i) of elements, such that for all $k \in \mathbb{N}$ we have

$$\bigvee_{i=0}^{k} a_i = \bigvee_{i=0}^{k} b_i,$$

and $b_i \perp b_j$ for $i \neq j$. It also follows that $\vee_{i=0}^{\infty} b_i$ exists and is equal to $\vee_{i=0}^{\infty} a_i$. Therefore by σ-additivity of ϕ we get

$$\phi(\bigvee_{i=0}^{\infty} a_i) = \phi(\bigvee_{i=0}^{\infty} b_i) = \sum_{i=0}^{\infty} \phi(b_i) = \lim \sum_{i=0}^{k} \phi(b_k) = \lim \sum_{i=0}^{k} \phi(a_k).$$

Hence, we proved that

$$\phi(\lim(a_0, a_0 \vee a_1, \ldots)) = \lim\,(\phi(a_0), \phi(a_0 \vee a_1), \ldots).$$

Now, if (c_i) is any sequence such that $\lim(c_i)$ exists, then by the definition of lim and in particular lim inf, there is another sequence (d_i) such that $\lim(c_i) = \vee_{i=0}^{\infty} d_i$. Using this and by the previous argument, it is not hard to see, that

$$\phi(\lim(c_i)) = \lim \phi(c_i).$$

Definition 4.2.1. A lattice $\mathcal{L}$ is called *σ-continuous* if the following two stipulations hold

1. $\lim(a_i) \wedge \lim(b_i) = \lim(a_i \wedge b_i)$.
2. $\lim(a_i) \vee \lim(b_i) = \lim(a_i \vee b_i)$.

whenever the limits $\lim(a_i)$ and $\lim(b_i)$ exist in $\mathcal{L}$. Note that the equation $(\lim(a_i))^{\perp} = \lim(a_i^{\perp})$ automatically holds in orthocomplemented lattices.

An important class of σ-continuous orthomodular lattices are the finite ones.

We can now state our main result:

Proposition 4.2.2. Let $(\mathcal{L}, \phi)$ be a general probability space where $\mathcal{L}$ is σ-continuous. Then there exists a dense extension $(\mathcal{L}', \phi')$ of $(\mathcal{L}, \phi)$.

This proposition entails:

Proposition 4.2.3. Let $(\mathcal{L}, \phi)$ be a general probability space where $\mathcal{L}$ is σ-continuous. Then $(\mathcal{L}, \phi)$ is common cause completable.

The proof of Proposition 4.2.2 consists of the following 3 steps (details can be found in the Appendix):

Step 1. Using an arbitrary classical probability space $\mathfrak{X}$ we construct first an orthomodular lattice $S = S(\mathfrak{X}, \mathcal{L})$ in such a way that $\mathcal{L}$ is a sub-orthomodular lattice of S.

Step 2. We define a measure ρ on S in a way that (S, ρ) becomes an extension of $(\mathcal{L}, \phi)$. Unfortunately S is not in general σ-complete, so we need the one more step.

Step 3. We extend S to a σ-complete orthomodular lattice while paying attention not to ruin ρ-measurability.

5 Common cause completability and the Common Cause Principle

Establishing results on common cause completability of general probability spaces is motivated by the need to assess the status of what is known as the Common Cause Principle. The Common Cause Principle is a claim about the causal structure of the world, and it states that if there is a probabilistic correlation between two events, then either there is a direct causal link between the correlated events that explains the correlation, or there is a third event, a common cause that brings about (hence explains) the correlation. This principle goes back to Reichenbach's work that defined the notion of common cause in terms of classical probability theory (Reichenbach 1956), and the principle was sharply articulated mainly by Salmon (Salmon 1978; 1984). There is a huge literature on the problem of whether this principle reflects correctly the causal structure of the world (see the references in Hofer-Szabó, Rédei, and Szabó 2013 and Wronski 2014). The difficulty in giving a definite answer to this question is that the Common Cause Principle has metaphysical character: it is a strictly universal claim containing two general existential statements. Standard arguments well known from the history of philosophy show that such general claims can be neither verified nor falsified conclusively; thus, as it was argued in in Hofer-Szabó, Rédei, and Szabó (2013, Chapters 1 and 10) and Rédei (2014), the only option one has when it comes to the problem of assessing the epistemic status of the principle is to have a look at the best available evidence relevant for the principle and see whether they are in harmony with the principle or not. Such evidence is provided by the empirical sciences.

Some of the (experimentally testable and confirmed) correlations are predicted by physical theories that apply non-classical probability spaces to describe the phenomena in their domain: Standard non-relativistic quantum mechanics of finite degrees of freedom is based on non-classical probability theory of the form $(\mathcal{P}(\mathcal{H}), \phi)$, where $\mathcal{P}(\mathcal{H})$ is the lattice of all projections of the von Neumann algebra $\mathcal{B}(\mathcal{H})$ consisting of all bounded operators on Hilbert space $\mathcal{H}$, and ϕ is a quantum state.

This type of quantum theory predicts the notorious EPR correlations (Hofer-Szabó, Rédei, and Szabó 2013, Chapter 9). Relativistic quantum field theory (in the so called algebraic approach (Horuzhy 1990; Haag 1992; Araki 1999)) is based on probability theory of the form $(\mathcal{P}(\mathcal{N}), \phi)$, where $\mathcal{P}(\mathcal{N})$ is the projection lattice of a type III von Neumann algebra and ϕ is a normal state on $\mathcal{N}$. Quantum field theory predicts an abundance of correlations between observables that are localized in spacelike separated spacetime regions. This is a consequence of violation of Bell's inequality in quantum field theory (for a review of the relevant theorems and references see Hofer-Szabó, Rédei, and Szabó 2013, Chapter 8). Given these correlations predicted by physical theories using non-classical probability spaces, the problem arises whether these correlations are in harmony with the Common Cause Principle. This is a very subtle and complicated matter. One has to specify very carefully what precisely the harmony would be and, given a specification, one can try to show that harmony obtains (or not).

A possible (and very natural Hofer-Szabó, Rédei, and Szabó 2013, Chapters 1 and 10; Rédei 2014) line of reasoning aimed at assessing the Common Cause Principle in this spirit is the following: Suppose a theory T applies a non-classical probability theory $(\mathcal{L}, \phi)$ to describe phenomena and predicts correlation between elements a, b in $\mathcal{L}$ such that, according to T, there is no (cannot be) a direct causal connection between a and b. The first question one would want to ask then: Is there a common cause c in $\mathcal{L}$ of the correlation between a and b? If there is, then theory T is a confirming evidence in favor of the Common Cause Principle: T is then a causally complete theory that can give an explanation of the correlations it predicts. We know from results about common cause incompleteness referred to earlier in this paper that it can happen that there is no common cause in $\mathcal{L}$ of the correlation – i.e. that $(\mathcal{L}, \phi)$ is common cause incomplete. This makes T a potentially disconfirming evidence for the Common cause Principle. But the mere fact that T is common cause incomplete does not falsify the Common Cause Principle because one can mount the following defence: The general probability theory $(\mathcal{L}, \phi)$ applied by T is just too meager, and there might exist "hidden" common causes of this correlation – hidden in the sense of being part of a larger probability theory $(\mathcal{L}', \phi')$ that extends $(\mathcal{L}, \phi)$. The conceptual significance of common cause extendability of general probability spaces should now be clear: common cause extendability entails that this kind of defence of the Common Cause Principle against potential falsifiers is *always* possible – and it is possible for an extremely wide class of general probability theories. This broad class includes both classical and quantum probability theories in particular.

The general common cause completability result and the above reasoning lead to the question of whether common cause completability also holds under more stringent conditions. A particularly relevant type of additional conditions are the ones that express "locality" understood as constraints on a physical theory T that make T to be compatible with the principles of the theory of relativity. A special such case is the quantum probability theory ($\mathcal{P}$ ($\mathcal{N}$), ϕ) with a type III von Neumann algebra $\mathcal{N}$: This quantum probability measure space theory describes relativistic quantum fields and it has a further internal (so called "quasi-local") structure. This local structure makes it possible to impose further, physically motivated locality conditions on the common causes. Under this further condition it becomes a highly non-trivial problem to decide whether the theory is common cause complete. One difficulty of this problem is that the locality conditions can be defined in different ways and causal completeness seems to depend sensitively on how the locality conditions are defined. The most natural locality condition leads to the problem of causal closedness of quantum field theory (Rédei 1997) that is still open. It could be proved however that quantum field theory is causally complete with respect to a notion of weakly localized common causes (Rédei and Summers 2002; see also the extensive discussion in Hofer-Szabó, Rédei, and Szabó 2013, Chapter 8). Since we do not have new results on local common causes in this paper, we do not give here the precise definitions of locality.

A further direction of possible research concerns strengthening the common cause completability results in this paper by taking into account the "type" of the common cause. A common cause c is said to have the type characterized by five positive real numbers (r_c, $r_{a|c}$, $r_{a|c^\perp}$, $r_{b|c}$, $r_{b|c^\perp}$) if these numbers are equal with the probabilities of events indicated by the subscript of the numbers, i.e. if $r_c = \phi(c)$, $r_{a|c} = \phi(a|c)$, $r_{a|c^\perp} = \phi(a|c\perp)$, $r_{b|c} = \phi(b|c)$, $r_{b|c^\perp} = \phi(b|c^\perp)$ (Definition 3.6 in Hofer-Szabó, Rédei, and Szabó 2013). One then can define *strong* common cause closedness of a general probability space ($\mathcal{L}$, ϕ) by requiring that for any correlation in this space and for any given possible type there exists in $\mathcal{L}$ a common cause of the given type. A general probability space can then be defined to be *strongly* common cause completable if it has an extension that is *strongly* common cause closed. It is not known whether general probability spaces are strongly common cause completable (cf. Problem 6.2 in Hofer-Szabó, Rédei, and Szabó 2013).

Appendix

Step 1.

Construction of $S(\mathcal{X}, \mathcal{L})$. Let $\mathcal{X} = (X, \Sigma, \mu)$ be a classical probability measure space, that is, Σ is a σ-algebra of subsets of X and μ is a probability measure on Σ, and let $\mathcal{L}$ be an orthomodular lattice.

For an element $a \in \mathcal{L}$ and a subset $B \subseteq X$ we define $X^a_B : B \to \mathcal{L}$ as

$$X^a_B(x) = a.$$

Next, we define *step functions*. A function $p : X \to \mathcal{L}$ is called a *step function* if it is of the following form

$$p = \bigsqcup_{i=0}^{\infty} \chi_{B_i}^{a_i},$$

where $a_i \in \mathcal{L}$, $B_i \subseteq X$ is measurable, i.e. $B_i \in \Sigma$ and $X = \bigsqcup_{i=0}^{\infty} B_i$ is a disjoint partition of X. So a step function is a function whose domain can be partitioned into countably many measurable sets and the function is constant on each partition.

Definition 5.0.1. For a classical probability space $\mathfrak{X} = (X, \Sigma, \mu)$ and an orthomodular lattice $\mathcal{L}$ we define

$$\begin{aligned} \mathcal{F}(X, \mathcal{L}) &\doteq \{p : p\colon X \to \mathcal{L} \text{ is a function}\} \\ S(\mathfrak{X}, \mathcal{L}) &\doteq \{p : p\colon X \to \mathcal{L} \text{ is a step function}\}. \end{aligned}$$

$\mathcal{F}(X, \mathcal{L})$ is an orthomodular lattice with the pointwise operations as follows: Suppose $f \in \{\wedge, \vee, \perp, 0, 1\}$ is an ℓ-ary operation and let $p_i \in \mathcal{F}(X, \mathcal{L})$ for $i < \ell$. Let the element $f^{\mathcal{F}}(p_0, \ldots, p_{\ell-1})$ be defined as

$$f^{\mathcal{F}}(p_0, \ldots, p_{\ell-1})\,(x) = f^{\mathcal{L}}(p_0(x), \ldots, p_{\ell-1}(x)),$$

for all $x \in X$. Note that $\mathcal{F}(X, \mathcal{L})$ is nothing else but the power $\mathcal{L}^X$.

Lemma 5.0.2.

(1) $S(\mathfrak{X}, \mathcal{L})$ is a subalgebra of $\mathcal{F}(X, \mathcal{L})$ and hence it is an orthomodular lattice.

(2) $\mathcal{L}$ can be completely embedded into $S\,(\mathfrak{X}, \mathcal{L})$.

Proof. (1) For simplicity denote $S\,(\mathfrak{X}, \mathcal{L})$ and $\mathcal{F}(X, \mathcal{L})$ respectively by S and $\mathcal{F}$. It is clear that $S \subseteq \mathcal{F}$, thus we have to prove that S is closed under the operations of $\mathcal{F}$.

Suppose $f \in \{\wedge, \vee, \perp, 0, 1\}$ is an ℓ-ary operation and let $p_i \in S$ for $i < \ell$. By definition of a step function, each p_i can be written in the following form

$$p_i = \bigsqcup_{j=0}^{\infty} \chi_{B_{i,j}}^{a_{i,j}},$$

where $a_{i,j} \in \mathcal{L}$, $B_{i,j} \in \Sigma$, such that for all i we have

$$X = \bigsqcup_{j=0}^{\infty} B_{i,j}.$$

Using that Σ is closed under countable intersections and unions, one can find a partition

$$X = \bigsqcup_{j=0}^{\infty} C_j$$

where $C_i \in \Sigma$ for all i and this partition is a common refinement of the partitions $\{B_{i,j}\}$. Thus we can conclude that each p_i is constant on each C_i:

$$p_i = \bigsqcup_{j=0}^{\infty} \chi_{C_j}^{a_{i,j}},$$

Then it is easy to see, that $f^{\mathcal{F}}(\mathrm{p}_0, \ldots, p_{\ell\text{-}1})$ is also constant on each Cj:

$$f^{\mathcal{F}}(p_0, \ldots, p_{\ell-1}) = \bigsqcup_{j=0}^{\infty} \chi_{C_j}^{f^{\mathcal{L}}(a_{0,j}, \ldots, a_{\ell-1,j})}.$$

But this is a step function, hence belongs to S. Thus we proved that S is closed under the operations of $\mathcal{F}$, and therefore it is a subalgebra of $\mathcal{F}$. Since subalgebras of orthomodular lattices are orthomodular lattices the proof is complete.
(2) Let $h : \mathcal{L} \to S$ be defined as

$$h(a) \doteq \mathcal{X}^{a}x. \tag{11}$$

Then h is an embedding which preserves every operation (infinite ones as well). Checking this is a routine and omitted. ▪

Step 2.

Construction of (S, ρ). Let $(\mathcal{L}, \phi)$ be a general and let $\mathfrak{X} = (X, \Sigma, \mu)$ be a classical probability space. Let $\mathcal{F} = \mathcal{F}(X, \mathcal{L})$ and $S = S(\mathfrak{X}, \mathcal{L})$ be as above. A function $f : X \to \mathcal{L}$ is called *μ-integrable* if

$$\phi \circ f : X \to [0, 1]$$

is (Σ, Λ)-measurable, where Λ is the Lebesgue σ-algebra of subsets of the unit interval. Then, because $\phi \circ f$ is non-negative, the integral $\int_X \phi \circ f \, d\mu$ exists. For μ-integrable functions $f \in \mathcal{F}$ we define

$$\rho(f) \doteq \int_X \phi \circ f \, d\mu$$

Every step function is μ-integrable, hence $\rho(f)$ is defined for every element $f \in S$. We prove that ρ is a probability measure on S.

Lemma 5.0.3. Suppose that $f_0, f_1, \ldots$ are μ-integrable functions and suppose $\lim(f_i)$ exists in $\mathcal{F}$. Then

$$\rho(\lim(f_i)) = \lim_{i \to \infty} \rho(f_i). \tag{12}$$

Specifically, $\lim(f_i)$ is μ-integrable and ρ is a probability measure on S.

Proof. Fix an arbitrary $x \in X$. Then

$$\phi(\lim f_i(x)) = \lim \phi(f_i(x)), \tag{13}$$

because ϕ is a measure on $\mathcal{L}$ and hence

$$\phi \circ (\lim f_i) = \lim \phi \circ f_i. \tag{14}$$

Then by definition

$$\rho(\mathsf{lim} f_i) = \int \phi \circ (\mathsf{lim} f_i) = \int \lim \phi \circ f_i \overset{!}{=} \lim \int \phi \circ f_i = \lim \rho(f_i),$$

where the equality marked with ! is a consequence of the monotone convergence theorem. $\rho(1) = 1$ and $\rho(0) = 0$, so ρ is normalized. ■

Next, we prove that (S, ρ) is an extension of $(\mathcal{L}, \phi)$.

Lemma 5.0.4. (S, ρ) is an extension of $(\mathcal{L}, \phi)$.

Proof. We need to show that ρ extends $\phi \circ h$, where h is defined by (11):

$$\rho(h(a)) = \int_X \phi \circ \chi_X^a \, d\mu = \int_X \phi(a) \, d\mu = \phi(a) \cdot \int_X 1 \, d\mu = \phi(a)$$

■

Step 3.

σ-complete extension of (S, ρ). Observe that (S, ρ) is not necessarily a general probability space because S is not, in general, σ-complete (albeit ρ is σ-additive). In what follows we prove that S can be extended to a σ-complete orthomodular lattice $\mathcal{Q}$ in such a way that $(\mathcal{Q}, \rho)$ is a general probability space (and thus an extension of $(\mathcal{L}, \phi)$).

Lemma 5.0.5. Let $\mathcal{B} \leq \mathcal{F}$ be a subalgebra of $\mathcal{F}$ such that every element of $\mathcal{B}$ is μ-integrable. Let $\{a_i\}_{i\in\mathbb{N}} \subseteq \mathcal{B}$ be a sequence such that $\lim(a_i)$ exists (in $\mathcal{F}$). Let

$$\mathcal{Y} = \langle \mathcal{B}, \lim(a_i) \rangle$$

be the generated subalgebra of $\mathcal{F}$. Then every element of $\mathcal{Y}$ is μ-integrable.

Proof. For simplicity, let $a = \lim(a_i)$. If $a \in \mathcal{B}$ then there is nothing to prove. Since $\mathcal{Y}$ is a generated algebra, every element $y \in \mathcal{Y}$ can be written in the form

$$y = t(b_0, \ldots, b_n),$$

where $b_i \in \mathcal{B} \cup \{a\}$ and t is a term in the algebraic language of $\mathcal{B}$. Clearly, if a doesn't occur in t, then $y \in \mathcal{B}$. So we may assume that

$$y = t(\overline{b}, a).$$

We need to prove that $t(\overline{b}, a)$ is μ-integrable. This we do by induction on the complexity of the term t. If t is a constant $b \in \mathcal{B}$ then it is μ-integrable since it belongs to $\mathcal{B}$. If t is the constant a, then it is μ-integrable by Lemma 5.0.3.

Now, suppose that $t_1(\overline{b}_1, a), \ldots, t_n(\overline{b}_n, a)$ are μ-integrable and let $f \in \{\wedge, \vee, \perp\}$ be any n-ary operation. Then, by continuity of $\mathcal{L}$, we have

$$f\Big(t_1(\bar{b}_1, a), \ldots, t_n(\bar{b}_n, a)\Big) = \mathsf{lim}_{i\to\infty} f\Big(t_1(\bar{b}_1, a_i), \ldots, t_n(\bar{b}_n, a_i)\Big),$$

which is a limit of elements of $\mathcal{B}$, hence by Lemma 5.0.3, it is μ-integrable. ■

Define now a set K of subalgebras of $\mathcal{F}$ as follows:

$$\mathsf{K} = \{\mathcal{B} : \mathcal{S} \leq \mathcal{B} \leq \mathcal{F} \text{ and every elements of } \mathcal{B} \text{ are } \mu \text{ — integrable}\}.$$

If $\mathcal{B}_\alpha$, $(\alpha < \kappa)$ is a chain from K, then $\bigcup_{\alpha < \kappa} \mathcal{B}_\alpha$ also belongs to K, whence the assumption of Zorn's lemma is satisfied. Consequently, there exists a maximal subalgebra $\mathcal{Q} \in \mathsf{K}$. Because every elements of $\mathcal{Q}$ are μ-integrable the measure ρ is defined on $\mathcal{Q}$. We claim that $(\mathcal{Q}, \rho)$ is a general probability space.

Lemma 5.0.6. $(\mathcal{Q}, \rho)$ is a general probability space.

Proof. We only have to show that $\mathcal{Q}$ is σ-complete. Let $a_i \in \mathcal{Q}$ for $i \in \mathbb{N}$ such that $\bigvee_{i=0}^{N} a_i \in \mathcal{Q}$ for all natural number N. We should prove

$$\bigvee_{i=0}^{\infty} a_i \in \mathcal{Q}.$$

For if not, by Lemma 5.0.5, it follows that

$$\mathcal{Q} \leq \langle \mathcal{Q}, \bigvee_{i=0}^{\infty} a_i \rangle$$

is a proper superalgebra of $\mathcal{Q}$ such that any of its elements are μ-integrable. This contradicts to the maximality of $\mathcal{Q}$. ▪

Lemma 5.0.7. If $\mathfrak{X} = (X, \Sigma, \mu)$ is dense then so is $(\mathcal{Q}, \rho)$.

Proof. It is enough to prove that for all $f \in \mathcal{Q}$ with $0 < \rho(f)$ there exists $g \in \mathcal{Q}$ such that $g \leq f$ and $0 < \rho(g) < \rho(f)$ since using σ-completeness of $\mathcal{Q}$ and σ-additivity of ρ this ensures denseness.

Suppose f is a step function. Then there exists a measurable subset A of X (i.e. $A \in \Sigma$) with $\mu(A) > 0$ such that if

$$g(x) = \begin{cases} f(x) & \text{if } x \in A \\ 0 & \text{otherwise} \end{cases}$$

then g is a step function with $\rho(g) > 0$ and $g \leq f$. By denseness of μ, there is a smaller set $B \subset A$ with $0 < \mu(B) < \mu(A)$. Let

$$f'(x) = \begin{cases} 0 & \text{if } x \in A \setminus B \\ g(x) & \text{otherwise} \end{cases}$$

Then $0 < \rho(f') < \rho(f)$ and $f' \prec f$ is in S.

In the general case if $f \in \mathcal{Q}$ then by the construction of $\mathcal{Q}$ there is some step function $f' \in \mathcal{Q}$ with $f' \leq f$ and $0 < \rho(f') \leq \rho(f)$. Applying the previous argument to f' completes the proof. ▪

Proof of Proposition 4.2.2. Let $(\mathcal{L}, \phi)$ be a σ-continuous general probability space. Then $(\mathcal{Q}, \rho)$ constructed above with the choice $\mathfrak{X} = ([0, 1], \Lambda, \lambda)$ is a dense extension of $(\mathcal{L}, \phi)$. ▪

5.1 The extension in the classical and quantum case

We illustrate the extension method presented in Subsection 4.2 by showing how it works in classical and quantum probability spaces. Since for these cases a simpler method exist (see Subsection 4.1) we will be brief and sketchy.

Classical Case. Since every distributive lattice is σ-continuous, Proposition 4.2.2 can directly applied to semi-classical spaces. For classical spaces it is enough to observe that the space $(\mathcal{Q}, \rho)$ is set-represented provided $(\mathcal{L}, \rho)$ is set-represented. A somewhat detailed form of the construction in this case can be found in Gyenis and Rédei (2011).

Quantum Case. Fix a quantum probability space $(\mathcal{P}(\mathcal{M}), \phi)$, where $\mathcal{M} \subseteq \mathcal{B}(\mathcal{H})$ is a von Neumann algebra acting on the separable Hilbert space $\mathcal{H}$, $\mathcal{P}(\mathcal{M})$ is the lattice of projections of $\mathcal{M}$ and ϕ a normal state on $\mathcal{M}$. We may assume, in fact, that ϕ is a normal state on $\mathcal{B}(\mathcal{H})$. Further fix a classical probability space $\mathfrak{X} = (X, \Sigma, \mu)$, where X is a set, Σ is a σ-algebra of subsets of X and μ is a probability measure on Σ. Throughout we will have $\mathfrak{X} = ([0, 1], \Lambda, \lambda)$ in mind, where Λ is the σ-algebra of Lebesgue-measurable subsets of the unit interval, and λ is the Lebesgue measure.

We start by recalling some definitions and facts from Hilbert space theory (e.g. from Fonseca and G. Leoni 2007; Kadison and Ringrose 1986; and Takesaki 2003).

Definition 5.1.1. A function $s : X \to \mathcal{H}$ is called *simple* if it is of the following form:

$$s = \sum_{i=1}^{\ell} h_i \chi_{E_i},$$

where $\ell \in \mathbb{N}$, $h_i \in \mathcal{H}$ and the sets E_i form a partition of X such that each E_i is measurable ($E_i \in \Sigma$). χE is the characteristic function of the set E.

A function $u : X \to \mathcal{H}$ is called *strongly measurable*, if there is a sequence $\{s_n\}_{i\in\mathbb{N}}$ of simple functions such that

$$\lim ||s_n(x) - u(x)||_{\mathcal{H}} = 0 \qquad \text{for } \mu\text{-almost all } x \in X.$$

If $u : X \to \mathcal{H}$ is strongly measurable, then the map $x \mapsto ||u(x)||_{\mathcal{H}}$ is measurable in the classical sense, i.e. (Σ, Λ)-measurable.

Definition 5.1.2. We define $\mathfrak{H} = L^2(\mathfrak{X}, \mathcal{H})$ as follows:

$$L^2(\mathfrak{X}, \mathcal{H}) = \left\{u : X \to \mathcal{H} : u \text{ is strongly measurable and } ||u||_{\mathfrak{H}} < \infty\right\},$$

where $||u||_{\mathfrak{H}}$ for a strongly measurable $u : X \to \mathcal{H}$ is defined as

$$||u||_{\mathfrak{H}} = \left(\int_X ||u(x)||^2_{\mathcal{H}}\, d\mu(x)\right)^{1/2}.$$

As usual, we identify functions which are almost everywhere the same. If, for the elements $x, y \in \mathfrak{H}$ we let

$$(x, y) = \int_X (x(t), y(t))_{\mathcal{H}} \, d\mu(t),$$

then this defines a scalar product which generates the norm $|| \bullet ||_{\mathfrak{H}}$.

By Theorem 2.110 of [2], $(\mathfrak{H}, || \cdot ||_{\mathfrak{H}})$ is a Banach space in which the set of simple functions is dense. Consequently, $\mathfrak{H}$ is a Hilbert space. Further, if X is a separable metric space, μ is a Radon measure and $\mathcal{H}$ is separable, then $\mathfrak{H}$ is separable, too. This will be the case when $\mathfrak{X}$ is the Lebesgue-space.

If $\mathfrak{H}$ is separable, then it is the direct integral of $\{\mathcal{H}\}_{x\in X}$ over $\mathfrak{X}$ in the sense of Definition 14.1.1 of Kadison and Ringrose (1986):

$$\mathfrak{H} = \int_X^{\oplus} \mathcal{H} \, d\mu.$$

For the rest part of this section, we assume that $\mathfrak{H}$ is separable.

Definition 5.1.3. An operator $T \in \mathcal{B}(\mathfrak{H})$ is said to be *decomposable* if there are operators $T_x \in \mathcal{B}(\mathcal{H})$ for $x \in X$ such that for each $a \in \mathfrak{H}$ we have

$$(Ta)(x) = T_x a(x) \qquad \text{for almost all } x \in X.$$

In this case, the system $\langle T_x : x \in X\rangle$ is called a *decomposition* of T and we write

$$T = \langle T_x : x \in X\rangle.$$

Components of the decomposition are almost everywhere unique.

In particular, the identity operator $I_{\mathfrak{H}} \in \mathcal{P}(\mathfrak{H})$ is decomposable with the decomposition

$$I_{\mathfrak{H}} = \langle I_{\mathcal{H}} : x \in X\rangle.$$

In a similar manner we can define for each $T \in \mathcal{B}(\mathcal{H})$ an operator $h(T)$ as follows:

$$h(T) = \langle T : x \in X\rangle$$

that is, for $a \in \mathfrak{H}$ the action of $h(T)$ is defined as

$$(h(T)a)\,(x) = Ta(x) \qquad \text{for all } x \in X.$$

It is not hard to see that $h(T) \in \mathcal{B}(\mathfrak{H})$ for all $T \in \mathcal{B}(\mathcal{H})$.

By Theorem 14.1.10 of Kadison and Ringrose (1986), the set $\mathcal{R} \subseteq \mathcal{B}(\mathfrak{H})$ of decomposable operators is a von Neumann algebra acting on $\mathfrak{H}$. Now the map h : $\mathcal{B}(\mathcal{H}) \hookrightarrow \mathcal{R}$ is a $*$-algebra embedding with $h[\mathcal{P}(\mathcal{H})] \subseteq \mathcal{P}(\mathcal{R})$ and, in particular, as a function $h : \mathcal{P}(\mathcal{M}) \to \mathcal{P}(\mathcal{R})$ it is a lattice embedding preserving $\perp$ as well.

Finally, we define a normal state ψ of $\mathcal{R}$ as follows:

$$\psi(T) = \int_X \phi(T_x) \, d\mu(x) \qquad \text{for all } T = \langle T_x : x \in X\rangle \in \mathcal{R}.$$

Then $(\mathcal{P}(\mathcal{R}), \psi)$ is an extension of $(\mathcal{P}(\mathcal{M}), \phi)$ in the sense of Definition 2.6. If $\mathfrak{X}$ is dense then so is $(\mathcal{P}\,(\mathcal{R}), \psi)$.

Acknowledgement

Research supported by the National Research, Development and Innovation Office, Hungary. K 115593. Zalan Gyenis was supported by the Premium Postdoctoral Grant of the Hungarian Academy of Sciences.

References

Araki, Huzihiro. 1999. *Mathematical Theory of Quantum Fields*. Oxford: Oxford University Press.

Fonseca, Irene, and Giovanni Leoni. 2007. *Modern Methods in the Calculus of Variations: Lp Spaces*. New York: Springer.

Gyenis, Balázs, and Miklós Rédei. 2004. "When Can Statistical Theories Be Causally Closed?" *Foundations of Physics* 34: 1285–1303. doi:10.1023/B:FOOP.0000044094.09861.12

Gyenis, Balázs, and Miklós Rédei. 2011. "Causal Completeness of General Probability Theories." In *Probabilities, Causes and Propensities in Physics*, edited by Mauricio Suarez, 157–171. London, New York: Springer.

Gyenis, Zalán, and Miklós Rédei. 2011. "Characterizing Common Cause Closed Probability Spaces." *Philosophy of Science* 78: 393–409. doi: 10.1086/660302

Gyenis, Zalán, and Miklós Rédei. 2014. "Atomicity and Causal Completeness." *Erkenntnis* 79: 437–451. doi: 10.1007/s10670–013–9456–1

Haag, Rudolf. 1992. *Local Quantum Physics: Fields, Particles, Algebras*. Berlin and New York: Springer.

Hofer-Szabó, Gábor, Miklós Rédei, and László E. Szabó. 1999. "On Reichenbach's Common Cause Principle and Reichenbach's Notion of Common Cause." *The British Journal for the Philosophy of Science* 50(3): 377–398.

Hofer-Szabó, Gábor, Miklós Rédei, and László E. Szabó. 2000. "Common Cause Completability of Classical and Quantum Probability Spaces." *International Journal of Theoretical Physics* 39(3): 913–919. doi:10.1023/A:1003643300514

Hofer-Szabó, Gábor, Miklós Rédei, and László E. Szabó. 2013. *The Principle of the Common Cause*. Cambridge: Cambridge University Press.

Horuzhy, Sergey Sergeevich. 1990. *Introduction to Algebraic Quantum Field Theory*. Dordrecht: Kluwer Academic Publishers.

Kadison, Richard V., and John Robert Ringrose. 1986. *Fundamentals of the Theory of Operator Algebras* (volume I and II). Orlando: Academic Press.

Kitajima, Yuichiro. 2008. "Reichenbach's Common Cause in an Atomless and Complete Orthomodular Lattice." *International Journal of Theoretical Physics* 47(2): 511–519. doi:10.1007/s10773–007–9475–2

Kitajima, Yuichiro, and Miklós Rédei. 2015. "Characterizing Common Cause Closedness of Quantum Probability Theories." *Studies in the History and Philosophy of Modern Physics* 52(B): 234–241. doi: 10.1016/j.shpsb.2015.08.003

Marczyk, Michał, and Leszek Wronski. 2015. "A Completion of the Causal Completability Problem." *The British Journal for the Philosophy of Science* 66(2): 307–326. doi: 10.1093/bjps/axt030

Rédei, Miklós. 1997. "Reichenbach's Common Cause Principle and Quantum Field Theory." *Foundations of Physics* 27(10): 1309–1321. doi: 10.1007/BF02551514

Rédei, Miklós. 2014. "Assessing the Status of the Common Cause Principle." In *New Directions in the Philosophy of Science*, edited by Maria Carla Galavotti, Dennis Dieks, Wenceslao J. Gonzalez, Stephan Hartmann, Thomas Uebel, and Marcel Weber, 433–442. Wien and New York: Springer.

Rédei, Miklós, and Stephen J. Summers. 2002. "Local Primitive Causality and the Common Cause Principle in Quantum Field Theory." *Foundations of Physics* 32(3): 335–355. doi:10.1023/A:1014869211488

Rédei, Miklós, and Stephen J. Summers. 2007. "Quantum Probability Theory." *Studies in History and Philosophy of Modern Physics* 38: 390–417.

Reichenbach, Hans. 1956. *The Direction of Time*. Los Angeles: University of California Press.

Salmon, Wesley C. 1978. "Why Ask 'Why?'?" *Proceedings and Addresses of the American Philosophical Association* 51: 683–705.

Salmon, Wesley C. 1984. *Scientific Explanation and the Causal Structure of the World*. Princeton: Princeton University Press.

Takesaki, Masamichi. 2003. *Theory of Operator Algebras*. New York: Springer.

Wronski, Leszek. 2010. "The Common Cause Principle. Explanation via Screeing off." PhD diss., Institute of Philosophy, Jagiellonian University, Cracow, Poland.

Wronski, Leszek. 2014. *Reichenbach's Paradise. Constructing the Realm of Probabilistic Common "Causes"*. Warsaw, Berlin: De Gruyter.

Aristidis Arageorgis
Technical University of Athens

Original Scientific Paper
UDK 115:530.145
530.1:164

SPACETIME AS A CAUSAL SET: UNIVERSE AS A GROWING BLOCK?

Abstract: *The causal set programme towards a quantum theory of gravity is situated* vis-à-vis *the long-standing debate between eternalism (block theory) and past-presentism or possibilism (growing block theory) in the philosophy of time. It is argued that despite "appearances" and declarations to the contrary, the programme does not side with growing block theorists when it comes to harboring a robust notion of Becoming – at least, not more than familiar relativistic theories on continuous spacetime manifolds. The problem stems mainly from the postulate of discrete general covariance – a requirement imposed upon the only fully worked out kind of dynamics for causal sets to date, a dynamics of a classical stochastic process.*

1. Introduction

The philosophy of time has long been troubled by the (at least, apparent) inhospitality of spacetime physics towards a robust notion of Becoming. The representation of a spacetime and of its material contents by means of 4-dimensional differentiable manifold hosting a family of geometric object fields seems biased *ab initio* in favor of the eternalist metaphysical thesis that Donald Williams ([1951] 1967) has dubbed "the doctrine of the manifold". Of course, some philosophers have tried to salvage a notion of Becoming, even in the less hospitable environments of relativistic worlds, either by weakening the philosophical notion (e.g., by localizing the present) or by dismantling some of the physics (e.g., by introducing a preferred inertial frame). Still, till recently, theory construction in physics has proceeded without paying much attention to incorporating a mechanism for realizing genuine Becoming.

A significant exception to this ethos in theoretical physics is the causal set programme (CSP) in quantum gravity, which made its first official appearance in 1987 (Bombelli et al. 1987). The CSP has two core postulates concerning the deep structure of spacetime: that it is *discrete* and that it can be interpreted with reference to *causal* concepts and principles.[1] In more detail, a spacetime as envisioned by classical general relativity is just a "coarse" approximation to a causal set, a locally finite partially ordered set. From a philosophical perspective, the CSP may be regarded as an outgrowth of the causal theories of time or

[1] Dowker (2005; 2006) and Wallden (2010) are three exceptionally helpful overviews.

spacetime – and, arguably, an outgrowth that overcomes significant hurdles of antecedent lines of attack.[2] Hence one could claim that the CSP deserves more attention by philosophers than the little it has attracted so far (Stachel 2006; Butterfield 2007; Earman 2008; Wüthrich 2012; Callender and Wüthrich 2014).

But I shall confine attention only to the question whether the CSP harbors a genuine notion of Becoming. In fact, the CSP promises to support a *dynamic* conception of time in a *growing block model of the universe*, similar in *basic* principle to the one championed by Charlie Dunbar Broad (1923) or the one defended by Michael Tooley (1997). According to this conception, referred to in the literature under the rubric *past-presentism* or *possibilism* (Savitt 2014), the totality of the existent state of affairs depends on time and at each time all past and present objects, events, etc. exist but not those in the future. Analogously, according to the CSP, a causal set grows via a stochastic process of addition ("birth") of new elements, while the addition of a new element should not be regarded as taking place *in* time but rather as *constituting* time. This suggests the image of a universe in which the "sum total of existence is always increasing" – to quote Broad (1923, 66–67). In fact, Rafael Sorkin (2007, 157n), one of the leading proponents of the CSP, has made an explicit reference to Broad's view of time and John Earman has pointed out that the CSP "promises to transmute Becoming from a piece of speculative metaphysics to one of naturalized metaphysics" (Earman 2008, 159). On the other hand, Jeremy Butterfield has, in a passing remark, deflated these prospects (Butterfield 2007, 859).

By developing Butterfield's remark, which itself rests on a technical point well known to the proponents of the CSP, into a fully articulated argument, I shall argue that the CSP, with its *classical* stochastic process dynamics, does *not* substantiate past-presentism – at least, not at a higher degree than the familiar relativistic theories on continuous (smooth) spacetime manifolds. The proviso "with its *classical* stochastic process dynamics" is crucial here because there is as yet considerable uncertainty as to what a *quantum* dynamics for causal sets will look like.

The plan of the paper is as follows. Section 2 sketches the philosophical issue. Section 3 reviews the basics of the CSP – in particular, those elements that are relevant to the argument I shall offer. Section 4 expounds the central argument. And Section 5 wraps it up by stating more fully and more rigorously the conclusion.

[2] As it is well known, the full geometric structure of a general relativistic spacetime cannot be recovered from its causal structure (viz. the structure induced by the order relation of connectibility of manifold points via timelike curves). The limitations for such a recovery were established by Malament (1977): roughly, the metrics of two general relativistic spacetimes that share the same causal structure may differ up to a conformal factor. (For a rigorous exposition, consult also Malament 2007, 270–271.) The CSP removes this "conformal ambiguity" by estimating the volume of a region in a general relativistic spacetime with the number of causal set elements "contained" in that region. The slogan is "geometry = order + number". For a more precise account, see Sorkin (2005). In this paper, philosophers will find explicitly stated philosophical motivations drawn from Riemann's views on physical geometry –views exploited later by Grünbaum (1973, 8–18) toward the thesis of the intrinsic metrical amorphousness of space and time continua– as well as references to Reichenbach and Robb.

2. World and Becoming

According to a common intuition, time "passes" and the history of the universe gradually unfolds with the coming-to-be of new events. Modern physics, however, appears to be hostile toward this "intuitive metaphysics" of time. The familiar spacetime theories, relativistic or not, have models of the type $\langle M, G_1, ..., G_n \rangle$, where M is a smooth 4-dimensional differentiable manifold and $G_1, ..., G_n$, for some natural number $n > 0$, are geometric object fields specifying the geometry and the matter-energy content of spacetime. The points of the manifold M, as spacetime locations of idealized possible events, are assumed to be given, "once and for all", from the "beginning" in all four dimensions.

This mode of construction of physical theories about space and time sides, at first blush at least, with the metaphysical thesis that Williams ([1951] 1967) named "the doctrine of the manifold". Here is an excerpt from another paper by Williams in which the "doctrine" is proclaimed:[3]

> What I advocate as 'the doctrine of the manifold,' ... is simply a philosophical acceptance, as an ultimate literal truth about the way things are in themselves, of the conception that nature, all there is, was, or will be, 'is' (tenselessly) spread out in a four dimensional scheme of location relations which intrinsically are exactly the same, and hence in principle commensurate, in all directions, but which happen to be differentiated, in our neighborhood at least, by the *de facto* pattern of the things and events in them – by the lie of the land, so to speak. We are all perfectly familiar with the fact that the prodigious difference of the vertical dimension of space, with its terrifying asymmetry of up and down, above and below, from all those comparatively indifferent directions we call horizontal, is not due to any intrinsic difference between vertical and horizontal distances but only to a certain characteristic complex of matter and force in our vicinity whose 'grain,' so to speak, runs one way and not the other. Just so, I argue, there is a somewhat more pervasive pattern of physical qualities and relations which constitutes the even more momentous oddity of the temporal direction, with its even more striking asymmetry of earlier and later, in contrast with all the so-called spatial directions (Williams 1965, 465).

The underlying philosophical debate is that between *static* and *dynamic conceptions of the world* and may be summarized thus.[4] On a static conception of the world, what states of affairs exist does *not* depend upon time. Accordingly, change is not conceived as change in the totality of existent states of affairs over time; it is conceived, rather, in terms of the possession, by some object, or by the world as a whole, of different properties at different times. By contrast, according to a dynamic conception of the world, what states of affairs exist *does* depend upon time. Consequently, the totality of existing states of affairs a given object participates in may differ from one time to another. Such a difference

[3] I owe the "discovery" of this particularly clear statement of the "doctrine of the manifold" to my reading of Savitt's (2002) – a paper devoted to unearthing a common ground between Williams's and Broad's later views.

[4] Here I rely heavily on Tooley 1997, 13–16.

is precisely what is conceived as change in the given object, as opposed to the mere possession by the object of different properties at different times. Similarly, change in the world as a whole is thought of, *not* in terms of the possession of different properties by different temporal slices of the world, but rather as a difference in the totality of states of affairs that exist as of different times.

Eternalists or *block theorists* ("Blockheads" in Earman's 2008 inspiring terminology) advocate a static conception of the world whereas *past-presentists* or *growing block theorists* (called "Broadheads" by Earman 2008) espouse a dynamic conception of the world. Eternalists, like Williams, believe that all past and future objects, events, etc. are as real as those of the present and that there neither were nor will be objects, events, etc. that do not exist now. By contrast, past-presentists, like Broad or Tooley, affirm the existence of all past or present objects, events, etc. but deny the existence of future objects, events, etc.[5] Here is a telling excerpt from Broad's *Scientific Thought*:

> When an event, which was present, becomes past, it does not change or lose any of the relations which it had before; it simply acquires in addition new relations which it *could* not have before, because the terms to which it now has these relations were then simply non-entities.
>
> It will be observed that such a theory as this accepts the reality of the present and the past, but holds that the future is nothing at all. Nothing has happened to the present by becoming past except that fresh slices of existence have been added to the total history of the world. The past is thus as real as the present. On the other hand, the essence of a present event is, not that it precedes future events, but there is quite literally *nothing* to which it has the relation of precedence. The sum total of existence is always increasing, and it is this which gives the time-series a sense as well as an order. A moment t is later that a moment t' if the sum total of existence at t includes the sum total of existence at t' together with something more (Broad 1923, 66–67).

Indeed, Broad distinguished three senses of the word 'change', with the third –the one he dubbed "*absolute Becoming*"– being the most fundamental:

> I think that we must recognise that the word "change" is used in three distinct senses, of which the third is the most fundamental. These are (i) Change in the attributes of things, as where the signal lamp changes from red to green; (ii) Change in events with respect to pastness, as where a certain event ceases to be present and moves into the more and more remote past; and (iii) Change from future to present. I have already given an analysis of the first two kinds of change. It is clear that they both depend on the third kind (Ibid., 67).

As already noted, a similar view, but with significant differences,[6] has been articulated recently by Tooley (1997). According to Tooley, the events in the world are connected via an asymmetric causal relation, with the causes producing their effects by "giving birth" to them, by "bringing them into existence". This

[5] Cf. Rea 2003, 247.

[6] Mainly as regards the relative priority of *tenseless* over *tensed* concepts and facts and the crucial role of causation.

conception of causation is combined with a substantivalist thesis according to which spacetime consists of points, thought of as contingent entities, with each spacetime point being the cause of birth of other spacetime points.

But I shall not dwell into more details of either Broad's or Tooley's views. I shall come directly to a well-known problem encountered by every attempt to implement a growing block model of the universe within classical general relativistic cosmology.[7] It appears plausible to require that in a spacetime manifold, Becoming be represented by a family of spacelike hypersurfaces, indexed by the values of a global time function. But some models $\langle \mathrm{M}, g_{ab}, T_{ab} \rangle$ of classical general relativity do not admit a global time function $t : M \rightarrow \square$,[8] while those that do admit one admit an infinity of them. And one might make a case to deal with the first horn of the problem, i.e., the non-existence of a global time function in some general relativistic spacetimes. A necessary and sufficient condition for a spacetime of general relativity to admit a global time function is that it be stably causal.[9] So one might confine attention to stably causal spacetimes and appeal to arguments that only such spacetimes are physically realizable.[10] But the horn of the "embarrassment of riches" is thornier. Past-presentism cannot tolerate such a "democracy". A global time function must be ontologically distinguished inasmuch as it demarcates "existence" from "non-existence" by providing the temporary locus of "absolute Becoming".

As Earman has pointed out, the CSP suggests a promising way to deal with this problem (Earman 2008, 149). It offers a "mechanism" generating an increase in what Broad would regard as the "sum total of existence" and *then* examines whether this increase corresponds to the accretion of layers of "now" indexed by a global time function. So let me turn immediately to the basics of the CSP.

[7] Earman presents this problem as a dilemma and canvasses the possible ways out, none of which appears trouble-free for the "Broadheads" (Earman 2008, 147–150). Earman's paper (2008) is the place to look for an admirably clear exploration of the prospects for a growing block model of the universe in both Newtonian and relativistic settings.

[8] A *global time function* in a time-orientable spacetime of general relativity $\langle M, g_{ab} \rangle$ is a differentiable map $t : M \rightarrow \square$ whose gradient $\nabla^a t$ is a past-directed timelike vector field so that t strictly increases along every future-directed timelike curve.

[9] A time-orientable general relativistic spacetime $\langle M, g_{ab} \rangle$ is called *stably causal* if and only if there exists on M a continuous, nonvanishing, timelike vector field t^a such that the spacetime $\langle M, \tilde{g}_{ab} \rangle$ with $\tilde{g}_{ab} = g_{ab} - t_a t_b$ does not exhibit closed timelike curves. That a time-orientable general relativistic spacetime admits a global time function (as defined in the preceding Note) just in case it is stably causal is the content of Theorem 8.2.2 in Wald 1984, 198. See also Earman 1995, 166.

[10] For example, Hawking and Ellis ([1973] 1989, 197) have argued that spacetimes that are not stably causal are not physically realizable along these lines: given that general relativity is expected to be the classical limit of a quantum theory of spacetime in which the metric does not have a definite "value" at each point, in order for a property of the spacetime to have physical significance it must be characterized by some "stability", i.e., must be as well a property of "nearby" spacetimes. Of course, such extrinsic justifications of causality conditions ultimately hinge upon *debatable* assumptions concerning putative "exotic possibilities" such as time travel. I shall not pursue this issue further. To start with, the reader may consult Earman 1995, ch. 6 and Arntzenius and Maudlin 2013.

2. Deep Structure of Spacetime: Causal Sets

Discreteness and *causal relatedness*, the two pillars of the CSP, are reflected in the very definition of a causal set. A *causal set* or *causet* is a locally finite, partially ordered set ("poset"). Explicitly, a causal set is a structure $\langle c, \prec \rangle$, where c is a set and $\prec$ is a binary relation on c satisfying the following conditions: (i) for every $x, y, z \in c$, if $x \prec y$ and $y \prec z$, then $x \prec z$ (*transitivity*); (ii) for each $x \in c$, it is not the case that $x \prec x$ (*irreflexivity*);[11] and (iii) for any $x, y, \in c$, $\{z \in c : x \prec z \prec y\}$ is a finite set (*local finiteness*). Clearly, given a causet $\langle c, \prec \rangle$, the relation $\preceq$ defined by

For all $x, y \in c$, $x \preceq y$ if and only if ($x \prec y$ or $x = y$)

is a partial order on c . All the causal sets $\langle c, \prec \rangle$ we shall be considering below in the context of classical sequential growth satisfy the stronger than local finiteness property of being *past finite*: for every $x \in c$, the set $\{y \in c : y \preceq x\}$ is finite. The statement '$x \prec z$' is expressed by several suggestive locutions borrowed from mathematical, physical or genealogical terminology: '*x* precedes *y*', '*y* follows *x*', '*x* is in the past of *y*', '*y* is in the future of *x*', '*x* is an ancestor of *y*', '*y* is a descendant of *x*', etc.

The core idea of the CSP is that the elements of a causal set may be thought of as events in a discretized spacetime and the order relation as a causal relation.[12] The basic assumption is that in the regime of very small scales (of the order of the Planck length $l_p = (G\hbar/c^3)^{1/2} = 10^{-33}$ cm) spacetime is no longer described by a semi-Riemannian metric on a smooth manifold, but by a causal set. To use an analogy suggested by Dowker (2005), just as ordinary matter appears to us "continuous" while in reality it consists of molecules and atoms, so spacetime appears to us "continuous" at large scales while in reality it is a causal set and the spacetime "continuum" of our experience is just an approximation to the underlying discrete ordered structure.

As an ordered set, every finite causal set can be represented graphically by a Hasse diagram. Draw a dot (vertex of a graph) to represent each element of the finite causal set and draw a line (edge of a graph) to represent each irreducible relation $x \prec y$,[13] with the preceding element x being represented by a dot drawn below the one representing the following element y. Figure 1 illustrates a very simple concrete example, with the elements of the causal set labeled by natural numbers.

[11] The expression '$x \nprec y$' is often used to abbreviate 'It is not the case that $x \prec y$'. Axiom (ii) for causal sets also goes by the name '*acyclicity*'.

[12] Philosophers should not read too much into the word 'causal' here. No worries should be triggered as to whether causation is indeed a transitive relation (see, e.g., Hitchcock 2001) or as to whether the intended use of the term captures a single robust notion of causation familiar from the philosophical tradition, if indeed there is one (see, e.g., Psillos 2007). One should have, rather, in mind the concept of causal relatedness or causal connectibility as it is employed in relativistic spacetime physics.

[13] The relation between two causet elements is said to be irreducible –in the jargon, a *link*– if and only if it is not implied by other such relations via transitivity.

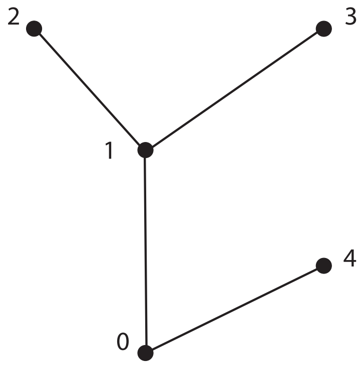

Figure 1. Hasse diagram of a labeled finite causal set

If you have trouble imagining the “spacetime continuum of our experience” to “arise” from a simple discrete structure like the one depicted in Figure 1, take into consideration that, according to rough estimates (Dowker 2006, 5), in the causal set underlying $1\text{cm}^3 \times 1\text{s}$ of the spacetime continuum you should expect some 10^{143} elements. (Not that such an estimate can really assist the imagination!)

Leaving aside flights of the imagination, I turn immediately to the tedious chore of reviewing a few basic definitions needed in the rest of the paper. Let c be a causal set. The *past of an element* $x \in c$ is the set of all elements that precede it in c, $\text{past}_c(x) = \{y \in c : y \prec x\}$. And an element $x \in c$ is called *maximal* if and only if it is to the past of no element – i.e., $x \notin \text{past}_c(y)$ for every $y \in c$. Let, now, a be a subset of c, $a \subseteq c$, . The *past of the subset a* of c is just the union of the pasts of its elements, $\text{past}_c(a) = \cup_{x \in a} \text{past}_c(x)$. Of course, a is said to be a *chain* just in case it is linearly ordered – i.e., any two distinct elements of a are related by the relation $\prec$ of precedence ($x \prec y$ or $y \prec x$ for any $x, y \in a$ with $x \neq y$). On the other hand, a is called a *stem* if and only if it is finite and contains its own past (past_c $(a) \subseteq a$).[14]

Now, as I have mentioned already, the exact final formulation of a *quantum* dynamics for causal sets is still an open question. But as a first step in the direction of formulating such a dynamics, Sorkin and collaborators have worked out a dynamics of a classical stochastic process in discrete stages, a *classical sequential growth* (CSG) dynamics.[15] The basic idea is this. Starting from the empty set, a causal set grows via a stochastic process at each stage of which a new element is added to the already existing causal set, with the process running to infinity. Accordingly, the elements of any causal set, be it finite or infinite, may be labeled by natural numbers. Specifically, a *natural labeling* of a causal set c

[14] To illustrate, if c is the simple 5-element causet depicted in Figure 1, then $\text{past}_c(0) = \emptyset$, past (2) {0,1}, past_c (4) = {0}, past_c ({1,3}) = {0,1}, past_c ({0,1}) = {0}, and so on. The elements 2, 3, 4 are the maximal elements of c. The subset {0, 1, 2} is a chain whereas {2, 4} is not. And the subsets {0, 1, 2}, {0, 1, 3}, {0, 4}, and {0, 1} are stems whereas {1, 2}, {0, 3}, and {4} are not. Note also that Rideout and Sorkin (1999, 2) use the term ‘*partial* stem’ for what is here called simply a stem.

[15] The dynamics I sketch in what follows is the one formulated by Rideout and Sorkin 1999.

is a bijective mapping of the initial segment of the set of natural numbers $\square = \{0,1,2,...\}$ whose cardinality equals that of c –hence, of $\square$ itself, if c is infinite– onto c that preserves order. That is, for a causal set c with n elements, a natural labeling is just a 1–1 and onto map

$$l : \{0,1,..., n-1\} \to c$$

such that for all $x, y \in c$, if $x \prec y$, then $l(x) < l(y)$. By convention, a natural labeling of a causet expresses an order of birth *we* attribute to its elements, but carries no intrinsic physical significance (see below).[16]

Obviously, a causal set may admit more than one labelings. A *labeled causal set* $\tilde{c} = \langle c, l_{\tilde{c}} \rangle$ is just a causal set, c, together with one of its labelings, $l_{\tilde{c}}$.[17] Given two labeled causal sets, $\tilde{c} = \langle c, l_{\tilde{c}} \rangle$ and $\tilde{b} = \langle b, l_{\tilde{b}} \rangle$, of the same cardinality, say k,[18] we shall say that $\tilde{c}$ and $\tilde{b}$ are k – *label variants* and we shall write $\tilde{c} \,\square_k\, \tilde{b}$ if and only if $l_{\tilde{b}} \circ l_{\tilde{c}}^{-1} : c \to b$ is an isomorphism with respect to the order relations in the causal sets c and b. In such a case, $\tilde{c}$ and $\tilde{b}$ are isomorphic ordered structures differing only in their labelings. Evidently, for each $k = 0,1,2,..., \infty$,[19] the binary relation $\square_k$ of k – label variance is an equivalence relation on the set of labeled causal sets with cardinality k.

The dynamical law in a CSG setting for causal sets reduces to an assignation of probability to each transition from every finite causal set to each causal set that can be formed from it by the addition of exactly one element. The *Rideout-Sorkin* (RS) *models* of such a dynamics are required to satisfy four postulates I shall briefly present next.

First is the requirement of *internal temporality*: each element of a causal set is born either to the future of, or unrelated to, all existing elements; no element can arise to the past of an existing element. Rideout and Sorkin phrase the rationale thus:

> The phenomenological passage of time is taken to be a manifestation of this continuing growth of the causet. Thus we do not think of the process as happening "in time," but rather as "constituting time" ... (Rideout and Sorkin 1999, 2).

Put another way, *physical* time is defined by the *intrinsic* ordering of a causal set.

The second postulate is the requirement of *discrete general covariance*: the net probability of formation of any particular finite causal set is independent of the order of birth we attribute to its elements (i.e., of each natural labeling of its

[16] Henceforth, I shall consider only natural (i.e., order-preserving) labelings of causets but I shall often drop, for brevity, the qualification 'natural'.

[17] A note on notation. I use small case letters from the beginning of the Latin alphabet (a, b, c,...), sometimes embellished with accents, to denote causets. For each $n \in \square$, $\Omega(n)$ is the set of all causets with n elements. $\Omega(\square) = \cup_{n \in \square} \Omega(n)$ is the set of all finite causets. And Ω is the set of all infinite "completed" causets, i.e., those that result "when" the sequential process of accretion of new elements "runs to completion". A tilde above the symbols for individual causets ($\tilde{a}$, $\tilde{b}$, $\tilde{c}$,...) and above the symbols for sets of causets ($\tilde{\Omega}(n)$, $\tilde{\Omega}(\square)$, $\tilde{\Omega}$) signifies natural labeling.

[18] Clearly, by the cardinality of a labeled causet $\tilde{c} = \langle c, l_{\tilde{c}} \rangle$ I just mean the cardinality of c.

[19] Mathematicians will forgive this notation and, in particular, my opting for '∞' instead of '$\aleph_0$'.

elements). Consider the set Ω(□) of all finite causal sets, partially ordered by the relation $\lhd$, where $b \lhd c$ just in case c can grow out of b via CSG.[20] And visualize the growth of a causal set in terms of directed paths in the graph of ⟨Ω(□),$\lhd$⟩ (a Hasse diagram of Hasse diagrams). Then discrete general covariance entails that any two paths with the same initial and final end points have the same product of transition probabilities. Figure 2 illustrates this in a simple example.[21]

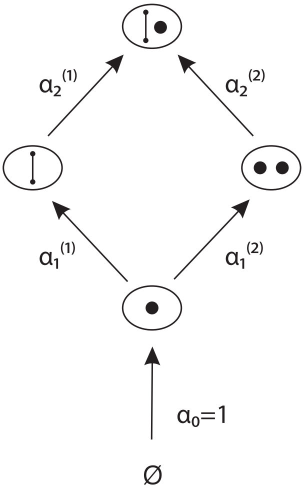

Figure 2. Discrete general covariance in CSG

CSG starts with the transition from the empty set to the 1-element causet – a transition assigned the probability value $\alpha_0 = 1$ because "the universe does exist". Thereafter there are two paths, indicated by the superscripts '(1)' and '(2)', that can lead to the formation of the 3-element causet depicted at the top, each involving the intermediate formation of a different 2-element causet. Discrete general covariance demands that the products of transition probabilities along these two paths be equal: $\alpha_1^{(1)}\alpha_2^{(1)} = \alpha_1^{(2)}\alpha_2^{(2)}$.

Thus a labeling of a causal set, even if natural, carries no physical significance. It stands for a kind of "external" or "extrinsic" time and introduces a *gauge element* into the theory because *we* use it to describe the dynamics of the growth of causal sets. The condition of discrete general covariance expresses what would here correspond to "invariance under gauge transformations". Indeed, the counterpart of a natural labeling of a causal set in a spacetime continuum would be a coordinate system $(x^i)_{i=0,1,2,3}$ whose first coordinate x^0 would be everywhere timelike so that it could serve to foliate spacetime into a family of spacelike hypersurfaces.[22] And, of course, physics is independent of the choice of coordinates.

[20] More precisely, for any b, $c \in$ Ω(□), $b \lhd c$ if and only for some $n \in$ □ there exist a_0, a_1, ..., $a_n \in$ Ω (□) such that a_0 is isomorphic to b, a_n is isomorphic to c, and for every $i \in \{0,1, ..., n\}$, a_{i+1} can be formed by accreting a single element to a_i (a_{i+1} is a "child" of a_i, to adopt the genealogical vocabulary dear to the proponents of the CSP).

[21] It is a mere reproduction of Figure 3 of Earman's (2008, 157).

[22] See also Rideout and Sorkin 1999, 2n3.

The third requirement imposed on the RS models for CSG dynamics implements the *classical* idea that events occurring in some “part” of a causal set are influenced only by that “portion” of the causal set lying to their past. It is the condition of *Bell causality*: the probability of a particular addition of a new element to a causal set depends only on the past of the new element and is not affected by elements in spacelike separation. The mathematical formulation provided by Rideout and Sorkin looks a bit convoluted at first blush (Rideout and Sorkin 1999, 6–7). It starts with a definition: for every $n \in$ □, the *precursor set* of the transition from $b \in \Omega(n)$ to $c \in \Omega(n+1)$ induced by the birth of a new element x is defined by

$$\text{precursor}(b \rightarrow c) = \text{past}_c(x) \subseteq b.$$

Bell causality demands that if for each $j \in \{1,2\}$, $c \rightarrow c_j$ is a transition from $c \in \Omega(n)$ to $c_j \in \Omega(n+1)$, then

$$\frac{\Pr(c \rightarrow c_1)}{\Pr(c \rightarrow c_2)} = \frac{\Pr(b \rightarrow b_1)}{\Pr(b \rightarrow b_2)},$$

where $b = \text{precursor}(c \rightarrow c_1) \cup \text{precursor}(c \rightarrow c_2)$ ($b \in \Omega(m)$ for some $m \leq n$) and b_j is the causal set resulting from b via the addition of a new element in the same manner as in the transition $c \rightarrow c_j$ ($b_j \in \Omega(m+1)$). Figure 3 offers a simple illustration: the diagram on the left depicts two possible transitions from a 3-element causet c to two different 4-element causets, c_1 and c_2, while the diagram on the right depicts the causets b, b_1, b_2 associated with these transitions as in the aforementioned stipulations.

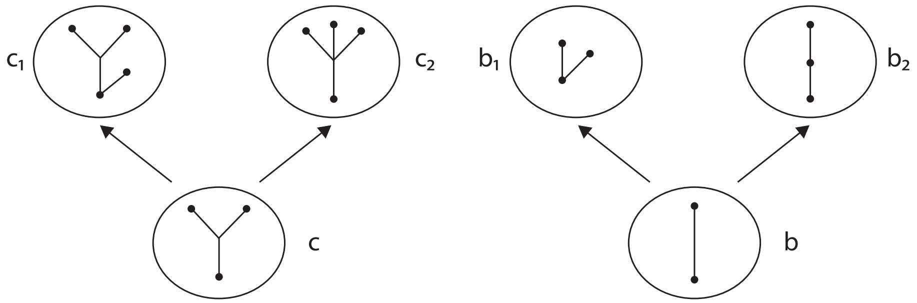

Figure 3. Bell causality in CSG

In this example, of course, the satisfaction of Bell causality amounts again to the satisfaction of relation (4).

The last requirement assumed to be satisfied by the RS models ensures that a CSG dynamics for causal sets be a Markov process.[23] It is the *Markov sum rule*: the sum of all transition probabilities issuing from a given causal set must be equal to 1.

[23] It might be helpful to recall here the slogan characterizing Markov processes: in a Markov process, the future and the past are independent if one conditionalizes on the present.

It has been shown that these four conditions are sufficiently restrictive to determine every transition probability from a causal set with n elements to a causal set with $n+1$ elements, for any $n\in$□, leaving just one free parameter for each stage of the stochastic process. Under the assumption that no transition probability vanishes,[24] one can derive that an arbitrary transition probability α_n from a member of $\Omega(n)$ to a member of $\Omega(n+1)$ may be computed by employing the formula

$$\alpha_n = \frac{\sum_{j=m}^{\varpi} \binom{\varpi - m}{\varpi - j} t_j}{\sum_{k=0}^{n} \binom{n}{k} t_k},$$

where ϖ stands for the cardinality of the precursor set of the transition, m is the number of maximal elements in the precursor set of the transition, and t_i ($i = 0,1,2,\ldots$) are free parameters satisfying $t_0 = 1$ and $t_i \geqslant 0$.

Needless to say, all of the above provide no more than a mere sketch of a central theme developed within the CSP. But I shall forego the discussion of other technical aspects of the CSP and turn to the exploration of its relation to growing block views of the universe.

3. General Covariance: Block Universe

Again, recall what the promise is: the CSP will provide a physical mechanism implementing a genuine notion of Becoming. Here is, in more detail, Earman's forceful way of putting the matter:

> Apart from the hoary philosophical debates about the reality of the future, the growing block model is given new life by the causal set approach to quantum gravity. By providing a physical mechanism for producing growth and Becoming, this approach promises to transmute Becoming from a piece of speculative metaphysics to one of naturalized metaphysics. Furthermore, the causal set approach avoids any appearance of parasitism on block models, since it makes classical relativistic spacetime an emergent feature of Becoming, something Broad would have no doubt applauded (Earman 2007, 159–160).

And, certainly, this assessment agrees not only with the imagery residing at the core of the CSP but also with declarations made by some of its proponents. I shall now argue that the enthusiasm all this may instigate in the hearts of growing block fans should be tempered, once all the consequences of the principle of discrete general covariance are unraveled.

In the context of classical general relativity on differentiable manifolds, the requirement of general covariance enforces diffeomorphism-invariance on

[24] Rideout and Sorkin (1999), whose approach and results I present here, imposed this assumption but Varadarajan and Rideout (2006) removed it. Of course, once the assumption of nonzero transition probabilities is removed, some of the physical postulates (notably, Bell causality) must be reformulated to make mathematical sense.

the dynamical models and coordinate-independence on the observables. We have mentioned already the way in which discrete general covariance restricts the stochastic dynamics of causal sets. What can we say about *observables*? Alternatively, what sort of *questions* is the available causal set theory expected to give answers to and how?

For comparison, let us start from a more familiar probabilistic physical theory – namely, quantum mechanics. The elementary questions quantum mechanics can give answers to have the form: 'Will a measurement of the observable O conducted on a quantum system in the state $|\Psi\rangle$ yield a result in the (measurable) set Δ of real numbers?' And the answers are probabilistic: 'Yes, with probability $\Pr_{|\Psi\rangle}(O, \Delta)$', with the probability value being computed courtesy of some variant of Born's statistical algorithm. But the causal set theory under discussion here appears to be a physical theory of a quite different ilk.[25] Its "empirical claims" –i.e., the claims associating (some of) its models with the world– are intended to pertain to the universe as a whole and the dynamics is supposed to describe possible entire histories of the universe (possible universes). Accordingly, the relevant elementary questions inquire as to whether a given causal set c belongs to *this* or *that* possible history. Such questions assume the general form: 'Will the CSG of the causal set c "end up" forming an infinite ("completed") causal set that satisfies the predicate Φ?' where 'Φ' stands for some predicate of interest. For example, the symbolic statement '$\Phi(a)$' may stand for 'a contains a 4-chain' or 'a contains an element whose ancestors and descendants jointly exhaust the remainder of a',[26] etc. And the answers are, again, of probabilistic nature: 'Yes, with probability $\mu(A)$', where A is that subset of Ω whose members satisfy the appropriate logical combination of predicates at hand.

However, we have not as yet shown how the probability measure μ is constructed. In essence, we have not as yet defined *mathematically* the relevant stochastic process. Indeed, the family $\{\alpha_n : n \in \square\}$ of transition probabilities for finite causal sets, given by (5) above, does not by itself constitute a stochastic process. Recall that, in general, a *stochastic process* is a family $\{X_t : t \in T\}$ of random variables on a probability space $\langle \Omega, \mathsf{R}, \mu \rangle$, indexed by (continuous or discrete) "time" t.[27] Furthermore, for the measure μ to be defined via Kolmogorov's theorem, *all* the finite-dimensional distributions $\mu_{t_1 \dots t_k}$, $\langle t_1, \dots, t_k \rangle \in T$, $k = 1,2,\dots$, are needed.[28]

[25] The differences are mitigated if quantum mechanics is construed along the lines of a suitable version of the consistent histories interpretation, favored, of course, by adherents to the CSP.

[26] In the jargon, such an element is called a *post*. From the cosmological viewpoint, whether the universe's causal set will develop a post amounts to whether the universe will recollapse.

[27] Ω is the *sample space*, R is a *σ-algebra* (*σ-field*) in Ω , and $\mu : \mathsf{R} \to \square$ is a *probability measure* on R. I shall not rehearse here the well-known definitions of these concepts or of the concept of a random variable.

[28] If $\{X_t : t \in T\}$ is a stochastic process in a probability space $\langle \Omega, \mathsf{R}, \mu \rangle$, then for each k-tuple $\langle t_1, \dots, t_k \rangle$ of distinct elements of T, the random vector $\langle X_{t_1}, \dots, X_{t_k} \rangle$ has, over $\square^k$, some distribution $\mu_{t_1 \dots t_k} : \mathfrak{R}^k \to \square : H \mapsto \mu(\{\omega \in \Omega : \langle \mathrm{X}_{t_1}(\omega), \dots, X_{t_k}(\omega) \rangle \in H\})$, where $\mathfrak{R}^k$ is the σ-algebra of k-dimensional Borel sets ($\mathfrak{R} = \mathfrak{R}^1$). The probability measures $\mu_{t_1 \dots t_k}$ are the *finite-dimensional distributions of the stochastic process*. Kolmogorov's existence theorem says, conversely,

It follows that in order to define with mathematical consistency the stochastic process representing CSG for causal sets, one should envision the birth of new elements to run *ad infinitum* as is, indeed, the case in RS universes. And this amounts to a first move of retreat *vis-à-vis* the block-universe idea as Sorkin himself readily concedes:

> In order to define [the measure] consistently, one must take [the sample space] to be a space of infinite causets, ones for which the growth process has "run to completion". We meet here with an echo of the block-universe idea, that is in effect built into mathematicians' formalisation of the concept of stochastic process (Sorkin 2007, 160n8).

Be this as it may, the proponents of the CSP have done all the necessary mathematical work.[29]

In short, the first step toward the construction of the relevant probability space has as follows. Take T = □ as the index set for "external" or "extrinsic" "time". Let the sample space be the set $\tilde{\Omega}$ of *labeled* infinite ("completed") causal sets. For each labeled finite causal set $\tilde{b}$ of cardinality card(b) = $n \in$ □, the *cylinder set with base* $\tilde{b}$ (of rank n) is just the set of all labeled infinite causal sets whose first n elements (those labeled 0, 1, ..., n – 1) form a causal set isomorphic to $\tilde{b}$ (with the same labeling) – symbolically:

cyl($\tilde{b}$) = {$\tilde{c} \in \tilde{\Omega}$: the first card(b) elements of $\tilde{c}$ form an isomorphic copy of $\tilde{b}$ with the same labeling}.

Accordingly, the σ-algebra $\tilde{\mathsf{R}}$ in $\tilde{\Omega}$ is defined as the σ-algebra generated by all the cylinder sets cyl($\tilde{b}$), $\tilde{b} \in \Omega($□$)$, i.e., as the smallest σ-algebra of subsets of $\tilde{\Omega}$ containing all these cylinder sets. It remains to define the appropriate probability measure on $\tilde{\mathsf{R}}$. The transition probabilities α_n, $n \in$ □, given by (5) above, determine the probability of formation of each labeled finite causal set (equal to the product of the transition probabilities corresponding to the individual births described by the labeling) and, consequently, a real function $\tilde{\mu}$ on the set of all cylinder sets. This function can, then, be extended to a probability measure $\tilde{\mu}$ on the entire σ-algebra $\tilde{\mathsf{R}}$ via the standard mathematical procedures

that if a given system of such measures satisfies two consistency conditions, then there exists a stochastic process having these finite dimensional distributions. More precisely, *if* T is an arbitrary set and to each k = 1, 2, ... and each $\langle t_1, \ldots, t_k\rangle \in T^k$ there corresponds a joint distribution function $\mu_{t_1\ldots t_k} : \mathfrak{R}^k \to$ □ so that all the $\mu_{t_1\ldots t_k}$ satisfy the following two conditions:

(a) For all $H_1, \ldots, H_k \in \mathfrak{R}$ and every permutation π of $\langle 1, 2, \ldots, k\rangle$,
$\mu_{t_1\ldots t_k}(H_1 \text{ x}\ldots\text{x } H_k) = \mu_{t_{\pi(1)}\ldots t_{\pi(k)}}(H_{\pi(1)} \text{ x}\ldots\text{x } H_{\pi(k)})$, and
(b) For all $H_1, \ldots, H_k \in \mathfrak{R}$,
$\mu_{t_1\ldots t_{k-1}}(H_1 \text{ x}\ldots\text{x } H_{k-1}) = \mu_{t_1\ldots t_{k-1}\, t_k}(H_1 \text{ x}\ldots\text{x } H_{k-1x} \text{ x } H_k)$,

then there exists on some probability space $\langle \Omega, \mathsf{R}, \mu\rangle$ a stochastic process $\{X_t : t \in T\}$ having the $\mu_{t_1\ldots t_k}$ as its finite-dimensional distributions. Billingsley (1995, ch. 7, sec. 36) is a standard reference for all these mathematical niceties.

[29] See Brightwell et al. 2002; 2003; as well as Dowker and Surya 2006.

underpinning Kolmogorov's theorem. Thus is obtained the probability space $\langle\tilde{\Omega}$, $\tilde{\mathsf{R}}$, $\tilde{\mu}\rangle$ pertaining to the labeled infinite ("completed") causal sets.[30]

Still, this approach is not "covariant".[31] The probability space $\langle\tilde{\Omega}$, $\tilde{\mathsf{R}}$, $\tilde{\mu}\rangle$ affords probabilistic answers to questions that are not covariant inasmuch as they implicate, for admitting a definite probabilistic answer, some particular labeling.[32] Yet, the postulate of discrete general covariance requires independence from labeling. The elements of a causal set are not intrinsically individuated and, consequently, for each labeled causal set the only representation endowed with physical significance is the isomorphism equivalence class it belongs to.

The switch to a covariant probability space is brought about in the following manner. Take the set Ω of *unlabeled* infinite ("completed") causal sets for sample space. Consider the relation $\Box_\infty$ of ∞-label variance in $\tilde{\Omega}$ discussed in the previous Section and deem a subset A of $\tilde{\Omega}$ to be *covariant* just in case along with each of its members, A contains all members of $\tilde{\Omega}$ equivalent (isomorphic) to it: for all $\tilde{c}, \tilde{c}' \in \tilde{\Omega}$, if $\tilde{c} \in A$ and $\tilde{c} \Box_\infty \tilde{c}'$, then $\tilde{c}' \in A$. Let R be the collection of all subsets A of $\tilde{\Omega}$ that are both $\tilde{\mu}$-measurable ($A \in \tilde{\mathsf{R}}$) and covariant in the aforementioned sense. Note that each $A \in \mathsf{R}$, although "originally" a subset of $\tilde{\Omega}$, it may also be regarded as a subset of Ω because it is relabeling invariant. Further, it is not hard to prove that R constitutes a σ-algebra of subsets of $\tilde{\Omega}$ (and of Ω) and, indeed, a *sub-σ-algebra* of $\tilde{\mathsf{R}}$ since clearly $\mathsf{R} \subseteq \tilde{\mathsf{R}}$. Therefore, the restriction of the measure $\tilde{\mu}$ to R yields a measure $\mu = \tilde{\mu}|\mathsf{R}$ on the measurable space $\langle\Omega, \mathsf{R}\rangle$ for unlabeled infinite ("completed") causal sets (i.e., $\mu(A) = \tilde{\mu}(A)$ for every $A \in \mathsf{R}$). Thus we get the covariant probability space $\langle\Omega, \mathsf{R}, \mu\rangle$ for the RS models of CSG – "covariant" in the sense that the members of R correspond exactly to the physical meaningful, according to discrete general covariance, questions the dynamics provides probabilistic answers to by means of μ.[33]

However, the *physical significance* of the members of the covariant σ-algebra R is not clear. The question or predicate an arbitrary $A \in \mathsf{R}$ corresponds to cannot,

[30] The relevant stochastic process involves the family $\{X_n : n \in \Box\}$ –recall $T = \Box$ – of random variables that are defined in this fashion. Identify each $\tilde{c} \in \tilde{\Omega}$ with the sequence $\langle\tilde{c}_0, \tilde{c}_1, \tilde{c}_2, ..., \tilde{c}_n, ...\rangle$, $\tilde{c}_n \in \tilde{\Omega}(n)$, $n \in \Box$, representing the successive stages of its formation out of the empty set according to its labeling ($\tilde{c}_0 = \varnothing$ and $\tilde{c}_1$ is the singleton comprising the element labeled 0). Then, for each $n \in \Box$, set $X_n : \tilde{\Omega} \rightarrow \tilde{\Omega}(\Box) : \tilde{c} \mapsto \tilde{c}_n$.

[31] Henceforth, I shall unashamedly apply the predicate 'covariant' to questions, σ-algebras, probability spaces, etc., hoping that what is meant in each case becomes manifest from what I write in the context.

[32] 'Does $\tilde{c}$ comprise a chain before the 13th stage of its growth?' is an example of such a question.

[33] Alternatively, we could construct the *quotient space* $\langle\tilde{\Omega}/\Box$, $\tilde{\mathsf{R}}/\Box$, $\tilde{\mu}/\Box\rangle$ of the probability space $\langle\tilde{\Omega}$, $\tilde{\mathsf{R}}$, $\tilde{\mu}\rangle$ with respect to the equivalence relation $\Box$. (Here I write '$\Box$' instead of '$\Box_\infty$' to reduce clutter in the notation.) For each $\tilde{c} \in \tilde{\Omega}$, let $\tilde{c}/\Box$ be the equivalence class of $\tilde{c}$ with respect to $\Box$, i.e., the set of all $\tilde{b} \in \tilde{\Omega}$ such that $\tilde{b} \Box \tilde{c}$. Identify $\tilde{\Omega}/\Box$ with the set $\{\tilde{c}/\Box : \tilde{c} \in \tilde{\Omega}\}$ and consider the function $p : \tilde{\Omega} \rightarrow \tilde{\Omega}/\Box : \tilde{c} \mapsto p(c) = \tilde{c}/\Box$. Define $\tilde{\mathsf{R}}/\Box$ by stipulating that for every $A \subseteq \tilde{\Omega}/\Box$, $A \in \tilde{\mathsf{R}}/\Box$ if and only if $p^{-1}[A] \in \tilde{\mathsf{R}}$, where $p^{-1}[A] = \{\tilde{c} \in \tilde{\Omega} : p(\tilde{c}) \in A\}$. Lastly, define the measure $\tilde{\mu}/\Box$ on the quotient measurable space $\langle\tilde{\Omega}/\Box$, $\tilde{\mathsf{R}}/\Box\rangle$ by demanding that $(\tilde{\mu}/\Box)(A) = \tilde{\mu}(p^{-1}[A])$ for each $A \in \tilde{\mathsf{R}}/\Box$. As expected, $\langle\Omega, \mathsf{R}, \mu\rangle$ can be naturally identified with $\langle\tilde{\Omega}/\Box$, $\tilde{\mathsf{R}}/\Box$, $\tilde{\mu}/\Box\rangle$ via the identification of each member c of Ω with the member $\tilde{c}/\Box$ of $\tilde{\Omega}/\Box$ where $\tilde{c} = \langle c, l_{\tilde{c}}\rangle$ with $l_{\tilde{c}}$ some labeling of c.

in general, be couched in terms familiar to the available theory describing causal set structure and growth. Yet, some members of R do have an apparent physical significance. These are the so-called "stem sets" defined by arbitrary unlabeled finite causal sets. For any $b \in \Omega(\square)$, define the *stem set of b* by

$$\text{stem}(b) = \{c \in \Omega : c \text{ contains a stem isomorphic to } b\}$$

That is, stem(b) comprises exactly those unlabeled infinite ("completed") causal sets for each of which there exists a natural labeling such that the first card(b) elements form a causal set isomorphic to b. That $\text{stem}(b) \in \mathsf{R}$, for every $b \in \Omega(\square)$, follows from the fact that stem(b) (viewed as a subset of $\tilde{\Omega}$) is a countable union of cylinder sets:

$$\text{stem}(b) = \cup\{\text{cyl}(\tilde{c}) : \tilde{c} \in \tilde{\Omega}(\square) \text{ and } b \text{ is a stem in } \tilde{c}\}.$$

Clearly, for each $b \in \Omega(\square)$, the associated member stem(b) of R corresponds to a question whose physical meaning is evident: for every $c \in \Omega$, the question '$c \in \text{stem}(b)$?' just means 'Does c contain a stem isomorphic to b?' And the answer afforded by the theory is, of course, probabilistic: 'Yes, with probability $\mu(\text{stem}(b))$'.

But are all infinite ("completed") causal sets characterized by their stems? The *exact* mathematical answer is "*Almost, yes!*" Here is how this answer is substantiated. Call an infinite ("completed") causal set a "rogue" if and only if there exists another infinite ("completed") causal set that is not isomorphic to the former although it shares with it exactly the same stems: $c \in \Omega$ is a *rogue* if and only if there exists a $c' \in \Omega$ such that c' is not isomorphic to c but for every $b \in \Omega(\square)$, $c \in \text{stem}(b)$ just in case $c' \in \text{stem}(b)$. Put the other way around, every infinite ("completed") causal set that is *not* a rogue is characterized up to isomorphism by its stems. Now, a remarkable theorem demonstrated by Brightwell et al. (2003, 4, Proposition 1) affirms that in any CSG dynamics, the set Θ of all rogues in Ω has measure zero, $\mu(\Theta) = 0$. Therefore *almost* every infinite ("completed") causal set produced by a CSG dynamics is characterized up to isomorphism by its stems.

The upshot of all this can be stated thus. With the exception of a set of infinite ("completed") causets of measure zero, all that can be meaningfully said about an infinite ("completed") causet may, once appropriately analyzed, be expressed by reference to its stems. In this sense, the stem questions of the form '$c \in \text{stem}(b)$?', with $c \in \Omega$ and $b \in \Omega(\square)$, "virtually" exhaust the set of *covariant* and *physically meaningful* elementary questions a causal set cosmology built on CSG is expected to tackle.

I shall now argue that precisely this aspect of CSG prohibits the causal set cosmology coupled to it from buttressing a robust notion of Becoming consonant with a growing block model of the universe. My argument rests on the following philosophical presuppositions.[34] A physical theory hosts a genuine notion of

[34] The argument is tailored after the so-called "truthmaker" or "grounding" objection(s) against presentism in the philosophy of time. Cf. Crisp 2003, 236–242; Rea 2003, 261–268.

Becoming, consonant with a growing block model of the universe, *only if* it distinguishes "stages of Becoming" so that at each such stage the entire history of the world is divided into a part that "has already become" or "is already definite" and a part that "has not as yet become" but is (temporarily) "indefinite". And if *truth supervenes on being*, as David Lewis (2001) has suggested, such a theory should accord physical significance to propositions whose truth-values supervene on states of affairs that have already become as of some such stage. Thus we arrive at the following necessary condition, I dub [N] for future reference.

> CONDITION [N]. *A physical theory T supports a notion of Becoming consonant with a growing block model of the universe* only if *T posits a family* $\{t_i : i \in I\}$ *of stages and assigns physical significance to at least one proposition p such that for some stage* t_i *the following holds: for all possible according to T worlds w and w', if p is true in w but false in w', then w and w' disagree with respect to facts up to* t_i.

The question we have to address now concerns the way in which "stages of Becoming" may be represented in a CSG dynamics. Of course, as causal sets are intended to portray *discrete* spacetimes, these stages should admit □ as index set. But what sort of structure in a growing causal set is going to represent the temporary locus of Becoming at the n^{th} stage for $n \in$ □? There appear only two choices. According to the first, call it "Choice 1", this structure may include (for sufficiently large *n*) more than one causet elements, none of which is an ancestor or descendant of another. According to the second, "Choice 2", the temporary locus of Becoming at each stage contains exactly one causet element. Intuitively, Choice 1 is intended to salvage, in a discrete spacetime, a surrogate for "hupersurface Becoming" in continuous spacetimes, whereas Choice 2 allies with the idea of "localizing" Becoming and the present. Neither choice supports the thesis that the CSP harbors a notion of Becoming that is both *genuine* by the standards of growing block theorists and *novel* in the sense that it has not been put forth in the context of relativistic theories of spacetime continua. Or so I shall argue.

Take up Choice 1 first. The most plausible way to spell it out mathematically seems to be this. Consider any infinite ("completed") causet *c*. Define the *level* of an element *x* in *c* as the maximum length of a chain in *c* with top element *x*.[35] Given that causets are past finite, every element in *c* has some finite level. For each $n \in$ □, define the n^{th} *level* of *c* as the set of all elements of *c* of level *n* and recognize it as a single stage of Becoming in the CSG producing *c*.[36] Along this line, the part of *c* that has already become as of a given stage comprises exactly those elements of *c* whose level is less than or equal to *n*, for some $n \in$ □. Let $c_{(n)}$

[35] Definitions and notation are borrowed from Brightwell et al. 2003, 4.

[36] Sure enough, the identification of the n^{th} level in a causet with the, say, n^{th} "stage of Becoming" in that causet's growth is somewhat arbitrary. However, not much hinges on this identification. The argument goes through as long as, for any causet $c \in \Omega$, what is taken to be the part of *c* that has already become as of any given stage has infinite complement with respect to *c*. Note, in addition, that there is no claim here that the levels of a causet approximated by a general relativistic spacetime will be mapped onto time slices of the embedding spacetime which correspond to sharp values of some global time function.

denote the thus demarcated subset of c – namely, the subset of all $x \in c$ with the property that every chain ending at x has length at most n (i.e., has at most n + 1 elements):

$$\max \{k \in \square : \exists x_0, x_1, \ldots, x_k \in c \text{ with } x_0 \prec x_1 \prec \ldots \prec x_k \text{ and } x_k = x\} \leq n.$$

Intuitively, $c_{(n)}$ is supposed to represent *past and present* as of a given stage of the growth of c, i.e., the cumulative accretion of events, or Broad's "sum total of existence", up to that stage.

Recall now that in causal set cosmology based on a CSG dynamics, for almost every world c (viz. for every $c \in \Omega \setminus \Theta$), the only propositions of physical significance that may be asserted about c can be analyzed into propositions of the form '$c \in \text{stem}(b)$', $b \in \Omega(\square)$. Moreover, no physical facts can distinguish between isomorphic causet worlds. Accordingly, on the above approach to expounding Choice 1, the necessary condition [N] is transcribed thus: *there exist* $b \in \Omega(\square)$ *and* $n \in \square$ *such that for every* $c, c' \in \Omega \setminus \Theta$, *if* $c \in \text{stem}(b)$ *but* $c' \notin \text{stem}(b)$, *then* $c_{(n)}$ *is not isomorphic to* $c'_{(n)}$. But, clearly, *this* is false! A stem that has not appeared as of any given stage *may* appear later on and, consequently, for every $b \in \Omega(\square)$ and every $n \in \square$ there exist $c, c' \in \Omega \setminus \Theta$ such that $c \in \text{stem}(b)$ and $c' \notin \text{stem}(b)$ *and* $c_{(n)}$ is isomorphic to $c'_{(n)}$.

The philosophical moral to be drawn is that, on this approach, causal set cosmology does not support a notion of Becoming consonant with a growing block model of the universe, since it does not bestow physical significance to any proposition referring exclusively to what might count as past and present. The crux of the issue is this, as the proponents of the CSP have acknowledged.[37] For a stochastic process that takes place against a non-dynamical temporal background, the predicates corresponding to measurable sets can be thought of as logical combinations of simpler predicates whose attribution can be decided in finite time. In the case of the stochastic growth of causal sets, this is true for the cylinder sets, which, however, are devoid of physical significance as they are not covariant. The closest covariant surrogates for the cylinder sets are the stem sets. But the attribution of the predicate corresponding to a stem set, even though it may be verified in finite time, it can be strictly falsified only in the limit of infinite number of stages of growth.

So let us turn to Choice 2. According to it, the temporary locus of Becoming at each stage of growth of a causet contains exactly one element; while the partial order relation $\preceq$ between causet elements admits the interpretation conferred by the locution 'x has already become (is already definite) as of y'. The relation $\preceq$ possesses the formal properties required for this interpretation: it is (i) *reflexive* ("every event has already become as of itself"), (ii) *transitive* ("for all events x, y and z, if x has already become as of y and y has already become as of z, then x has already become as of z"), and (iii) *non-universal* ("for every event x, there is at least one event that has not already become as of x").[38]

[37] Cf. Brightwell et al. 2002, 13–14.

[38] For the rationale underpinning these requirements, see also Stein 1991, 148. That $\preceq$ on causets fulfills the requirement of non-universality follows from the trivial fact that every $c \in \Omega$ satisfies the condition: for each $x \in c$ there exists a $y \in c$ such that $y \preceq x$ does *not* hold.

It is this conception of Becoming that has been explicitly advocated by some proponents of the CSP. Dowker, for example, in connection with the way in which the RS models can deal with "the problem of Now", has affirmed:

> There is growth and change. Things happen! But the general covariance means that the physical order in which they happen is a *partial* order, not a total order. This doesn't give any physical significance to a *universal* Now, but rather to events, to a Here-and-Now.
>
> I am not claiming that this picture of accumulating events (which will have to be reassessed in the quantum theory) would *explain* why we experience time passing, but it is more compatible with our experience than the Block Universe view (Dowker 2005, 458).

And, in a similar vein, she concluded one of her lectures with the following points:[39]

> "Becoming" and lack of a global time peacefully co-exist in these models. Things happen, but in a partial order.
>
> In a Sequential Growth Model, the causal past of any newly born element is real: reality accumulates, like sediment, with the events of the past fixed and unchanging and the future as yet unrealised potentiality.

But now there are two qualms one may justifiably have regarding the assertion that the CSP gives *new life* to the growing block view of the universe. First, *this* notion of Becoming is not novel: one can trace it in philosophical treatments of "good old" special relativity on Minkowski spacetime. And second, one may worry whether a partial ordering of events together with a solipsist view of each event's present is too weak a base to underpin a robust notion of Becoming that makes absolutely no concessions to eternalist views. I shall bracket here the second kind of concerns as they have been voiced and argued, for and against, in an extensive literature on the metaphysics of special and general relativity.[40]

Let me just go over the basic components of the relevant here conception of Becoming in the context of special relativity.[41] First, it involves the intention to graft the explication of the terms 'present' and 'temporal Becoming' on the intrinsic geometry of Minkowski spacetime. To this end, Stein (1991) demonstrated that the only plausible two-place relation '*Rxy*' between points of Minkowski spacetime that (i) exhibits the necessary formal properties (reflexivity, transitivity, non-universality) to admit the interpretation '*x* has already become as of *y*', and (ii) is invariant under automorphisms preserving time-orientation, is

[39] The referred to talk by Fay Dowker bears the title "Discrete spacetime: Things happen, they just happen in a partial order" and the slides for it are available in the Internet (Dowker 2015).

[40] Savitt's (2014) provides an overview and parts II and III of Dieks's (2006) are devoted exactly to the possibility of reconciling temporal Becoming with relativistic spacetime theories.

[41] See Stein 1968; 1991; Dieks 2006.

the relation '*x* lies in or on the past light cone of *y*'.[42] Second, on this approach, "in Einstein-Minkowski space-time *an event's present is constituted by itself alone*" (Stein 1968, 15). And, third, the process of Becoming is *local*, not only in the sense that it reduces to the happenings of individual local events, but also in the sense that the temporal relations in the network of these happenings arise from a partial ordering of non-global nature (Dieks 2006, 173). Becoming is just the successive happening of events along a timelike world line.

It is not my intention to defend (*or* criticize) this conception of Becoming in relativistic physics. I cannot resist, however, quoting one of Stein's witty remarks to its defense:

> At any rate, it is clearly a fact about the historical etymology of our language that the original meaning of the word "present" was not *now*, but *here-now*... That remains a current usage: When a soldier at roll call responds "Present!" upon hearing his name, he is not merely announcing that he still exists; he means that he is on the spot (Stein 1991, 159).

Still, my point is that the CSP has not as yet offered a conception of Becoming in physics that is not already familiar from philosophical discussions of relativistic theories on smooth spacetime manifolds. As a last piece of "textual evidence", let me cite an excerpt from Dieks's conclusions:

> So the natural view is that the history of our universe is realized by events that come into being; and that they come into being after and before each other as dictated by the partial ordering relation induced by the spacetime structure. According to this proposal the life of the universe is not one linear series of events, but a partially ordered set of events (Dieks 2006, 172–173).

The similarity with Dowker's views cited above is evident.

4. Conclusion

The CSP constitutes a vigorous approach towards a quantum theory of gravity that deserves more attention from philosophers than it has attracted so far. And one can argue in support of this claim even irrespectively of one's expectations as to whether the CSP will eventually succeed in "uncovering the Holy Grail", in producing a satisfactory theory of quantum gravity. At any rate, the current state of theoretical physics offers very feeble evidence to ground such expectations, not only for the CSP, but also for programmes that are more popular among physicists and seem to have attained a higher level of maturity, like the string theory programme or the loop quantum gravity programme.

What makes the CSP of particular interest to philosophers is that its development has been accompanied, if not motivated, by the explicitly stated

[42] For the exact statement of the result, see the theorem on p. 149 of Stein's (1991). It should be mentioned here that, as Callender and Wüthrich (2014) have shown, causal sets violate what would be an analogue of Stein's theorem, without, however, salvaging a notion of the "present" consistent with Lorentz invariance.

desire to respond to some challenges with venerable history in the philosophical tradition.[43] I tried to review and assess the CSP's response to one such philosophical challenge – namely, the challenge to incorporate a dynamic conception of the world, in harmony with a growing block model of the universe, in a (relativistic) physical theory about spacetime.

I argued that the CSP, at its present stage of development, does not meet the challenge for two reasons. First, the only kind of dynamics for causal sets that has been fully elaborated to date, the CSG type of dynamics, cannot be defined with mathematical consistency but in the limit of infinite time "when" causal set growth "has reached completion". This is due to the fact the relevant dynamical law amounts to the specification of the probability measure of a classical stochastic process. Still, this makes the growing block imagery the CSP aspires to flesh out within physics parasitic on a block universe view.

The second reason does not stem from some demand for mathematical consistency but from a physical principle lying at the core of the CSP, discrete general covariance. As I tried to show, this principle deprives the CSP from the conceptual resources to ground truths about the past or the present – resources that are plausibly required of any theory that aims to buttress the metaphysics of past-presentism. The only way out for a proponent of the CSP, who wishes to cling to a notion of Becoming that salvages a dynamic conception of the world, seems to be the stratagem of localizing Becoming and the present. But this move, whether on the right track or not, is not novel: it has been proposed and debated in the context of philosophical attempts to trace a viable notion of Becoming within relativistic spacetime theories on continuous (smooth) manifolds.

As I have mentioned right from the start, both of these obstacles to salvaging a robust notion of Becoming are acknowledged by the proponents of the CSP as well as by philosophers that have commented upon the issue. Why does, then, the CSP continue to retain its appeal *vis-à-vis* prospects of maintaining some growing block view of the universe? I see only two interrelated reasons for this (other than the obvious one that any metaphysical-ontological interpretations of the theories formulated within the CSP are apt to differ radically once causal sets are endowed with a *quantum* dynamics). The first has to do with the postulated *discreteness* of the deep structure of spacetime. In a discrete spacetime, there is not only "before" and "after", but also "previous" and "next": an event does not only have a "past" and a "future", but also a "predecessor" and a "successor". And this "animates" the motion of "now" along a timelike world line. The second reason has been recently pointed out by Callender and Wüthrich (2014): causal sets do exhibit a kind of *gauge-invariant growth* (measured by the *number* of elements). But this growth can be taken to graft a notion of temporal Becoming only at the expense of sacrificing various deep-seated intuitions and philosophical convictions.[44]

[43] One may adduce, as further reasons, the clarity and parsimony of the CSP's guiding physical principles as well as of its mathematics.

[44] For the ramifications, see Callender and Wüthrich 2014.

References

Arntzenius, Frank, and Tim Maudlin. 2013. "Time Travel and Modern Physics." In *The Stanford Encyclopedia of Philosophy* (Winter 2013 Edition), edited by Edward N. Zalta. Stanford: Metaphysics Research Lab, Center for the Study of Language and Information, Stanford University. https://plato.stanford.edu/archives/win2013/entries/time-travel-phys

Billingsley, Patrick. 1995. *Probability and Measure.* 3rd edition. New York: John Wiley & Sons.

Bombelli, Luca, Joohan Lee, David Meyer, and Rafael D. Sorkin. 1987. "Spacetime as a Causal Set." *Physical Review Letters* 59(5): 521–524. doi: 10.1103/PhysRevLett.59.521

Broad, C. D. 1923. *Scientific Thought*. New York: Harcourt, Brace & Co.

Brightwell, Graham, Fay Dowker, Raquel S. García, Joe Henson, and Rafael D. Sorkin. 2002. "General Covariance and the 'Problem of Time' in a Discrete Cosmology." In *Correlations. Proceedings of the Alternative Natural Philosophy Association 23 Conference* (August 16–21, 2001, Cambridge, England), edited by Keith G. Bowden, 1–17. London: ANPA.

Brightwell, Graham, Fay Dowker, Raquel S. García, Joe Henson, and Rafael D. Sorkin. 2003 "'Observables' in Causal Set Cosmology." *Physical Review D* 67: 084031–1–8. doi: 10.1103/PhysRevD.67.084031

Butterfield, Jeremy. 2007. "Stochastic Einstein Causality Revisited." *British Journal for the Philosophy of Science* 58(4): 805–867. doi: 10.1093/bjps/axm034

Callender, Craig, and Christian Wüthrich. 2014. "What Becomes of a Causal Set." Manuscript.

Crisp, Thomas M. 2003. "Presentism." In *The Oxford Handbook of Metaphysics*, edited by Michael J. Loux and Dean W. Zimmerman, 211–245. Oxford: Oxford University Press.

Dieks, Dennis. 2006. "Becoming, Relativity and Locality." In *The Ontology of Spacetime. Volume 1*, edited by Dennis Dieks, 157–176. Amsterdam: Elsevier.

Dieks, Dennis, ed. 2006. *The Ontology of Spacetime. Volume 1.* Amsterdam: Elsevier.

Dowker, Fay. 2005. "Causal Sets and the Deep Structure of Spacetime." In *100 Years of Relativity – Spacetime Structure: Einstein and Beyond*, edited by Abhay Ashtekar, 445–464. New York: World Scientific.

Dowker, Fay. 2006. "Causal Sets as Discrete Spacetime." *Contemporary Physics* 47(1): 1–9. doi: 10.1080/17445760500356833

Dowker, Fey. 2015. "Discrete Spacetime: Things Happen, They just Happen in a Partial Order." *Philosophy and the Sciences at the University of Manchester*, May 18. http://www.cicada.manchester.ac.uk/events/workshops/graphs-asynchronous-systems/slides-2.pdf

Dowker, Fay, and Sumati Surya. 2006. "Observables in Extended Percolation Models of Causal Set Cosmology." *Classical and Quantum Gravity* 23(4): 1381–1390. doi: 10.1088/0264–9381/23/4/018

Earman, John. 1995. *Bangs, Crunches, Whimpers, and Shrieks: Singularities and Acausalities in Relativistic Spacetimes*. Oxford: Oxford University Press.

Earman, John. 2008. "Reassessing the Prospects for a Growing Block Model of the Universe." *International Studies in the Philosophy of Science* 22(2): 135–164. doi: 10.1080/02698590802496680

Grünbaum, Adolf. 1973. *Philosophical Problems of Space and Time*. 2nd enlarged edition. Dordrecht: D. Reidel.

Hawking, Stephen W., and George F. R. Ellis. [1973] 1989. *The Large Scale Structure of Space-Time*. 8th reprint. Cambridge: Cambridge University Press.

Hitchcock, Christopher. 2001. "The Intransitivity of Causation Revealed in Equations and Graphs." *Journal of Philosophy* 98(6): 273–299. doi: jphil200198614

Lewis, David. 2001. "Truthmaking and Difference-making." *Noûs* 35(4): 602–615. doi: 10.1111/0029–4624.00354

Malament, David B. 1977. "The Class of Continuous Timelike Curves Determines the Topology of Spacetime." *Journal of Mathematical Physics* 18(7): 1399–1404. doi: 10.1063/1.523436

Malament, David B. 2007. "Classical Relativity Theory." In *Philosophy of Physics*, edited by Jeremy Butterfield and John Earman, 229–273. Part A. Amsterdam: Elsevier.

Psillos, Stathis. 2007. "What Is Causation?" In *Episteme Reviews: Research Trends in Science, Technology and Mathematics Education*, edited by Beena Choksi and Chitra Natarajan, 11–29. Bangalore: Macmillan India.

Rea, Michael C. 2003. "Four-dimensionalism." In *The Oxford Handbook of Metaphysics*, edited by Michael J. Loux and Dean W. Zimmerman, 246–280. Oxford: Oxford University Press.

Rideout, David P., and Rafael D. Sorkin. 1999. "Classical Sequential Growth Dynamics for Causal Sets." *Physical Review D* 61: 024002–1–16. doi: 10.1103/PhysRevD.61.024002

Savitt, Steven. 2002. "On Absolute Becoming and the Myth of Passage." In .), *Time, Reality and Experience* (Royal Institute of Philosophy Supplement 50), edited by Craig Callender, 153–167. Cambridge: Cambridge University Press.

Savitt, Steven. 2014. "Being and Becoming in Modern Physics." In *The Stanford Encyclopedia of Philosophy* (Summer 2014 Edition), edited by Edward N. Zalta. Stanford: Metaphysics Research Lab, Center for the Study of Language and Information, Stanford University. https://plato.stanford.edu/archives/sum2014/entries/spacetime-bebecome

Sorkin, Rafael D. 2005. "Causal Sets: Discrete Gravity." In *Lectures on Quantum Gravity (Proceedings of the Valdivia Summer School)*, edited by Andres Gomberoff and Donald Marolf, 305–327. New York: Plenum.

Sorkin, Rafael D. 2007. "Relativity Theory Does Not Imply that the Future Already Exists: A Counterexample." In .), *Relativity and the Dimensionality of the World*, edited by Vesselin Petkov, 153–161. Berlin: Springer.

Stachel, John. 2006. "Structure, Individuality, and Quantum Gravity." In *The Structural Foundations of Quantum Gravity*, edited by Dean Rickles, Steven French, and Juha Saatsi, 53–82. Oxford: Oxford University Press.

Stein, Howard. 1968. "On Einstein-Minkowski Space-time." *The Journal of Philosophy* 65(1): 5–23. doi: 10.2307/2024512

Stein, Howard. 1991. "On Relativity Theory and Openness of the Future." *Philosophy of Science* 58(2): 147–167. doi: 10.1086/289609

Tooley, Michael H. 1997. *Time, Tense, and Causation*. Oxford: Clarendon Press.

Varadarajan, Madhavan, and David Rideout. 2006. "General Solution for Classical Sequential Growth Dynamics of Causal Sets." *Physical Review D* **73**: 104021–1–10. doi: 10.1103/PhysRevD.73.104021

Wald, Robert M. 1984. *General Relativity*. Chicago and London: The University of Chicago Press.

Wallden, Petros. 2010. "Causal Sets: Quantum Gravity from a Fundamentally Discrete Spacetime." *Journal of Physics: Conference Series* 222(1): 012053. doi: 10.1088/1742–6596/222/1/012053

Williams, Donald C. [1951] 1967. "The Myth of Passage." In .), *The Philosophy of Time: A Collection of Essays*, edited by Richard M. Gale, 98–116. Garden City, New York: Doubleday and Company, Inc.

Williams, Donald. C. 1965. "Physics and Flux: Comment on Professor Čapek's Essay." In *Boston Studies in the Philosophy of Science*, *Vol. II: In honor of Philipp Frank* (Proceedings of the Boston Colloquium for the Philosophy of Science), edited by Robert S. Cohen and Marx W. Wartofsky, 464–475. New York: Humanities Press.

Wüthrich, Christian. 2012. "The Structure of Causal Sets." *Journal for General Philosophy of Science* 43: 223–241. doi: 10.1007/s10838–012–9205–1

Gábor Hofer-Szabó[*]
Research Center for the Humanities

Original Scientific Paper
UDK 530.1
167/168

THREE PRINCIPLES LEADING TO THE BELL INEQUALITIES

Abstract: *In the paper we compare three principles accounting for correlations, namely Reichenbach's Common Cause Principle, Bell's Local Causality Principle, and Einstein's Reality Criterion and relate them to the Bell inequalities. We show that there are two routes connecting the principles to the Bell inequalities. In case of Reichenbach's Common Cause Principle and Bell's Local Causality Principle one assumes a non-conspiratorial joint common cause for a set of correlations. In case of Einstein's Reality Criterion one assumes strongly non-conspiratorial separate common causes for a set of perfect correlations. Strongly non-conspiratorial separate common causes for perfect correlations, however, form a non-conspiratorial joint common cause. Hence the two routes leading the Bell inequalities meet.*

Keywords: *Einstein's reality criterion, Common Cause Principle, local causality, Bell inequalities*

1 Introduction

Many were pondering on the historical reasons of why it took thirty years to get from EPR argument to the Bell inequalities (see, for example: Bell 1964/2004; Howard 1985; Redhead 1987; Hájek and Bub 1992; Fine 1996; Norton 2004; Szabó 2008; Goldstein et al. 2011; Maudlin 2014; Lewis 2015). This paper has nothing to say about these historical and conceptual reasons. It rather intends to show that the route leading from Einstein's Reality Criterion to the Bell inequalities is no longer than the route starting off from two other principles standardly used to causally account for correlations, namely Reichenbach's Common Cause Principle and Bell's Local Causality Principle.

In the paper we will handle the three principles side by side and show how they relate to one another and to the Bell inequalities. In Section 2 we show how the principles are used to causally account for correlations; in Section 3 we use them to explain conditional correlations; and in Section 4 we trace the routes leading from the principles to the Bell inequalities. In the paper we deliberately keep the philosophical analysis short so that the formal parallelism will not be lost sight of.

[*] Research Center for the Humanities, Budapest; email: szabo.gabor@btk.mta.hu

2 Explaining correlations

Let A and B be two correlated but causally separated events represented in a classical probability space (Σ, p):

$$p(A \wedge B) \neq p(A)p(B) \tag{1}$$

One can invoke three principles to causally account for this correlation. If one is concerned only with the probabilistic aspects, one can apply

Reichenbach's Common Cause Principle: If there is a correlation between two events and there is no direct causal (or logical) connection between the correlated events, then there always exists a common cause of the correlation.

Formally, a common cause of the correlation (1) is a partition $\{C_k\}$ $(k \in K)$ in (Σ, p)—or in an extension of (Σ, p); see Hofer-Szabó, Rédei and Szabó 2013—such that for any $k \in K$:

$$p(A \wedge B|C_k) = p(A|C_k)p(B|C_k) \tag{2}$$

If one furthermore assumes that the events A and B also have spatiotemporal localization, for example they are located in spatially separated regions, V_A and V_B, respectively, then to causally account for them, one can invoke a further principle:

Bell's Local Causality Principle: "A theory will be said to be locally causal if the probabilities attached to values of local beables in a space-time region V_A are unaltered by specification of values of local beables in a spatially separated region V_B, when what happens in the backward light cone of V_A is already sufficiently specified, for example by a full specification of local beables in a space-time region V_C" (Bell 1990/2004, 239–240).

The figure Bell is attaching to this formulation is reproduced in Fig. 1 with the original caption. In a locally causal theory for any correlation between events A and B localized in spatially separated regions V_A and V_B, respectively, the *atomic partition* $\{C_k\}$ $(k \in K)$ in the probability space (Σ, p) associated to any region V_C causally shielding-off V_A from the common past of V_A and V_B as depicted Fig. 1 should satisfy (2).

Finally, suppose we interpret the correlation (1) epistemologically as a *prediction*. That is we interpret A as a *predicting event* and B as a *predicted event* and the prediction as a correlation between the two. After all, a prediction is ontologically nothing but an (ideally strong) correlation between two event types. Weather forecast is simply a correlation between the today announcement and the tomorrow weather. Moreover, in a prediction the predicted event cannot causally influence the predicting events. One can predict the tomorrow weather but not the yesterday weather.

Suppose furthermore that the following two requirements also hold: (i) The predicting event is also causally irrelevant for the predicted event. This can happen for example when

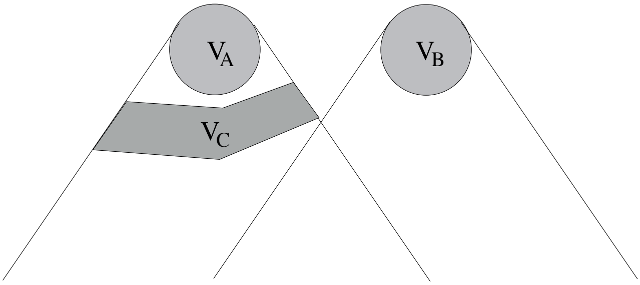

Figure 1: Full specification of what happens in V_C makes events in V_B irrelevant for predictions about V_A in a locally causal theory.

the two events are spatially separated. (ii) The correlation between A and B is *perfect*:

$$p(A \wedge B) = p(A) = p(B) \tag{3}$$

If all these hold, then we have a third principle to account for the correlation (3):

Einstein's Reality Criterion: "If, without in any way disturbing a system, we can predict with certainty (i.e. with probability equal to unity) the value of a physical quantity, then there exists an element of physical reality corresponding to this physical quantity" (Einstein, Podolsky, and Rosen 1935, 777–778).

Observe, that the term "without in any way disturbing a system" is just condition (i) above, and the term "predict with certainty" is just condition (ii). What Einstein's Reality Criterion requires is that in case of a perfect prediction, that is perfect correlation between causally separated events, an element of reality should account for the correlation.

What is an element of reality?

The distinctive feature of an element of reality (see Gömöri and Hofer-Szabó 2017 for the details) is that it determines the predicted event with certainty. Formally, an element of reality is a partition $\{C^+, C^-\}$ in (Σ, p) such that the following holds:

$$p(A \wedge B|C^+) = 1 \tag{4}$$

$$p(A \wedge B|C^-) = 0 \tag{5}$$

Now, let us go back to the Reichenbach's Common Cause Principle. It is well known that for perfect correlations a common cause that is a partition $\{C_k\}$ $(k \in K)$ satisfying (2) is *deterministic*: for any $k \in K$

$$p(A \wedge B|C_k) \in \{0, 1\} \tag{6}$$

Hence, the indices $k \in K$ can be grouped into two groups K^+ and K^- with $K^+ \vee K^- = K$ such that

$$C^+ = \vee_{k\in K^+} \mathrm{C_k} \tag{7}$$

$$C^- = \vee_{k\in K^-} \mathrm{C_k} \tag{8}$$

and $\{C^+, C^-\}$ satisfies (4)–(5). Common causes for perfect correlations understood as predictions are just elements of reality.

To sum up, a correlation between two events depending on whether we understand it purely probabilistically or spatiotemporally or in the context of predictions can be explained by three different principles: by Reichenbach's Common Cause Principle, by Bell's Local Causality Principle or by Einstein's Reality Criterion.

3 Explaining conditional correlations

Now, let us apply the above reasoning to measurements. Let a_i and b_j ($i \in I, j \in J$) be *measurement choices* and let $\{A_i, A'_i\}$ and $\{B_j, B'_j\}$ be binary *measurement outcomes* on two spatially separated systems. We will represent the measurement choices as two partitions $\{a_i\}$ ($i \in I$) and $\{b_j\}$ ($j \in J$) in a classical probability space (Σ, p), and the measurement outcomes by further partitioning the appropriate measurement choices a_i and b_j, respectively:

$$A_i \wedge A'_i = 0 \qquad A_i \vee A'_i = a_i \tag{9}$$

$$B_j \wedge B'_j = 0 \qquad B_j \vee B'_j = b_j \tag{10}$$

Suppose that for a given $i \in I$ and $j \in J$ the measurement outcomes A_i and B_j are *conditionally* correlated in the following sense:

$$p(A_i \wedge B_j|a_i \wedge b_j) \neq p(A_i|a_i)\, p(B_j|b_j) \tag{11}$$

What is the causal explanation of this conditional correlation?

Before we turn to the above principles, we make the following stipulation: Whatever explains the above correlations, it has to be causally and hence probabilistically independent of the measurement choices. In other words, in applying the above principles we will always require:

No-conspiracy: If a partition $\{C_k\}$ ($k \in K$) represents a set of events explaining the correlation (11), then for any $k \in K$ the following relation is required:

$$p(a_i \wedge b_j \wedge C_k) = p(a_i \wedge b_j)\, p(C_k) \tag{12}$$

Next, we formulate the three principles causally accounting for the conditional correlation between the measurement outcomes given certain measurement choices:

Reichenbach's Common Cause Principle: The common cause of the conditional correlation (11) is a partition $\{C_k\}$ in (Σ, p) such that for any $k \in K$:

$$p(A_i \wedge B_j|a_i \wedge b_j \wedge C_k) = p(A_i|a_i \wedge C_k)\, p(B_j|b_j \wedge C_k) \tag{13}$$

$$p(a_i \wedge b_j \wedge C_k) = p(a_i \wedge b_j)\, p(C_k) \tag{14}$$

Bell's Local Causality Principle: Suppose there is a conditional correlation (11) between measurement outcomes A_i and B_j given measurement choices a_i and b_j. Suppose further that A_i and a_i are localized in regions V_A and B_j and b_j are localized in regions V_B spatially separated from V_A. Then, if the theory accounting for this correlation is locally causal, then the atomic partition $\{C_k\}$ ($k \in K$) in (Σ, p) associated to the region V_C (see Fig. 1) should satisfy (13)–(14).

Einstein's Reality Criterion: Suppose that the conditional correlation (11) represents now a prediction. That is let A_i denote the outcome of a predicting event a_i and let B_j denote the outcome of the predicted event b_j. Suppose furthermore that A_i, a_i and B_j, b_j are causally separated. Also suppose that we can predict the outcome B_j of the measurement b_j by obtaining outcome A_i for the prediction a_i for sure. In other words, suppose that the conditional correlation is perfect:

$$p(A_i \wedge B_j | a_i \wedge b_j) = p(A_i | a_i) = p(B_j | b_j) \tag{15}$$

Then Einstein's Reality Criterion claims that there are elements of reality that is a partition $\{C^+, C^-\}$ in (Σ, p) explaining correlation (15) in the following sense:

$$\begin{aligned} p(A_i \wedge B_j | a_i \wedge b_j \wedge C^+) &= 1 & (16) \\ p(A_i \wedge B_j | a_i \wedge b_j \wedge C^-) &= 0 & (17) \\ p(a_i \wedge b_j \wedge C^+) &= p(a_i \wedge b_j)\, p(C^+) & (18) \\ p(a_i \wedge b_j \wedge C^-) &= p(a_i \wedge b_j)\, p(C^-) & (19) \end{aligned}$$

Just as above, in case of a perfect correlation a common cause $\{C_k\}$ ($k \in K$) satisfying (13)–(14) is deterministic, hence a suitable grouping of the C_k-s *via* (7)–(8) will yield the elements of reality C^+ and C^-. In short, Einstein's Reality Criterion is a special case of Reichenbach's Common Cause Principle when the correlation is perfect (for the details see Gömöri and Hofer-Szabó 2017).

To sum up, the core of all three principles is to account for correlations in terms of a non-conspiratorial common cause. In case of Reichenbach's Common Cause Principle only the probabilistic aspects (13)–(14) of the common cause are taken into consideration. In case of Bell's Local Causality Principle both the correlated events and also the common cause have a spatiotemporal localization. In case of Einstein's Reality Criterion the whole correlation scenario is interpreted in the framework of a prediction and the correlation is taken to be perfect.

Before we move on to the relation of the principles to the Bell inequalities, let us see how the conditional and unconditional correlations and their explanations relate to one another.

First, observe that if the measurement choices are causally and therefore probabilistically independent, that is if for any $i \in I$ and $j \in J$:

$$p(a_i \wedge b_j) = p(a_i)\, p(b_j) \tag{20}$$

and the algebraic inclusions (9)–(10) hold, then the outcomes A_i and B_j are correlated in the *conditional* sence

$$p(A_i \wedge B_j|a_i \wedge b_j) \neq p(A_i|a_i)\, p(B_j|b_j) \tag{21}$$

if and only if they are correlated in the *unconditional* sense

$$p(A_i \wedge B_j) \neq p(A_i)\, p(B_j) \tag{22}$$

Second, given (9)–(10) and (20), $\{C_k\}$ is a non-conspiratorial common cause of the *conditional* correlation (21):

$$p(A_i \wedge B_j|a_i \wedge b_j \wedge C_k) = p(A_i|a_i \wedge C_k)\, p(B_j|b_j \wedge C_k) \tag{23}$$

$$p(a_i \wedge b_j \wedge C_k) = p(a_i \wedge b_j)\, p(C_k) \tag{24}$$

if and only if $\{C_k\}$ is a non-conspiratorial common cause of the *unconditional* correlation (22):

$$p(A_i \wedge B_j|C_k) = p(A_i|C_k)\, p(B_j|C_k) \tag{25}$$

$$p(a_i \wedge b_j \wedge C_k) = p(a_i \wedge b_j)\, p(C_k) \tag{26}$$

(For the proof see Hofer-Szabó, Rédei and Szabó 2013, Lemma 9.8.). Therefore, on the assumptions (9)–(10) and (20), the common causal explanations (23)–(24) and (25)–(26) are interchangeable.

4 From the principles to the Bell inequalities

How the above three principles serving for a causal explanation of correlations relate to the Bell inequalities? The crucial point is to see how the different principles relate to the common causal explanation of *more correlations*. Principally, there are two possible ways: either the different correlations are explained by a *joint common cause* or each correlation is explained by a *separate common cause*. The standard derivation of the Bell inequalities from Reichenbach's Common Cause Principle and Bell's Local Causality Principle assumes a joint common cause; whereas the derivation of the Bell inequalities from Einstein's Reality Criterion assumes only separate common causes. Since the assumption of separate common causes is weaker than that of a joint common cause, the derivation of the Bell inequalities from Einstein's Reality Criterion needs a stronger version of no-conspiracy.

Let us see the derivations in turn:

Reichenbach's Common Cause Principle. Suppose that $I = J = \{1, 2\}$ and the events A_i and B_j are all conditionally correlated that is for any $i, j \in I$:

$$p(A_i \wedge B_j|a_i \wedge b_j) \neq p(A_i|a_i)p(B_j|b_j) \tag{27}$$

The four correlations are said to have a *non-conspiratorial joint common cause* if there is a single partition $\{C_k\}$ $(k \in K)$ in (Σ, p) (or in an extension of (Σ, p)) such that for all $i, j \in I$ and $k \in K$ the following hold:

$$p(A_i \wedge B_j | a_i \wedge b_j \wedge C_k) \ = \ p(A_i | a_i \wedge C_k)\, p(B_j | b_j \wedge C_k) \tag{28}$$

$$p(a_i \wedge b_j \wedge C_k) \ = \ p(a_i \wedge b_j)\, p(C_k) \tag{29}$$

We claim that the events A_i, B_j, a_i and b_j with a non-conspiratorial joint common causal explanation satisfy the *Clauser-Horne inequalities* that is for any $i, i', j, j' \in I$ and $i \neq i'$, $j \neq j'$:

$$\begin{aligned} &-1 \leqslant p(A_i \wedge B_j | a_i \wedge b_j) + p(A_i \wedge B_{j'} | a_i \wedge b_{j'}) \\ &+p(A_{i'} \wedge B_j | a_{i'} \wedge b_j) - p(A_{i'} \wedge B_{j'} | a_{i'} \wedge b_{j'}) \\ &-p(A_i | a_i) - p(B_j | b_j) \leqslant 0 \end{aligned} \tag{30}$$

For the proof see the Appendix.

Bell's Local Causality Principle. Again, let $I = J = \{1, 2\}$. Suppose that the events A_i and B_j localized in spatially separated regions V_A and V_B respectively, are all conditionally correlated in the sence of (27). In a *locally causal theory* the atomic partition of the local algebra associated to V_C (see again Fig. 1) is a non-conspiratorial joint common cause in the sense of (28)–(29). Hence the Clauser-Horne inequalities (30) follow, just as in the case of Reichenbach's Common Cause Principle.

Einstein's Reality Criterion. Suppose now that $I = J = \{1, 2, 3, 4\}$ and there is a perfect conditional correlation between (the predicting events) A_i and (the predicted events) B_j for any $i = j \in$ I:

$$p(A_i \wedge B_i | a_i \wedge b_i) = p(A_i | a_i) = p(B_i | b_i) \tag{31}$$

First, observe that the four correlations in (31) are not the same as the correlations (27) above. In (27) $I = J = \{1, 2\}$ and the four correlations were not necessarily perfect; in (31) $I = J = \{1, 2, 3, 4\}$ and the four correlations are the $i = j$ perfect correlations.

Now, Einstein's Reality Criterion does *not* assume that all four correlations in (31) have a *joint* common cause. All it assumes is that there are *separate* elements of reality to each correlation, that is for any $i \in I$ there is a partiton $\{C_i^+, C_i^-\}$ satisfying

$$p(A_i \wedge B_i | a_i \wedge b_i \wedge C_i^+) = 1 \tag{32}$$

$$p(A_i \wedge B_i | a_i \wedge b_i \wedge C_i^-) = 0 \tag{33}$$

However, instead of simply requiring no-conspiracy:

$$p(a_i \wedge b_j \wedge C_k^+) = p(a_i \wedge b_j)\, p(C_k^+) \tag{34}$$

$$p(a_i \wedge b_j \wedge C_k^-) = p(a_i \wedge b_j)\, p(C_k^-) \tag{35}$$

($i, j, k \in I$) one requires *strong no-conspiracy*, namely that *any* element C in the Boolean algebra generated by the four pairs of elements of reality $\{C_k^{\pm}\}$ should be independent of *any* combination of the measurement choices:

$$p(a_i \wedge b_j \wedge C) = p(a_i \wedge b_j)\, p(C) \tag{36}$$

In short, in case of more correlations Einstein's Reality Criterion requires *less* than the other two principles since it requires only separate elements of reality for the different correlations, but also requires *more* since it requires all Boolean combinations of the elements of reality to be independent of the measurement choices.

The derivation of the Clauser–Horne inequalities (30) from a strongly non-conspiratorial separate common casual explanation is straightforward. From (31), (32)–(33) and (36) it follows that for any $i, j \in I$:

$$p(A_i|a_i) = p(B_i|b_i) = p(C_i^+) \tag{37}$$

$$p(A_i \wedge B_j|a_i \wedge b_j) = p(C_i^+ \wedge C_j^+) \tag{38}$$

Now, it is an elementary fact of classical probability theory that for any four events C_i^+, $C_{i'}^+$, C_j^+ and $C_{j'}^+$ in (Σ, p) we have:

$$\begin{aligned} -1 \leqslant\; & p(C_i^+ \wedge C_j^+) + p(C_i^+ \wedge C_{j'}^+) + p(C_{i'}^+ \wedge C_j^+) \\ & - p(C_{i'}^+ \wedge C_{j'}^+) - p(C_i^+) - p(C_j^+) \leqslant 0 \end{aligned} \tag{39}$$

Substituting (37)–(38) into (39) one arrives at (30).

What one proves here is that the atomic partition composed of the intersections of strongly non-conspiratorial separate common causes *for perfect correlations* form a non-conspiratorial joint common cause for all correlations. Note that in the general case that is *for non-perfect correlations* the relation between separate and joint common causes is not so straightforward and the relation of strongly non-conspiratorial separate common causes to the Bell inequalities is not known (see Hofer-Szabó, Rédei and Szabó 2013, Conjecture 9.11.).

To sum up, one can arrive at the Bell inequalities from the three principles on two different routes. In the standard derivation based on Reichenbach's Common Cause Principle or Bell's Local Causality Principle one takes four correlations and assumes that they have a non-conspiratorial joint common cause. In case of Einstein's Reality Criterion one takes four perfect correlations and assumes that each has a separate common cause which together are strongly non-conspiratorial. Both routes lead directly to the Clauser-Horne inequalities.

5 Conclusions

In this paper we compared three principles accounting for correlations and related them to the Bell inequalities. Reichenbach's Common Cause Principle, in the original sense at least, refers only to one correlation: it demands a common cause for a given correlation if the direct causal link between the correlata can be excluded. In the derivation of the Bell inequalities, however, the principle had to be used in a stronger sense, namely demanding one and the same cause for a set of correlations. Bell's Local Causality Principle has already been formulated originally in this strong sense: all correlations localized in spatially separated regions were to be screened-off by the "full specification" of an appropriately localized third spacetime region. In this sense Bell's Local Causality Principle is a stronger principle than Reichenbach's Common Cause Principle. Finally, Einstein's Reality Criterion again assumes elements of reality to each correlation separately, similar to Reichenbach's Common Cause Principle. Moreover, it does so only in case of correlations. In this sense Einstein's Reality Criterion seems to be even weaker than Reichenbach's Common Cause Principle.

Note, however, that not even the strongest of the three principles, namely Bell's Local Causality Principle implies the Bell inequalities on its own. Even this principle needs to assume that the common causes or elements of reality causally responsible for the correlations are causally and hence probabilistically independent from the measurement choices. To be sure, no-conspiracy seems to be a natural requirement for an element of reality to deserve its name. No-conspiracy, however, can be defined in different strength. And this is the point where the principles faring worse at the beginning can catch up. Even though Einstein's Reality Criterion provides only separate elements of reality for the correlations, if these elements of reality are strongly non-conspiratorial, then they suffice to derive the Bell inequalities. In short, no-conspiracy together with joint elements of reality and strong no-conspiracy together with separate elements of reality fare equally well in the derivation of the Bell inequalities.

Appendix

Proof. It is an elementary fact of arithmetic that for any $\alpha, \alpha', \beta, \beta' \in [0, 1]$ we have

$$-1 \leqslant \alpha\beta + \alpha\beta' + \alpha'\beta - \alpha'\beta' - \alpha - \beta \leqslant 0 \tag{40}$$

Now, let $\alpha, \alpha', \beta, \beta'$ be

$$\begin{aligned} \alpha &= p(A_i|a_i \wedge C_k) && (41)\\ \alpha' &= p(A_{i'}|a_{i'} \wedge C_k) && (42)\\ \beta &= p(B_j|b_j \wedge C_k) && (43)\\ \beta' &= p(B_{j'}|b_{j'} \wedge C_k) && (44) \end{aligned}$$

Substituting (41)–(44) into (40) we get

$$\begin{aligned} -1 \leqslant p(A_i|a_i \wedge C_k)\, p(B_j|b_j \wedge C_k) + p(A_i|a_i \wedge C_k)\, p(B_{j'}|b_{j'} \wedge C_k)\\ +p(A_{i'}|a_{i'} \wedge C_k)\, p(B_j|b_j \wedge C_k) - p(A_{i'}|a_{i'} \wedge C_k)\, p(B_{j'}|b_{j'} \wedge C_k)\\ -p(A_i|a_i \wedge C_k) - p(B_j|b_j \wedge C_k) \leqslant 0 \end{aligned} \tag{45}$$

Using the screener-off condition (28) we obtain

$$\begin{aligned} -1 \leqslant p(A_i \wedge B_j|a_i \wedge b_j \wedge C_k) + p(A_i \wedge B_{j'}|a_i \wedge b_{j'} \wedge C_k) \\ +p(A_{i'} \wedge B_j|a_{i'} \wedge b_j \wedge C_k) - p(A_{i'} \wedge B_{j'}|a_{i'} \wedge b_{j'} \wedge C_k) \\ -p(A_i|a_i \wedge C_k) - p(B_j|b_j \wedge C_k) \leqslant 0 \end{aligned} \tag{46}$$

Multiplying by $p(C_k)$, using no-conspiracy (29) and summing up for k one arrives at (30).

Acknowledgements. I wish to thank Márton Gömöri and Balázs Gyenis for valuable discussions. This work has been supported by the Hungarian Scientific Research Fund, OTKA K-115593.

References

J. S. Bell, "On the Einstein-Podolsky-Rosen paradox," *Physics*, 1, 195–200 (1964); reprinted in (Bell 2004, 14–21).

J. S. Bell, "La nouvelle cuisine," in: J. Sarlemijn and P. Kroes (eds.), *Between Science and Technology*, Elsevier, (1990); reprinted in (Bell 2004, 232–248).

J. S. Bell, *Speakable and Unspeakable in Quantum Mechanics*, (Cambridge: Cambridge University Press, 2004).

A. Einstein, B. Podolsky and N. Rosen, "Can Quantum Mechanical Description of Physical Reality be considered complete?," *Phys. Rew.*, 47, 777–780 (1935).

A. Fine, *The Shaky Game, Einstein, Realism and the Quantum Theory*, (Chicago: University of Chicago Press, 1996).

M. Gömöri and G. Hofer-Szabó, "On the meaning of EPR's Criterion of Reality," (in preparation, 2017).

S. Goldstein, T. Norsen, D. V. Tausk, and N. Zanghi, "Bell's theorem," *Scholarpedia*, 6(10), 8378 (2011).

A. Hájek and J. Bub, "EPR," *Found. Phys.*, 22, 313–331 (1992).

G. Hofer-Szabó, M. Rédei and L. E. Szabó, *The Principle of the Common Cause*, (Cambridge: Cambridge University Press, 2013).

G. Hofer-Szabó and P. Vecsernyés, "A generalized definition of Bell's local causality," *Synthese* 193(10), 3195–3207 (2016).

D. Howard, "Einstein on Locality and Separability," *Stud. Hist. Phil. Sci.*, 16, 171–201 (1985).

P. J. Lewis, "Bell's theorem, realism, and locality," URL = http://philsci-archive.pitt.edu/11372/ (2015).

T. Maudlin, "What Bell did," *J. Phys. A: Math. Theor.*, 47, 424010 (2014).

J. D. Norton, *Einstein for Everyone*, URL = http://www.pitt.edu/ jdnorton/teaching/HPS_0410/index.html

M. Redhead, *Incompleteness, Nonlocality, and Realism*, (Oxford: Clarendon Press, 1987).

L. E. Szabó, "The Einstein-Podolsky-Rosen Argument and the Bell Inequalities," *Internet Encyclopedia of Philosophy*, URL= http://www.iep.utm.edu/epr/ (2008).

O. Cristinel Stoica[*]
Horia Hulubei National Institute for Physics
and Nuclear Engineering,
Department of Theoretical Physics,
Iulian D. Toader[†]
The Research Institute,
University of Bucharest

Original Scientific Paper
UDK 114/115:530.12

SPACETIME SINGULARITIES AND INVARIANCE

1 Introduction

Spacetime singularities are an important topic in general relativity and in cosmology, but understanding their philosophical significance is rather a side-issue in contemporary debates in philosophy of physics. The prevailing attitude still seems to be that singularities constitute a breakdown of physical laws.[46] A notable exception to this rather unfortunate state of affairs is John Earman's book, *Bangs, Crunches, Whimpers, and Shrieks* (Earman 1995), published 20 years ago, focused on issues related to the definition, proper characterization, and existence of singularities, as well as on several problems and hypotheses that they are thought to have given rise to. Proposing a tolerant attitude towards singularities, Earman discusses cosmic censorship, supertasks, and the horizon problem, among other issues. What we discuss in the present paper is a novel approach to singularities, based on a recent extension of general relativity that shows why singularities do not actually constitute a breakdown of physical laws: it is not only the case that physical laws are valid, but they also remain invariant at singularities (Stoica 2013). We are interested here in describing this kind of invariance, as well as in drawing its consequences for our understanding of equivalence in general relativity. In particular, adopting a distinction recently introduced by Dennis Dieks (Dieks 2006), we point out that the difference between the metrics at singularities and those outside of singularities is factual,

[*] Horia Hulubei National Institute for Physics and Nuclear Engineering, Department of Theoretical Physics, Bucharest. Corresponding author: cristi.stoica@theory.nipne.ro

[†] The Research Institute, University of Bucharest; Descartes Centre for the History and Philosophy of the Sciences and the Humanities, Utrecht University. Contact: itoad71@gmail.com or i.d.toader@uu.nl

[1] See, e.g., this remark in a recent collection on the philosophy of general relativity: "If you think you have a singularity, then you can't use it in a physical model. You don't know how to include such an object in a physical system, either as the outcome of gravitational collapse or as an object that might affect other objects with its gravitational field. [...] you are paralyzed by incomprehension" (Schutz 2012).

rather than nomological, and that this justifies the extension of the principle of equivalence to singularities.

Singularities, let us recall, have been discovered in the simplest solution representing the spacetime outside a body — the solution given by Schwarzschild, soon after Einstein proposed general relativity, one hundred years ago. As typically characterized, singularities are regions where the metric tensor is no longer regular, so that the mathematical objects in the Einstein equation and other field equations become infinite or undefined. Most physicists, in particular Einstein, initially rejected the possibility of singular solutions. The hope was that they are due to idealizations like the perfect spherical symmetry, and would not occur in the real world, but the singularity theorems by Penrose (Penrose 1965; 1969) and Hawking (Hawking 1966a; 1966b; 1967; Hawking and Penrose 1970) show that they are in fact unavoidable. Singularities are indeed predicted by the theory to occur both inside black holes, and at the Big Bang. This confirmed the ideas of those who thought that general relativity contained the seeds of its own destruction. Penrose then proposed the cosmic censorship hypothesis, which states that although singularities exist, they are isolated beyond the event horizon and so don't affect the physics outside the black hole (Penrose 1979; 1998). Hawking, however, showed that if quantum effects are considered, black holes can evaporate, and so the problem persists and it is even aggravated (Hawking 1975; 1976). This is the well-known information loss paradox.

The novel approach to singularities, adopted here and briefly described further below, gives alternative formulations of the geometric objects and the field equations, which don't break down at singularities and also remain invariant. A new branch of geometry emerges in this way, namely an invariant and more general extension of singular semi-Riemannian geometry (Stoica 2014c), as well as a new physical theory: *singular general relativity* (Stoica 2013). This provides an alternative formulation of differential geometry and general relativity, one which is equivalent to the standard one, but which can be extended at singularities too. This formulation turns out to solve many of the problems related to singularities. But its philosophical significance has yet to be fully articulated and evaluated.

Famously, Einstein and Rosen had remarked that one could multiply the standard equations by a suitable quantity, which vanishes at the singularity and removes the infinities, so the singularities may be not harmful after all (Einstein and Rosen 1935). The novel approach adopted here can be seen as a mathematical follow-up on this remark, in the sense that it provides an invariant account leading to this multiplication, and justifies it mathematically and physically. This approach not only commends an attitude of tolerance towards singularities (Earman 1996), but is animated by the belief that, when correctly understood, they are a source of fruitful developments in general relativity and quantum gravity (Stoica 2014b).

As far as we can see, the novel approach to spacetime singularities has important consequences for several topics of interest in philosophy of physics. Since at singularities, the distance between distinct points, as well as the duration between distinct events, can vanish, a further revision of our notion of spacetime, already revised by relativity and quantum mechanics, seems to be required. Similarly, issues concerning causality arise as well. Since distinct events are no longer separated in space or time, how does that change the way we should think about causality? Can the evolution equations be extended beyond the singularity? Are singularities really the "end of the line," as is often believed? Standard treatments of singularities don't seem to help here, but the new approach allows the extension of spacetime and the fields beyond singularities. There are also related issues, having to do with the occurrence of Cauchy horizons, which seem to accompany charge and rotating black holes. Cauchy horizons causally disconnect some regions of spacetime, in the sense that the evolution equations defined on spacelike surfaces in the past of these regions don't reach them. The new approach to singularities allows black hole solutions to be compatible with global hyperbolicity, and hence with the absence of Cauchy horizons (Stoica 2012c).

Furthermore, formulating general relativity and the geometry of spacetime in a way which is defined at singularities seems to suggest that the variables used in this formulation are more fundamental than the standard ones, which fail at singularities. This raises a metaphysical question about fundamentality: what geometric and physical fields are really the fundamental ones, and why? In particular, some charts are singular and are not good for representing the black hole solutions around the event horizons, but others are not singular, and are proper for this task. So the choice of the differential structure of the spacetime manifold can be done in many ways, but this choice should receive a good physical, mathematical, and philosophical justification.

In this paper, we want to focus our discussion on the problem of invariance of the physical laws. Since at singularities the metric is no longer regular, but degenerate, Lorentz invariance (in the tangent space at each spacetime point) is violated and needs to be replaced by a different type of invariance. But what type, more precisely?

2 Singularities in General Relativity

According to special relativity, physical laws are invariant with respect to the Poincaré group (Einstein 1905). This is the group of isometries of the four-dimensional spacetime (the Minkowski spacetime), and includes translations and linear transformations that leave invariant the Lorentz metric

$$\Delta s^2 = -c^2\Delta t^2 + \Delta x^2 + \Delta y^2 + \Delta z^2, \qquad (1)$$

where t is the time coordinate, x, y, z are Cartesian space coordinates, and c is the speed of light. In the absence of gravitational forces, the equations expressing physical laws preserve their form when a Poincaré transformation is applied. In special relativity, the set of positions occupied at different moments of time by an inertially-moving particle represents a straight line. This line can be parametrized so as to give its *proper time*. To every instant of time corresponds a point on the time axis of the observer, and a three-dimensional space orthogonal (with respect to the inner product (1)) to the time axis at that point gives the *proper space* of the observer. An orthogonal *inertial reference frame* of the observer can be defined by choosing an origin on the time axis, a future-pointing vector on the time axis, and an orthogonal frame in the three-dimensional space orthogonal to the time axis at the origin. All such reference frames are related to one another by a Poincaré transformation.

Of course, it is not necessary for a reference frame to consist of orthogonal vectors. Any affine transformation of the Minkowski spacetime can transform any inertial frame into another inertial frame, but only the Poincaré transformations preserve the orthogonal character of inertial frames. In order to deal with inertial frames that are not orthogonal, one has to change the form of the metric, although the metric itself is an invariant mathematical object. The general form of the metric, valid in any inertial frame, is

$$\Delta s^2 = \sum_{a,b=0}^{3} g_{ab} \Delta x^a \Delta x^b = g_{ab} \Delta x^a \Delta x^b, \tag{2}$$

where $(x^0,x^1,x^2,x^3) := (t,x,y,z)$. By Einstein's summation convention, one can drop the $\Sigma^3{}_{a,b=0}$ symbol and sum over all indices which appear twice, in both lower and upper positions. In the case of non-inertial observers, one has to use curvilinear coordinates, such as the spherical coordinates, where the form of the metric tensor varies from point to point and the quantities Δx^a become infinitesimal. This means that while for orthogonal inertial frames we could identify the Minkowski spacetime with a four-dimensional vector space on which the metric is defined, for curvilinear coordinates one can no longer make this identification. Instead, one introduces at each point of spacetime a four-dimensional vector space, the so-called *tangent space* to the Minkowski spacetime at that point. The metric is then actually defined on the tangent space at that point. So, the infinitesimal length becomes

$$\mathrm{d}s^2 = g_{ab} \mathrm{d}x^a \mathrm{d}x^b, \tag{3}$$

where the coefficients g_{ab} are taken to depend on position.

Observing that curvilinear coordinates introduce additional inertial forces, Einstein realized that, if spacetime is curved, gravity can be included as such an inertial force, thereby obtaining the theory of general relativity. But, of course, on a curved spacetime the notion of Poincaré transformations no longer applies.

Spacetime can no longer be identified with a vector space, and the Lorentz metric is no longer the same in all spacetime. Due to the curvature, one can use only local curvilinear coordinates, defined on open neighborhoods, and the metric varies from point to point, being defined independently on the tangent space at each point of the spacetime manifold. The coordinate transformations are local diffeomorphisms. On the Minkowski spacetime, diffeomorphisms include the Poincaré transformations. The field equations, that is, the Einstein equation and the equations describing the behavior of matter fields, are all formulated in tensor formalism, and are invariant under the group of diffeomorphisms. The Einstein equation is

$$G_{ab} + \Lambda g_{ab} = \kappa T_{ab}, \tag{4}$$

where the Einstein tensor $G_{ab} = R_{ab} — \frac{1}{2} R g_{ab}$ is equated with the stress-energy tensor T_{ab} of matter, and the constant Λ is the cosmological constant, responsible for the accelerated expansion of the universe. This formulation does not violate Lorentz invariance, because diffeomorphisms also change the components of the metric tensor, so lengths remain invariant under diffeomorphisms. Therefore, in general relativity, Lorentz invariance is preserved in the tangent space at each point in spacetime.

However, as is well known, soon after Einstein gave this formulation of the equation, Schwarzschild found a solution describing the spacetime for a spherically symmetric gravitational field. The Schwarzschild metric is, in Schwarzschild coordinates,

$$\mathrm{d}s^2 = -\left(1 - \frac{2m}{r}\right)\mathrm{d}t^2 + \left(1 - \frac{2m}{r}\right)^{-1}\mathrm{d}r^2 + r^2\mathrm{d}\sigma^2, \tag{5}$$

where

$$\mathrm{d}\sigma^2 = \mathrm{d}\theta^2 + \sin^2\theta\mathrm{d}\phi^2. \tag{6}$$

It turns out that this metric has two singularities, one corresponding to r = *2m* (the so-called *event horizon*), and another to r = 0. As showed by Eddington (Eddington 1924) and Finkelstein (Finkelstein 1958), suitable coordinate transformations can remove the event horizon singularity, proving that this is an artefact of the Schwarzschild coordinates. But the same does not work for the r = 0 singularity, because no coordinate transformations can remove the infinite value of the Kretschmann scalar $R^{abcd}R_{abcd}$ at $r = 0$. Another exact solution was proposed by Friedmann, representing an expanding universe (Friedman 1922; 1924; 1999). Its modern form is the Friedmann-Lemaitre-Robertson-Walker (FLRW) metric,

$$\mathrm{d}s^2 = -\mathrm{d}t^2 + a^2(t)\mathrm{d}\Sigma^2, \tag{7}$$

where

$$\mathrm{d}\Sigma^2 = \frac{\mathrm{d}r^2}{1 - kr^2} + \mathrm{r}^2\left(\mathrm{d}\theta^2 + \sin^2\theta\mathrm{d}\phi^2\right) \tag{8}$$

The spacetime manifold is here the product between a one-dimensional manifold representing the time, and a three-dimensional *symmetric space* E, which can be the three-sphere S^3, the Euclidean space $\mathbb{R}^3$, or the hyperbolic space H^3. Also $k = 1$ for S^3, $k = 0$ for $\mathbb{R}^3$, and $k = —1$ for H^3. This solution has one singularity, at the *big-bang*, where $a(t) = 0$.

3 Singular General Relativity

The fact that these two solutions have singularities is not an accident due to the very high symmetry, as one initially hoped. The theorems offered by Penrose (Penrose 1965; 1969) and Hawking (Hawking 1966a; 1966b; 1967; Hawking and Penrose 1970) show that, in general relativity, singularities have to occur in very general situations, such as encountered in our universe. The $r = 0$ singularity of the Schwarzschild black hole is a spacelike singularity: time seems to end there. Any matter falling into the black hole reaches the singularity and vanishes, together with the information which it contains. The problem only gets bigger when quantum theory is taken into account. Because of particle creation in curved spacetime (Hawking 1975), black holes evaporate, and at the end, part of the information describing the state of the universe appears to be lost (Hawking 1976). But what happens to Lorentz invariance?

Before we answer this question, let's distinguish between *benign* singularities and *malign* singularities, corresponding to the two main ways in which the metric tensor g_{ab} can become singular: by having all its components smooth, but vanishing determinants, so that the metric becomes degenerate, and by having some of its components become infinite, respectively. In the case of a degenerate metric, the reciprocal metric g^{ab} is singular. This makes it impossible to use two important tensor operations: the contraction between covariant components, which is usually done by contracting with g^{ab}, and the covariant derivative, which is typically given in terms of the Christoffel symbol of the second kind,

$$\Gamma^{c}{}_{ab} = \tfrac{1}{2} g^{cs}(\partial_a g_{bs} + \partial_b g_{sa} - \partial_s g_{ab}). \tag{9}$$

Without the covariant derivative, in particular, one can neither write the field equations, nor can one define the Riemann curvature tensor, $\mathrm{R}^{a}{}_{bcd}$, in the usual way. In the case of a degenerate metric that has a constant signature (Barbilian 1939; Moisil 1940; Strubecker 1941; 1942a; 1942b; 1944; Vrănceanu 1942), one can define an operation similar to the covariant derivative (Kupeli 1987c; 1987a; 1987b; 1996).

However, the definition relies on choosing a subspace on the tangent space at each point, in a non-invariant way. Moreover, it could not be used for spacetime singularities, due to the constant signature of the metric.

More recently, however, one of the authors developed a novel approach that works, in an invariant way, with degenerate metrics that have a variable signature (Stoica 2014c). The problem of contraction between covariant components has been resolved for a special type of tensors, which belong to tensor products of the tangent space and a subspace of the cotangent space of covectors of the form $g_{ab}v^{b}$. It turns out that such tensors contain all that is needed to describe singularities in terms of finite quantities, and even to construct the Riemann curvature. This approach defines a different kind of covariant derivative, which remains finite, in terms of the Christoffel symbol of the first kind,

$$\Gamma_{abc} = \tfrac{1}{2}\,(\partial_a g_{bc} + \partial_b g_{ca} - \partial_c g_{ab}). \tag{10}$$

This allowed the definition of the Riemann curvature tensor R_{abcd}. As a consequence, Einstein's equation (4) could be rewritten, in a different form, which is however equivalent to (4) outside the singularities, but remains finite and smooth at singularities (Stoica 2014c; 2014a). The new formalism can handle the FLRW singularities as well, which turned out to be benign, resulting in a geometric and physical description in terms of finite quantities (Stoica 2015b; 2012b). Similarly, for the black hole singularities, coordinate transformations were found that transform the malign singularity at $r = 0$ into a benign one (Stoica 2012d; 2012a; 2015c). In the case of the Schwarzschild singularity, it has been shown that the coordinate transformation

$$\begin{cases} r &= \tau^2 \\ t &= \xi\tau^4. \end{cases} \tag{11}$$

results in the following form of the metric

$$\mathrm{d}s^2 = -\frac{4\tau 4}{2m - \tau^2}\,\mathrm{d}\tau^2 + (2m - \tau^2)\tau^4\,(4\xi\mathrm{d}\tau + \tau\mathrm{d}\xi)^2 + \tau^4\mathrm{d}\sigma^2, \tag{12}$$

which renders the $r = 0$ singularity benign. The Schwarzschild singularity is, of course, still a singularity, because the metric is degenerate, but the new formalism can be applied to obtain a geometric description of the singularity in terms of finite quantities. Also, the solution extends analytically beyond the singularity, so that in the case of Hawking evaporation, the spacetime is recovered. The upshot of the new approach is that, if they are benign, spacetime singularities do not constitute a problem for our physical theorizing, and in particular they should not be considered as a breakdown of physical laws.

Naturally, at those points in the spacetime manifold where the metric is de generate, Lorentz invariance is violated *because* the metric is degenerate. On account of this violation, the Lorentz group has to be replaced by a group of linear transformations of the tangent space that preserve the degenerate metric at those very points. But how does such a group look like?

A degenerate metric g on a vector space V, in particular on the tangent space at a spacetime point, defines an inner product on V, by $\langle u, v\rangle = g_{ab}u^a v^b$. The vectors $v \in V$ satisfying $\langle u, v\rangle = 0$ for any vector $u \in V$ are called degenerate, and they form a vector subspace of V, called the *radical* of g. The vector space V can be split as a direct sum between the radical and a complementary vector space of dimension equal to the rank of the metric, on which the restriction of $g$is non-degenerate. The vectors from the dual space V^* act on V as linear forms. Those co-vectors which vanish on the radical of g form a vector space of dimension equal to the rank of the metric, called the *annihilator* of the radical. A general linear transformation on V also acts on the dual V^*. If the linear transformation preserves the metric, it has to map the radical onto itself. On the dual space, it maps the annihilator onto itself while preserving the non-degenerate metric induced by g on the annihilator. These general linear transformations form a group which preserves the metric.

Such groups were studied by the Romanian mathematician Dan Barbilian (Barbilian 1939). In particular, if the metric is non-degenerate, the group is the orthogonal group of the metric. If the metric is degenerate in all directions, that is, if it vanishes, then the group preserving it is the general linear group of V. The group that is needed, in our case, to replace the Lorentz group at singularities is the intermediate case, i.e., a group that preserves the degenerate metric only at some points in spacetime. The replacement is not required outside the singularities, where the metric is non-degenerate. Thus, outside the singularities there is no violation of the Lorentz invariance (in the tangent space at every point). The replacement, moreover, does not have any physical consequences outside the singularities (at least if quantum effects are ignored).

4 Discussion

Now, in summary, the standard formulation of general relativity presents the Einstein equation as valid and generally covariant outside the singularities, but as breaking down at singularities. However, for the reasons expounded above, it turns out that this standard formulation can be circumvented. The extension of general relativity to singularities introduces an equivalent, generally covariant reformulation of the Einstein equation, which does not break down at (benign) singularities. Its fundamental equation, just like the Einstein equation, is invariant under the symmetry group of local diffeomorphisms. Unlike the Einstein equation, the fundamental equation of singular general relativity is also Barbilian invariant, i.e., invariant under a symmetry group that preserves only the degenerate metric at (benign) singularities, in the tangent space at every point.

Furthermore, singular general relativity also extends the principle of equivalence to singularities. In a recent paper, Dennis Dieks introduced an useful distinction between factual and nomological differences between reference frames, i.e., "differences that should be seen as *fact-like* rather than *law-like*" (Dieks 2006). He argued that, whereas in classical mechanics and in special relativity, the differences between inertial and accelerated frames should be understood as nomological differences, in general relativity, the differences between all frames should be understood as purely factual differences, since the spacetime manifold itself is a dynamical structure rather than a fixed background. This allows for an understanding of the equivalence between reference frames in the sense that there are no nomological differences between them, and justifies the claim, which had been repeatedly challenged, that Einstein was right in thinking that general relativity extends the principle of equivalence to accelerated motion. In other words, understanding equivalence in the way suggested by Dieks justifies the claim that Einstein was right in thinking that the equivalence of all arbitrary frames is a consequence of the general covariance of his equation. This only assumes that general covariance is regarded not as a merely formal requirement, in the sense that the Einstein equation should possess the same syntactic form under arbitrary coordinate transformations (Norton 1995; Earman 2006), but rather as a substantive requirement that the same laws hold in all reference frames, that is, that they contain no quantities locating the frame with respect to a fixed spacetime background. On this interpretation of general covariance, all arbitrary frames can be equivalent despite being factually distinct.

We agree with this interpretation of general covariance, since we think it true that the metric itself is a dynamical object, rather than a fixed background, as it is in classical mechanics and in special relativity. But, furthermore, we believe that distinction introduced by Dieks can be used in supporting the view that the principle of equivalence, understood in nomological terms, should be extended not only to accelerated motion, but to spacetime singularities as well. For if the same laws that hold in all reference frames are also valid at singularities, and if differences of metrics are factual rather than nomological, then the principle of equivalence can be extended to singularites. As described above, it is indeed the case that the same laws that hold outside of singularities are also valid at singularities, and it seems natural to take the degeneracy of the singular metric as a matter of fact, rather than as a matter of law. So the principle of equivalence can be extended farther than Einstein thought it would.

But some may want to deny that the degeneracy of the singular metric is a matter of fact. However, besides the dynamical character of the metric, which we take to strongly suggest that its degenerate character is factual rather than nomological, here is another reason that could justify this idea. It is known that the metric tensor of a distinguishing spacetime can be obtained from the causal structure (the collection of lightcones) up to a scaling factor (Zeeman 1964; 1967; Malament 1977). The scaling factor can be recovered by knowing a measure, which gives the volume form (Finkelstein 1969). Therefore, at the topological level, ignoring the differential structure, the metric can be expressed equivalently by the causal structure and the measure, which are topologically invariant. It has been shown that the topology of the lightcones is the same at singularities as it is outside them, at least for the Friedmann-Lemaitre-Robertson-Walker big–bang singularity and for the Schwarzschild and Reissner-Nordstrom black hole singularities (Stoica 2015a). The measure is also not manifestly special at these singularities. But if these two structures – the causal structure and the measure, which again depend only on the topology – are not manifestly special at singularities, then how does the degeneracy of the metric arises? The reason is that the differential structure "arranges" the spacetime events on the manifold in a certain way, and forces the metric to be degenerate in some cases. With respect to the causal structure, the lightcones become flattened in some directions of spacetime, although topologically they are equivalent. At the topological level the local diffeomorphisms are replaced by local homeomorphisms. So, while general covariance still preserves the degenerate or regular character of the metric, an extension of general covariance, where local diffeomorphisms are replaced by local homeomorphisms, obliterates any differences between the degenerate and regular metrics, which further supports the idea that the character of the metric is factual rather than nomological. However, extending the field equations, including the Einstein equation, so that they are invariant to homeomorphisms and not merely to diffeomorphisms, is of course still an open problem. The difficulty is due to the fact that all field equations are partial differential equations, which make sense for our current mathematical understanding only in the presence of a differential structure. Nevertheless, it is safe to conclude that even limited to local diffeomorphisms, the laws are the same outside the singularities and at singularities, and the differences between frames are only factual, which reinforces our point above that the principle of equivalence, interpreted in nomological terms, can be extended to singularities.[47]

References

Barbilian, Dan. 1939. "Galileische Gruppen und quadratische Algebren." *Bull. Math. Soc. Roumaine Sci.*, 41(1): 7–64.

Dieks, Dennis. 2006. "Another Look at General Covariance and the Equivalence of Reference Frames." *Stud. Hist. Philos. Sci. B: Stud. Hist. Philos. M. P.* 37(1): 174–191. doi: 10.1016/j.shpsb.2005.11.001

Earman, John. 1995. *Bangs, Crunches, Whimpers, and Shrieks-Singularities and Acausalities in Relativistic Spacetimes*. New York, Oxford: Oxford University Press.

Earman, John. 1996. "Tolerance for Spacetime Singularities." *Found. Phys.* 26(5): 623–640. doi:10.1007/BF02058236.

Earman, John. 2006. "Two Challenges to the Requirement of Substantive General Covariance." *Synthese* 148(2): 443–468. doi:10.1007/s11229-004-6239-x

Eddington, Arthur S. 1924. "A Comparison of Whitehead's and Einstein's Formulae." *Nature* 113(2832): 192. doi: 10.1038/113192a0

Einstein, Albert. 1905. "Zur elektrodynamik bewegter korper." *Annalen der Physik* 322(10): 891–921.

Einstein, Albert, and Nathan Rosen. 1935. "The Particle Problem in the General Theory of Relativity." *Phys. Rev.* 48(1): 73. doi: 10.1103/PhysRev.48.73

Finkelstein, David. 1958. "Past-future Asymmetry of the Gravitational Field of a Point Particle." *Phys. Rev.* 110(4): 965. doi: 10.1103/PhysRev.110.965

Finkelstein, David. 1969. "Space-Time Code." *Physical Review* 184(5): 1261. doi: 10.1103/PhysRev.184.1261

Friedman, Alexander. 1922. Über die Krümmung des Raumes. *Zeitschrift für Physik A Hadrons and Nuclei* 10(1), 377–386.

Friedman, Alexander. 1924. Über die Möglichkeit einer Welt mit konstanter negativer Krümmung des Raumes. *Zeitschrift für Physik A Hadrons and Nuclei* 21(1): 326–332.

[47] The Research Institute at the University of Bucharest, where the paper has been finalized, is gratefully acknowledged by the second author for financial and institutional support.

Friedman, Alexander. 1999. "On the Curvature of Space." *Gen. Relat. Grav.* 31(12): 1991–2000.

Hawking, Stephen W. 1966a. "The Occurrence of Singularities in Cosmology." *P. Roy. Soc. A-Math. Phy.* 294(1439): 511–521. doi: 10.1098/rspa.1966.0221

Hawking, Stephen W. 1966b. "The Occurrence of Singularities in Cosmology. II." *P. Roy. Soc. A-Math. Phy.* 295(1443): 490–493. doi: 10.1098/rspa.1966.0255

Hawking, Stephen W. 1967. "The Occurrence of Singularities in Cosmology. III. Causality and Singularities." *P. Roy. Soc. A-Math. Phy.* 300(1461): 187–201. doi: 10.1098/rspa.1967.0164

Hawking, Stephen W. 1975. "Particle Creation by Black Holes." *Comm. Math. Phys.* 43(3): 199–220. doi: 10.1007/BF02345020

Hawking, Stephen W. 1976. "Breakdown of Predictability in Gravitational Collapse." *Phys. Rev. D* 14(10): 2460. doi: 10.1103/PhysRevD.14.2460

Hawking, Stephen W., and Roger W. Penrose. 1970. "The Singularities of Gravitational Collapse and Cosmology." *Proc. Roy. Soc. London Ser. A* 314(1519): 529–548. doi: 10.1098/rspa.1970.0021

Kupeli, Demir. 1987a. "Degenerate Manifolds." *Geom. Dedicata* 23(3): 259–290. doi: 10.1007/BF00181313

Kupeli, Demir. 1987b. "Degenerate Submanifolds in Semi-Riemannian Geometry." *Geom. Dedicata* 24(3): 337–361. doi: 10.1007/BF00181606

Kupeli, Demir. 1987c. "On Null Submanifolds in Spacetimes." *Geom. Dedicata* 23(1): 33–51.

Kupeli, Demir. 1996. *Singular Semi-Riemannian Geometry*. Dordrecht: Kluwer Academic Publishers Group.

Malament, David. 1977. "The Class of Continuous Timelike Curves Determines the Topology of Spacetime." *Journal of mathematical physics* 18(7): 1399–1404. doi: 10.1063/1.523436

Moisil, Grigore C. 1940. "Sur les géodésiques des espaces de Riemann singuliers." *Bull. Math. Soc. Roumaine Sci.* 42(1): 33–52.

Norton, John. 1995. "Did Einstein Stumble? The Debate over General Covariance." *Erkenntnis* 42(2): 223–245. doi: 10.1007/BF01128809

Penrose, Roger. 1965. "Gravitational Collapse and Space-Time Singularities." *Phys. Rev. Lett.* 14(3): 57–59. doi: 10.1103/PhysRevLett.14.57

Penrose, Roger. 1969. "Gravitational Collapse: The Role of General Relativity." *Revista del Nuovo Cimento; Numero speciale* 1: 252–276.

Penrose, Roger. 1979. "Singularities and Time-asymmetry." In *General relativity: an Einstein centenary survey*, Volume 1, edited by Stephen W. Hawking and Werner Israel, 581–638. Cambridge: Cambridge University Press.

Penrose, Roger. 1998. "The Question of Cosmic Censorship." In *Black Holes and Relativistic Stars*, edited by Robert Wald, 233–248. Chicago, Illinois: University of Chicago Press.

Schutz, Bernard F. 2012. "Thoughts about a Conceptual Framework for Relativistic Gravity." In *Einstein and the Changing Worldviews of Physics*, Volume 12, edited by Cristoph Lehner, Jurgen Renn, Matthias Schemmel, 259–269. New York: Birkhaueser.

Stoica, Ovidiu Cristinel. 2012a. "Analytic Reissner-Nordstrom singularity." *Phys. Scr.* 85(5): 055004. doi: 10.1088/0031-8949/85/05/055004

Stoica, Ovidiu Cristinel. 2012b. "Beyond the Friedmann-Lemaitre-Robertson-Walker Big Bang singularity." *Commun. Theor. Phys.* 58(4): 613–616. doi: 10.1088/0253-6102/58/4/28

Stoica, Ovidiu Cristinel. 2012c. "Spacetimes with Singularities." An. *Şt. Univ. Ovidius Constanța* 20(2): 213–238. doi: 10.2478/v10309-012-0050-3

Stoica, Ovidiu Cristinel. 2012d. "Schwarzschild Singularity is Semi-regularizable." *Eur. Phys. J. Plus* 127(83): 1–8. doi: 10.1140/epjp/i2012-12083-1

Stoica, Ovidiu Cristinel. 2013. "Singular General Relativity." PhD. diss., University Politehnica Bucharest, Faculty of Applied Sciences. https://arxiv.org/pdf/1301.2231.pdf

Stoica, Ovidiu Cristinel. 2014a. "Einstein Equation at Singularities." *Cent. Eur. J. Phys* 12: 123–131. doi: 10.2478/s11534-014-0427-1

Stoica, Ovidiu Cristinel. 2014b. "Metric Dimensional Reduction at Singularities with Implications to Quantum Gravity." *Ann. of Phys.* 347(C): 74–91. doi: 10.1016/j.aop.2014.04.027

Stoica, Ovidiu Cristinel. 2014c. "On Singular Semi-Riemannian Manifolds." *Int. J. Geom. Methods Mod. Phys.* 11(5): 1450041. doi: 10.1142/S0219887814500418

Stoica, Ovidiu Cristinel. 2015a. "Causal Structure and Spacetime Singularities." Preprint. arXiv:1504.07110.

Stoica, Ovidiu Cristinel. 2015b. "The Friedmann-Lemaitre-Robertson-Walker Big Bang Singularities are Well Behaved." *Int. J. Theor. Phys.* 55(1): 71–80. doi: 10.1007/s10773-015-2634-y

Stoica, Ovidiu Cristinel. 2015c. "Kerr-Newman Solutions with Analytic Singularity and No Closed Timelike Curves." *U.P.B. Sci Bull. Series A* 77(1): 129–138.

Strubecker, Karl. 1941. "Differentialgeometrie des isotropen Raumes. I. Theorie der Raumkurven." *Sitzungsber. Akad. Wiss. Wien, Math.-Naturw. Kl., Abt. Ila* 150: 1–53.

Strubecker, Karl. 1942a. "Differentialgeometrie des isotropen Raumes. II. Die Flächen konstanter Relativkrümmung $K = rt — s^2$." *Math. Z.* 47(1): 743–777.

Strubecker, Karl. 1942b. "Differentialgeometrie des isotropen Raumes. III. Flächentheorie." *Math. Z.* 48(1): 369–427.

Strubecker, Karl. 1944. "Differentialgeometrie des isotropen Raumes. IV. Theorie der flächentreuen Abbildungen der Ebene." *Math. Z.* 50(1): 1–92.

Vrănceanu, Gheorghe. 1942. "Sur les invariants des espaces de Riemann singuliers." *Disqu. Math. Phys. București* 2: 253–281.

Zeeman, E. Christopher. 1964. "Causality Implies the Lorentz Group." *Journal of Mathematical Physics* 5(4): 490–493.

Zeeman, E. Christopher. 1967. "The Topology of Minkowski Space." *Topology* 6(2): 161–170.

V. S. Pronskikh
Fermi National Accelerator Laboratory

Original Scientific Paper
UDK 53:167/168

EXPERIMENTAL BACKGROUND AND THEORY-LADENNESS OF EXPERIMENTATION

Abstract: *In this work, I examine the roles of the experimental background (effects capable of mimicking the one under study) in cognition, and its relation to the problem of closedness of experimental system. Taking as examples the experiments in particle physics widely discussed in the philosophy of science (discoveries of muon and neutral currents), I suggest that determination of the experimental background often implies an explicit use of components of high-level theories. I argue that the neutron background in the neutral current experiments resulted from the same sort of phenomena as the events in the detector did although those phenomena occurred in the materials surrounding the detector rather than in the detector itself. Therefore, it is justified herein that due to the presence of background experimental outcomes are entertained with theory of phenomenon.*

Key words: *Social Epistemology; Theory-Ladenness; Experiments; High Energy Physics.*

Introduction

Questions regarding the role of the experimental background in scientific experiments are closely linked to the epistemological question of the relationship between theoretical and empirical in cognition, the problem of the theory-ladenness of experimentation, and the question of whether the experiment is an “open system” (and if so, to what extent). Background is generally defined as a phenomenon in a scientific experiment that can mimic the phenomenon under scrutiny, having a similar appearance even though its nature is different from the one under study. The first philosophical analyses of the experimental background’s role in physics experiments of the XX century belong to Peter Galison (Galison 1987) and Andrew Pickering (Pickering 1984; 1981).

Elucidation of the experimental background’s epistemic role is one of the subjects of a discipline within the philosophy of science called the philosophy of scientific experimentation. Its active development began 30 years ago, although the first work on this topic can be considered the book “The Aim and Structure of Physical Theory” of Pierre Duhem (Duhem [1906] 2007). That work was followed by seventy seven years of practical absence of attention to the problems of scientific experiment, interrupted by a book of Ian Hacking (Hacking 1983).

During the same years (the early– to mid-1980s), the aforementioned works on philosophy, history, and sociology of experiments in high-energy physics by Galison and Pickering were published. These works deal explicitly with the background problem.

At that same time, several works of Allan Franklin on the philosophy of experiments, which were based on modern physics, were published. He also published a later review on the topic of physics experiments in the *Stanford Encyclopedia of Philosophy* (Franklin 2012). In that review, background and its determination were defined as one of the strategies experimenters apply to ensure that their results do not contain errors. In 2003, a collection edited by philosopher Hans Radder was published, which sets out a number of issues emerging in philosophy of scientific experimentation; however, the volume does not consider background to be a separate issue. Nevertheless, the collection analyzed such issues as the nature of the experimental error (this question, as will be shown below, is also associated with the problem of the background), the role of theory, and computer modeling in the experimental practice.

To date, a new edition of Radder's earlier work (Radder 2003) has been published, in which he discusses the issue of material realization, theoretical description of the experiment, and the production of the phenomena in the experiment. The various possibilities for a realistic understanding of the experiment, in particular, formulate the condition of a "closed" experimental system that allows the researcher to treat the background as one of the threats of such "closedness".

In classical experimentation, the problem of background did not attract any significant attention because the distinction of the studied and the other phenomena were deemed self-evident. For example, in the case of Galileo's observation of the planets of the solar system with a telescope, background (phenomena that might have been confused with that under study) could be a spot on the telescope's glass or a planet belonging to another galaxy. Nonetheless, a distinction could be made quite easily. For example, a spot would not be subject to Kepler's law and would thus remain motionless while another planet would have a different character of motion than a planet of the solar system. Despite the simplicity of the practical implementation of such criteria, an assumption about the nature of planetary motion (Kepler's law; i.e., a theory of the phenomenon) is involved in discriminating the background.

However, in his analysis of the experiment of Einstein and de Haas in the first third of the XX century where they measured the gyromagnetic ratio of the electron and his description of experimental efforts to suppress this background source as the Earth's magnetism, Galison notes that, although the major background was the Earth's magnetic field in these experiments, it was still possible to eliminate it through the relatively simplistic design. Nevertheless, controlling the background became the central activity of these experimenters because a large part of their time and effort was spent on the identification of the background and the subsequent struggle to manage it (Galison 1987). That was one of the key observations regarding the nature of physics experiments in the XX century.

Background, coincidences, and causality

The next step in addressing problems of background can be related to cosmic ray research, namely to the experiments designed to detect a muon, which is a new particle (Galison 1987). Experimental physicists used previously developed Geiger-Muller counters connected to electroscopes to show that discharges occurred in the counters when the charged particles passed through them. These discharges could be caused not only by cosmic particles, but also by any other charged particles entering the counter and thus creating background.

To distinguish penetrating muons from other charged particles accidentally entering the counter, experimenters suggested a scheme. They separated two counters with a metal bar. Each counter was then connected to a separate electroscope that measured the electric discharge caused by the passage of particles through it. The experimenters expected that muons, which have higher penetrating power, would penetrate both the counter and the separating block, causing simultaneous operation of both counters. That served as an indication of the studied phenomenon through the coincidence of both counters.

In that experiment, a background manifested itself through a simulating process where the counters operated simultaneously due to the fact that they both were penetrated by independently charged particles so that the simultaneous operation of the counters was random. Thus, the background of random coincidences was created, and the demonstration of the results (muon detection) consisted of measuring an excess number of true coincidences (i.e., those caused by the passage of a muon through both counters) over the number of random coincidences (i.e., those caused by accidental simultaneous entering of both counters by different causally unrelated particles).

This is one of the first descriptions of the experiment that used such a scheme to demonstrate the causal relationships between the experimental results by means of coincidences. It is important that later, with the development of elementary particle physics and nuclear physics, such schemes and arguments grew to be relied upon more often, and in modern experimental physics, those arguments are used frequently. The appearance of such arguments suggests that in quantum physics, experimental devices lose their selectivity with respect to the phenomena that they had in the classical experiment. That fact is primarily due to the features of the studied objects (quantum particles) that are indistinguishable because of the multitude processes in the microworld where objects with similar properties can be created.

Another distinctive feature of the interpretation of the experiment coincidence is the need to distinguish between true (causally related) and accidental (causally unrelated) coincidences for subsequent comparisons and inferences about the presence or absence of the effect sought. One such possibility is the measurement of the number of coincidences in a notorious absence of the studied effect (muons) and its subsequent comparison to the number of coincidences in its putative presence. But for that, one needs to create the conditions under which the studied phenomenon is known to be absent. The basis of this possibility is offered by a theory of the phenomenon and the set of instrumental theories upon which the operation principles of the experimental apparatus are based.

Instrumental theories and background

The term *instrumental theory* here, as well as in the works of Galison and Pickering, is used in a different context than it is in Popper's criticism of the instrumentalist position (Popper 1959). In contemporary experiment, a set of processes occurs in the apparatus intended for the preparation and measurement of a phenomenon that can be described by previously created and validated theories. It is the reliance upon such theories and confidence in their applicability to the interpretation of the processes occurring in the device that allows experimentalists to use the apparatuses as they search for new phenomena and makes a device an apparatus. These theories, upon which the apparatus's operation is based, are referred to as instrumental herein. Franklin refers to the use of well-corroborated (instrumental) theories in apparatuses as one of the epistemic strategies of experimentation (Franklin 2012).

Instrumental theories must be distinguished from theories of the phenomenon under study, upon which the application of instruments is focused. For instance, the theory under scrutiny can be the theory of electroweak interaction discussed here while the operation of the instruments can be described by such disciplines as classical mechanics, thermodynamics, or electromagnetism. A more complicated case in which the instrumental theory is a non-classical theory is discussed below. Nevertheless, as a rule, those are theories other than the theory of the studied phenomenon. The question of application of the theory as a tool is not related to its descriptive properties or to the presence of philosophical issues related to its foundations, and requires only empirical adequacy of this theory in the area in which it is used in such a capacity.

A further possibility is the use of temporal arguments, such as the study of time dependences of the coincidences occurring in the apparatus and conclusion based upon which part of them is random and which part is true (causal). For example, it can be assumed that the operation of the counters is simultaneous (up to the device's resolution) if it is true coincidence (the particle has flown both counters in a raw), or it is not (the counters were penetrated by two non-causally related particles). Such an argument presupposes a theoretical understanding of the processes in which there is background and its certain temporal properties. These theories are, thus, included in the obtained experimental results forming its theoretical components. Compared to the classic experiment, in contemporary high-energy physics we observe an increase in the number of theoretical components included in the experimental results as well as a growth of their complexity.

A qualitatively different example is the experiments on the detection of the muon in cosmic rays carried out with a condensation chamber with a thick lead plate (Galison 1987). The experimenters observed two types of particles and related phenomena in the chamber; one produced particle showers while others left tracks penetrating through it. Instrumental theory, which, in this case, was early quantum electrodynamics, could explain the emergence of particle showers, although its predictions for that energy were unreliable. In this regard, most scientists believed that the showers were caused by a new type of particle while the penetrating particles were deemed to be electrons. However, in the course of further development of the theory of showers, theorists showed that the penetrating particles could not be electrons, and must, therefore, be some new particles. This (Galison 1987) changed the very formulation to the opposite of what it had been. It now became necessary to explain what the shower particles were in the assumption that the new particles were penetrating. In the course of further theoretical analysis, it was revealed that the shower particles were electrons as described by the Bethe-Heitler theory. Thus, the development of theory in the experiment changed the original idea of the observed phenomenon (shower particles—muons, penetrating—electrons) to the opposite (shower particles—electrons, penetrating—muons). In other words, what was considered a phenomenon was proven to be background and what was considered the background became the phenomenon that should be studied and explained. This example (Galison 1987) illustrates how the theory determines what the background and the phenomenon to be questioned are.

However, in the analysis of Galison (Ibid.), a sufficiently clear distinction of what type of theories is meant is not explicitly present. In our account of the theoretical components in the accelerator-based experiments in high-energy physics (Lipkin and Pronskikh 2009), we identified a variety of roles in the structure of the experiment for two types of theories: 1) the theory of the phenomenon (a theory scrutinized in the experiment) and 2) the instrumental theories (upon which the operation of apparatuses is based). For the muon detection experiment, Galison points to quantum mechanics, electrodynamics, and models based on quantum theories as theories determining what the background and the phenomenon are.

In other words, (Galison 1987) did not distinguish between the theory of the muon as the studied phenomenon (what it is and how to interact with it) and an instrumental theory (manifested as interactions of known particles—photons and electrons in the material of the condensation chamber with a metal plate—and the muon appears already as a result of exclusion explanation of the phenomenon), discussing not the former but the latter. Fundamentally new here is the fact that in elementary particle physics, for the first time, a quantum theory plays a role of instrumental theory. The novelty of the fact that a quantum theory is now acting as the instrumental theory can explain the absence of distinction theory types. However, in the modern physics of the microworld, such roles of quantum theories are quite trivial. In the age of the experiments described, such use of quantum theories was the leading edge of science, and the theory of the phenomenon evolved simultaneously and in conjunction with the experiment's technique. That can explain the shift of the interpretation to the opposite.

Background in neutron current experiments

More interesting examples of experiments in high energy physics, which are widely discussed in the literature on the philosophy of science (Galison 1987; Pickering 1984), are the experiments carried out at CERN (European Organization for Nuclear Research) on confirmation of the theory of electroweak interaction (the Glashow-Weinberg-Salam theory), in particular a series of experiments to search for neutral currents . As an analogy, we can look to quantum electrodynamics, where the electron creates the current interaction with the electromagnetic field by emitting or absorbing a quantum of the field (photon). In the electroweak theory of Glashow, Weinberg, and Salam, an interaction, such as the scattering of neutrinos by an electron or nucleon, is also represented as the exchange of virtual quanta of the field—bosons of two types. The charged current (process with a change of the electric charge) in the theory was carried out through the exchange of W+ and W– bosons while the neutral current (process without changing the charge) in that theory was associated with exchange of neutral Z0-bosons.

The experiment on detection of neutral currents performed using an accelerator involves some qualitative differences in addition to significantly increased technical complexity, in particular a growing role of the background. In this experiment, a beam of protons from the accelerator hit the target, and the produced particles passed through a layer of soil, leaving only the muonic neutrino. The flow of neutrinos reached the bubble chamber and caused interactions in it, which manifested themselves particle showers. The reaction products left tracks that were recorded on film.

Confirming the electroweak theory consisted of attributing recorded images to a particular process (emission of charged particles in the decay of neutral Z0– or charged W± boson) and the subsequent statistical analysis. Some neutrinos inevitably propagated in the materials surrounding the camera, such as magnets, flooring, construction materials, and radiation protection, and formed a neutron background inside of it. These background neutrons, which also penetrated the chamber (detection unit of the apparatus), could have caused showers similar to those produced by neutrinos.

Thus, if one of these secondary neutrons collided with a neutron or a proton of a nucleus in the material of the bubble chamber, it might have resulted in a shower of hadrons, which looked similar to one of the possible neutrino events (appearance of the charged W± currents). Another background source were the neutrons emitted along with other hadrons in the neutrino showers, which could subsequently cause secondary showers elsewhere in the camera that were indistinguishable from those made by the neutrino. Because of the presence of a muonic neutrino in previous events, they could have been associated with this event (and thus referred to the background) and, thus, were called associative.

Therefore, Galison’s suggestion that a snapshot of a neutrino-electron scattering event (this process, along with the showers, could be explained by the exchange of Z0-bosons), which became a turning point in the confirmation of the existence of neutral currents and, therefore, provide confirmation of the electroweak theory, was initially interpreted as a manifestation of the background and did not attract any attention (Galison 1987). That was due to the fact that initially weak currents were not key to this theory, and the problem of their search was not set as a priority for the experimenters. Only later, when the search for neutral currents (due to their increased theoretical importance) became significant, their signature was found among the previously obtained images and interpreted as a confirmation of the theory.

Here, many authors, including James Bogen and James Woodworth (Bogen and Woodworth 1988), came to discuss the importance of statistical arguments for the analysis of experimental data. From a qualitative perspective, the background neutrons, which could mimic similar images, appeared from the materials surrounding the cell, meaning the probability of background events was higher near the chamber walls and lower at the chamber's center. Nevertheless, for each specific image, it was impossible to determine whether it was the background or the phenomenon under study. One could only estimate the probability of such an identification on the basis of the model analysis. Similar considerations gave Bogen and Woodward grounds to believe that theories predict phenomena, but not the data (i.e., not individual images or measurements).

Pickering examines the neutral current experiment and its individual phases, and he draws attention to the fact that in the earlier stages of the experiment, the value of the background was calculated by computer simulation (Pickering 1984). Pointing to the imperfection of the phenomenon model the experimenters had used, he argues that before the confirmation of the electroweak theory became the central task of the experiment, the simulation results had been interpreted as evidence of background. A few years later, the simulation results obtained by the same method were considered to be evidence in favor of the observation of neutral currents.

One reason for the ambiguity in the interpretation is the complexity of modeling the entire design of the facility. However, the description of the neutron transport technique underwent a significant development over a few years in 1970s and the accuracy of the model predictions increased. That probably determined the more reliable understanding of the processes taking place in the chamber and the surrounding materials. The complexity of the experimental simulation of large designs in large experiments, as discussed by Pickering, is rather exaggerated, as it is a technical matter rather than a fundamental one.

Background and phenomenal theories

Nevertheless, one of the epistemic problems that was overlooked in the discussions about this experiment was that the background neutrons that partially arose in the materials surrounding the chamber were a result of the same processes as those forming showers in the neutrino scattering within the chamber. Thus, it described the same electroweak theory (either the Glashow-Weinberg-Salam theory or another a priori theoretical conception of what is happening in the process) and, therefore, deployed the theory of the phenomenon at the background identification stage of the experiment.

Having undergone scattering in the materials surrounding the chamber, neutrinos can participate in the process of the exchange of electroweak bosons (which was later confirmed by experiments) in the same way as they can in the chamber substance, giving rise to the neutron background. However, the accuracy of the electroweak theory was demonstrated by analyzing the events in the chamber quantitatively. For such analysis, a quantification of the background in these surrounding materials was necessary, and for that, a preceding theoretical model of the process under study was indispensable.

Thus, to determine the number of background neutrons, it is essential to assess in some way the number of neutrons that occur in accordance with a theory of that phenomenon. The number of events with manifestations of the studied neutral currents is defined as the excess over the background ones, and the measured value of the phenomenon, as noted by Galison, is what remains after subtracting background. Background theory, therefore, ultimately quantitatively enters the result of the measurements. I argue that in addition to the instrumental theories, the experimental results involve prior knowledge of the background model based on the theory of the phenomenon. This fact has not been given sufficient attention in the philosophy of science. Such theory does not necessarily have to be the Glashow-Weinberg-Salam theory, but it has to be some theory (model) of how neutrons originate in the scattering of neutrinos on nucleons (i.e., a theory describing the same scope of phenomena that the electroweak theory itself does).

The occurrence of the instrumental theories in experimental results highlights the general theory-laden nature of the experiment, but does not fundamentally challenge the interpretation of experimental results. However, the theory-laden nature of the phenomenal one allows constructivist interpretation and indicates a potentially important area of research in the philosophy of experimentation. I argue that the experimental background in the example of neutral currents is one of the mechanisms by which the theory of the phenomenon may be included in a significant experimental result.

Including neutral currents in the scope of discussed experiments, Bogen and Woodward (1988), much like their forerunners Galison and Pickering, did not focus on the fact that the theory of the phenomenon and the theory of background were, essentially, the theory of the same phenomenon, concluding (based on the distinction of data and phenomena) that the theory of the phenomenon (the Glashow-Weinberg-Salam theory) was not given consideration in the detection of neutral currents. Deborah Mayo (Mayo 1994) also reduces the analysis of experimental data to consideration of the role of statistical methods and excludes the role of the theory being tested on that basis. Samuel Schindler (Schindler 2014) emphasizes the impact of high-level theory (the theory being tested) on the result of the experiment, in particular the willingness of experimenters to register neutral currents based on a theoretical need rooted in a high scientific value of the Glashow-Weinberg-Salam theory.

Schindler also points out that the credibility of the background modeling in these experiments “depended on the entire set of assumptions” (Ibid., 508) in the calculations (as I show in this paper, the theory of the phenomenon had to be one of those assumptions). In this connection, he claims that it is not necessarily a reliable separation of the background from the signal that served as an argument in favor of the discovery of neutral currents. More important were many other theoretical advantages of the Glashow-Weinberg-Salam theory–, which predicted these currents and influenced the choice of this theory and the discovery of experimental evidence in its support. One such evidence, according to Schindler, was the choice of the magnitude of the background noise allowing one to make the statement about the discovery of neutral current events. Schindler agreed with Galison that background was only possible to estimate in the experiment whereas the lower the estimated value of the background is, the greater the number of events that might have been mistaken for a manifestation of neutral currents. Schindler admits that there is an influence of knowledge by experimentalists of theory prediction on the estimation of background, but without an indication of a direct entry of the scrutinized theory in the background calculations, such experimentalists’ assumptions do not appear decisive. Slobodan Perović considers the arguments in favor of the experiment’s theory-ladenness and calls the processes in the chamber true and the processes caused by background neutrons artifacts while not giving too much importance to the commonality of their physical nature with background (Perović 2013).

In addition, the necessary distinction between the background and noise is not often made. This complicates the interpretation of numerous assertions because the noise in the experimental context is usually associated with electromagnetic oscillatory processes in the electronic parts of the apparatus. Despite the fact that background can, generally speaking, also be the cause of noise, the background discussed by Galison and Pickering is not noise in its most common understanding and is caused by the processes occurring in the scattering of neutrinos. A more detailed examination of the problem of the distinction between background and noise goes beyond the scope of this paper.

Background and experimental closedness

Another question of the philosophy of experimentation that arises in connection with the problem of the background is the question of the “openness” of the experimental system (i.e., whether it is possible to take all factors affecting the experimental result into account). It is concerned with the question of whether the system may experience unpredictable processes that can change the outcome for which the experimenter will not be able to account. This question was considered by Pickering (Pickering 1981) in relation to experiments in the search for free quarks using a facility similar to that used in the Robert Millikan experiments. Pickering studied the work of the two groups of experimenters who have come to different conclusions as to the existence of fractional electric charges. During the experiments, the groups found, in themselves and each other, all the new factors that influence the results and identified problems in experimental techniques that they could not solve.

In that regard, Pickering raised the question of whether it is possible, in principle, to take into account all the processes that determine the experiment's outcome. Radder (Radder 2012), regarding the issue of closedness, introduces the concept of the theoretical description of experiment that echoes with the notion of theoretical components used herein. He believes that the closed experimental system is only possible under a given theoretical description. In addition, one has to avoid or take into consideration only those factors that can affect a certain outcome. As a result, an experiment was looking for an answer to very specific questions.

Radder formulates two conditions of a closed experimental system: 1) all episodes taking place in a system should not have sufficient conditions outside the system, and 2) all necessary conditions of reproducibility outside the system must be met. As we can see in relation to the problem of background, the first of these conditions cannot always be fulfilled because, for example, a shower in the camera can be caused by background neutrons of certain energy, which is a sufficient condition for the occurrence of the shower, just like the presence of neutrinos. Thus, the major problem regarding background is the lack of justification for the absence of unknown background sources in the experiment.

The presence of background violates the closedness of the experimental system, but unlike the experiments performed before the first third of the XX century, it is impossible to eliminate in modern experiments by design. This creates, as discussed above, a sort of theory-ladenness that can be associated with the design of experiments and, therefore, can be called performative. However, in addition to exclusion by design, Galison also considered such methods of accounting for background as its measurement and calculation.

Conclusion

As in the experiment on the detection of neutral currents, because of the fact that the background neutrons were created by the neutrino flux and were also the phenomenon under study, it was impossible to measure the background in the chamber separately from that phenomenon. Nevertheless, it is often possible to calculate the background, and it is essentially the same phenomenon as that under scrutiny. Given that in these experiments, the means of that calculation were created and developed simultaneously with the experiment, the results were not always stable, which does not create any difficulty in principle. It is obvious, however, that such means should include any a priori information about the process under study because it entails the creation of the background neutrons in the chamber.

Thus, the epistemic role of background processes that mimic the physical phenomenon under scrutiny is one of the important and widely discussed issues in the philosophy of experimentation. As I show in this study, background can, in addition to its other roles, introduce the theoretical components of high-level theories in the experiment's outcome. The error associated with underestimation of background (e.g., its unknown nature or magnitude) in the experiment can be attributed to the so-called systematic errors (i.e., errors of a non-statistical nature caused by insufficient knowledge of the experimental system). The complexity of the background issue indicates the need for a detailed epistemological analysis of modern experiments in physics and other sciences to examine the role of background in knowledge production.

References

Bogen, James, and James Woodward. 1988. "Saving the Phenomena." *The Philosophical Review* 97(3): 303–352.

Duhem, Pierre. [1906] 2007. *La Théorie Physique. Son Objet, sa Structure*. Paris: Chevalier & Riviére. New Edition, Paris: Vrin.

Franklin, Allan. 2012. "Experiment in Physics." In The Stanford Encyclopedia of Philosophy (Winter 2012 Edition), edited by Edward N. Zalta. Stanford: Metaphysics Research Lab, Center for the Study of Language and Information, Stanford University. http://plato.stanford.edu/archives/win2012/entries/physics-experiment/.

Galison, Peter. 1987. *How Experiments End*. Chicago: University of Chicago Press.

Hacking, Ian. 1983. Representing and Intervening: Introductory Topics in the Philosophy of Natural Science. Cambridge: Cambridge University Press.

Lipkin, Arkadiy I, and Vitaly S. Pronskikh. 2009. "Interlacing of Theory, Experiment and Instrument in Accelerator-based Experiments: The "Theoretical-operational" Model." *Investigated in Russia* 44: 511–521.

Mayo, Deborah G. 1994. "The New Experimentalism, Topical Hypotheses, and Learning from Error." In *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association* 1994(1): 270–279.

Perović, Slobodan. 2013. "Theory-driven Experimentation in Particle Physics." *Belgrade Philosophical Annual* 26: 51–61.

Pickering, Andrew. 1981. "The Hunting of the Quark." *Isis* 72(2): 216–236.

Pickering, Andrew. 1984. *Constructing Quarks: A Sociological History of Particle Physics*. Chicago: University of Chicago Press.

Popper, Karl. 1959. *The Logic of Scientific Discovery*. London: Hutchinson.

Radder, Hans. 2003. *The Philosophy of Scientific Experimentation*. Pittsburgh: University of Pittsburgh Press.

Radder, Hans. 2012. *The Material Realization of Science: From Habermas to Experimentation and Referential Realism* (Boston Studies in the Philosophy and History of Science, Book 294). Dordrecht: Springer.

Schindler, Samuel. 2014. "A matter of Kuhnian theory-choice? The GWS model and the neutral current." *Perspectives on Science* 22(4): 491–522.

Milan M. Ćirković[*]
Astronomical Observatory of Belgrade,
Volgina 7, 11000 Belgrade, Serbia

Original Scientific Paper
UDK 167/168:165
520/524
113/119

ANTHROPIC ARGUMENTS OUTSIDE OF COSMOLOGY AND STRING THEORY

Abstract: *Anthropic reasoning has lately been strongly associated with the string theory landscape and some theories of particle cosmology, such as cosmological inflation. The association is not, contrary to multiple statements by physicists and philosophers alike, necessary. On the contrary, there are clear reasons and instances in which the anthropic reasoning is useful in a diverse range of fields such as planetary sciences, geophysics, future studies, risk analysis, origin of life studies, evolutionary theory, astrobiology and SETI studies, ecology, or even strategic studies and global policy. The association of anthropic reasoning with string theory and particle cosmology has not only become the standard wisdom, but has been often construed in a negative way, in order to demonstrate or insinuate that such reasoning is too abstract or even belongs to "fringe science", remote from run-of-the-mill research practices in any other more "mundane" and less theoretical scientific discipline. The purpose of this paper is to (i) analyse some of the counter-examples to the standard wisdom, (ii) suggest that the anthropic reasoning is more flexible, more general, and less fashion-driven than the detractors state. In addition, we consider some historical and/or extrascientific motivation for this persistent prejudice.*

Keywords: *history and philosophy of physics, epistemology, anthropic reasoning, planetary science, risk analysis, bioethics*

1. Introduction: Why Anthropic Reasoning?

Anthropic reasoning, understood in a loose enough sense, has been present in the history of ideas since ancient times.[1] In modern science, it appears in Boltzmann's famous 1895 retort to Zermelo on the topic of the origin of the Second Law of Thermodynamics, or the temporal asymmetry of the universe:

> If we assume the universe great enough, we can make the probability of one relatively small part being in any given state (however far from the state of thermal equilibrium), as great as we please. We can also make the probability great that, though the whole universe is in thermal equilibrium, our world is in

[*] Also at Future of Humanity Institute, Faculty of Philosophy, University of Oxford, Suite 8, Littlegate House, 16/17 St Ebbe's Street, Oxford, OX1 1PT, UK. E-mail: mcirkovic@aob.rs.

[1] For some specific instances, see Ćirković 2004.

> its present state. It may be said that the world is so far from thermal equilibrium that we cannot imagine the improbability of such a state. But can we imagine, on the other side, how small a part of the whole universe this world is? Assuming the universe great enough, the probability that such a small part of it as our world should be in its present state, is no longer small.
>
> If this assumption were correct, our world would return more and more to thermal equilibrium; but because the whole universe is so great, it might be probable that at some future time some other world might deviate as far from thermal equilibrium as our world does at present. Then the afore-mentioned H-curve would form a representation of what takes place in the universe. The summits of the curve would represent the world where visible motion and life exist (Boltzmann 1895, 415).

The issue at hand was whether the thermodynamical asymmetry (popular "arrow of time") is a local or truly global, universe-wide phenomenon. The idea behind Boltzmann's – or Boltzmann-Schuetz's – approach was to explain the local thermodynamical disequilibrium by appealing to the size of the universe and the conditions necessary for our existence as living beings and observers. What can be "more anthropic" than this? The whole polemic is interesting not only from the point of view of history of science (Steckline 1983), but also since Boltzmann's ideas could be recast in the modern multiverse context (Ćirković 2003). However, it went largely unnoticed for quite a long time, and Boltzmann's inception of the anthropic reasoning was recognized only about the time of the publication of the comprehensive monograph of Barrow and Tipler (1986).

In the orthodox history of science, it was Dicke's (1961) rejoinder to Dirac's (in)famous "large-number hypothesis" (LNH) which first takes into account explicit condition for existence of observers at present.[2] The locution "anthropic principle" comes from the study of Carter (1974), which defined anthropic principles and went some way to impose some order on the nature and quality of anthropic arguments used in cosmology thus far. For most people and sources, Carter's study is the ultimate origin of anthropic reasoning, although it is already clear from the above (and detailed, although hardly unbiased, historical survey could be found in Barrow and Tipler 1986) that it only followed from previous considerations.

From the start the debate revolved about the infamous problem of fine-tuning of fundamental constants of nature and cosmological parameters. While Dirac argued that some numerical relationships between those (understood as pure numbers, in order to avoid confusion stemming from man-made units of measurement) hold in the absolute sense throughout the history of the universe, Dicke has turned the explanatory issue "upside down" by asking when and how such approximate relationships could be *observed*. For example, we observe stars of age similar – within an order of a magnitude – to the age of the universe itself, since we are observers dependent on stars and stellar evolution for our existence. Either much earlier or much later, itwould be impossible that observers like us

[2] See Dirac 1937; 1961.

evolve in the universe, thus there is no surprise that we observe approximate relationship between these timescales. This conclusion of Dicke could be generalized to other fine-tuning "coincidences", as has been subsequently clarified. In any case, Dicke pioneered an approach which is *evolutionary* in nature, relying on the concept of observership, rather than (at least partly) metaphysical concepts of absolute laws.

So, historically, anthropic reasoning *has* indeed emerged from cosmological considerations (for further historical summary see Ellis 2011). But it went much farther than this circumscribed domain in the 40+ years since Carter's seminal work. And there is really no reason why it should not have, since what Carter himself understood well (e.g., Carter 1983; 1993), but what subsequent critics *and* supporters often failed to understand and appreciate, anthropic reasoning deals with *observation-selection effects*, and has no teleological meaning whatsoever. It simply applies whenever the number or other properties of observers come into play and influence particular scientific result or claim. This disteleological nature of the anthropic reasoning has been best explicated in the seminal monograph by Bostrom (2002).

Thus, it follows from the above outlined general argument that in many fields of science in which observation selection is likely to play some role, it is possible to use anthropic reasoning. An obvious example is biological evolution, which is itself the process through which observers have emerged so far.[3] Obviously, the number and properties of observers at present impose constraints – possibly stringent constraints – on processes such as abiogenesis and subsequent unfolding of evolution. Another relevant domain is computer science, including the fields such as complexity theory, artificial intelligence (AI). About the same time when Dicke replied to Dirac's LNH, the great pioneer of computer science, Norbert Wiener, published his famous *Cybernetics*. It contains a beautiful discussion of the observation-selection effects in the problem of the arrow of time:

> Our observations of the stars are through the agency of light, of rays or particles emerging from the observed object and perceived by us. We can perceive incoming light, but can not perceive outgoing light, or at least the perception of outgoing light is not achieved by an experiment as simple and direct as that of incoming light. ... This being the case, we can see those stars radiating to us and to the whole world; while if there are any stars whose evolution is in the reverse direction, they will attract radiation from the whole heavens, and even this attraction from us will not be perceptible in any way, in view of the fact that we already know our own past but not our future. Thus the part of the universe which we see must have its past-future relations, as far as the emission of radiation is concerned, concordant with our own. The very fact that we see a star means that its thermodynamics is like our own... Within any world with which we can communicate, the direction of time is uniform (Wiener 1961, 34–35).

[3] Without prejudice that this is how all or even majority of observers emerge. For instance, our current fledgling AI efforts might be crowned by success in the future and it is quite possible that most of the observers in the universe are of machine origin (e.g. Dick 2003).

Today, post-Barrow and Tipler and post-Bostrom, it is not difficult to recognize anthropic reasoning in Wiener's discourse. This is not a "Whiggish" interpretation of history – to forestall such criticism – since the issues Wiener discusses are still parts of open research programs and heterodox strategies should be not only permitted, but desirable; there is no established view which could enforce misreading of the historical evidence.

Therefore, in stark contrast to the prevailing belief, anthropic arguments have an important tradition, and a role to play *outside* highly speculative fields of string theory and particle/quantum cosmology. In what follows I review some examples and argue for the role for anthropic arguments and explanatory reasoning much larger than hitherto assumed. The primary goal of this study is to dispense with the prejudice that anthropic reasoning is limited to extremely general and "abstract" or "remote" fields of particle cosmology and string theory. Detractors of anthropic reasoning (and they are legion; e.g., Maddox 1984; Earman 1987; Gould 1987; Wilson 1994; Pagels 1998; Manson 2000; Mosterín 2000; 2005; Klee 2002; Smolin 2007; to mention just a few examples[4]) have traditionally argued that anthropic reasoning is quite limited in scope, and while most of them have not pressed this scope limitation as the main reason for rejecting the anthropic reasoning, the meme has spread around. To show that it is wrong, therefore, and that the anthropic reasoning is much broader in scope as an explanatory strategy, leads to weakening the entire sceptical case. In particular, it includes weakening the sceptical strategy in the domain of fundamental physics, where currently the greatest battles are waged.

Let us consider just a couple of examples of the sceptical discourse. Earman (1987) writes in his well-publicized – and largely outdated – critique of the anthropic reasoning:

> But to be legitimate, the anthropic reasoning must be backed by substantive reasons for believing in the required worlds-within-worlds structure... Neither classical general relativity nor quantum mechanics provide any firm grounds for taking worlds-within-worlds models seriously, and while various speculative versions of the new inflationary cosmology may eventually provide such ground, the verdict is at present very much in doubt (Earman 1987, 316).[5]

Note that not only is the application of the anthropic reasoning strictly circumscribed in Earman's account, but specific advances in cosmology are required as *prerequisites* for anthropic explanations.

Pagels goes along:

> My own view is that although we have not yet discovered the most basic physical laws, if we do, the possibility of life in a universe governed by those laws will in some sense be written into them. The existence of life is not a selective principle acting on those laws; rather, it is a consequence of them. Whether or

[4] Not to mention the cottage industry of anti-anthropic web sites or blogs, e.g., Woit 2004–2016; Motl 2004–2017.

[5] Here, one should note that what Earman calls "worlds-within-worlds" is what we nowadays call the multiverse.

> not I am right, it is simply premature to invoke the anthropic principle until the origin of the universe is much better understood (Pagels 1989, 185).

So, we have a clear pattern here: it is alleged that the anthropic reasoning makes at least some sense in cosmology, but is strictly rejected in other fields. This limitation is then used to diminish and denigrate the usage of anthropic reasoning *even in cosmology.* Note that Pagels' argument is at best a special pleading: that the possibility of life is allowed by the most basic laws is a truism, since what else could it be (so long as we are metaphysical naturalists)? At worst, it is an outright straw-man argument, since we would be hard pressed to find anybody suggesting that the existence of life is not *a* consequence of the fundamental laws of physics. The indefinite article betrays Pagels' crucial omission – what he fails to address is the question how could life be *a* consequence of the most fundamental laws and not *the* consequence of them. The most fundamental laws – whatever precisely they are – almost certainly permit both universes without life and those with life; since we are obviously in the latter category, it is uncontroversial that our existence acts as a "selective principle" (to use Pagels' charged terminology) not on laws themselves, but on *instantiations of these laws*; not on the equations of dynamics, but on the multiple solutions of those equations. To believe that in contrast to all other theories ever devised in physics these equations will have just a single solution, smacks of unsupported quasireligious faith. If they have multiple solutions then the question '*Why do we observe solution A, rather than solutions B, C, ...?*' is entirely legitimate,scientific, pertinent, and unavoidable problem. So, Pagels' criticism is, in the final analysis, based on a *quid pro quo* (cf. Barnes 2012).

Further, Mosterín in a spectacularly hostile review of the anthropic thinking argues:

> The universe (as far as we can know it) is something unique ('uni-verse' and 'uni-que' come from the same root 'uni', one). We can learn a posteriori how the universe is, but it makes no sense to speculate on how it should be on the basis of a priori statistical considerations. This is the reason why John Leslie's (1989) firing squad argument is flawed. He compared our existence to the survival of a sentenced man, because each of the guns in his execution squad misfires. Has someone (God?) tinkered with the guns beforehand? Of course, there have been lots of firing squads and seldom have all the guns misfired. There is a grim statistics of firing squads. But the universe is a unique historical fact. There are no statistics of universes. Besides, the components of the firing squad are people with the intention of shooting, but there are no intentions in the fabric of the universe. At least in usual language, fine tuning implies intentionality and multiplicity of cases (Mosterín 2005, 448XX).

Here we can follow the reasoning of the detractors a few steps further: since it is not *really* acceptable *even* in cosmology, then we do not need to consider anything further in declaring the anthropic reasoning completely bogus and devoid of sense. Some of the gaping holes in Mosterín's argument are obvious (metaphysical commitment to a single universe, *quid pro quo* reference to the "usual language" in describing fine tunings, etc.), others are somewhat more subtle (commitment to a verificationist account of truth), yet others stem from ignorance or suspicion toward recent developments in science, especially particle and quantum cosmology and string theory. The best recent rebuttals of the claims of Mosterín and comp. are given by Carroll (2006) and Barnes (2012) in a technical manner, and by the great Leonard Susskind in his popular book (Susskind 2006). Part of the arguments Mosterín uses against the anthropic reasoning is from an auto-ironically imprecise paper by Klee (2002); see the comprehensive rebuttal in Walker and Ćirković (2006). However, the detractors gained much traction in science, philosophy, and popular media. For example, anti-anthropic papers are much easier to publish than pro-anthropic ones. Even corrections of obvious *numerical errors* in the anti-anthropic literature are lacking (as testified by the never corrected errors regarding the size and mass of the Milky Way in Klee's paper), while detractors like Martin Gardner or Pagels relished in pointing out even trivially simple mistakes/typos in the encyclopaedic monograph of Barrow and Tipler.

Nevertheless, the criticism often sounds better than it really is, since it invokes abstract and remote issues of cosmology and fundamental physics. A few people feel comfortable discussing "the most basic physical laws" or the issues of cosmological origins, since there is no intuitive grasp of them. Therefore, it pays off for the opponents of the anthropic reasoning to promote the dogma that this reasoning is inextricably linked to fundamental physics and cosmology. So, there is not only rhetorical but a game-theoretical motivation for the anti-anthropic campaign. In the following, I shall use examples from planetary science and risk analysis in order to refute the dogmatic prejudice that the anthropic reasoning is limited to highly abstract and non-intuitive domains of fundamental physics and cosmology. Emancipation from the dogma leads not only to our increased explanatory freedom, but also to some quite practical consequences, as the examples from the domain of risk analysis will show. Therefore, the whole of this debate has profound ethical ramifications, which is too often conveniently ignored by the mainstream philosophers. I also speculate on the origin of hostile prejudices about anthropic reasoning in the concluding section.

2. Anthropic arguments in planetary sciences and astrobiology

In the domain of planetary sciences which experienced explosive growth in the last couple of decades there is a substantial room for anthropic argument. To a large degree, the development of planetary sciences, both within Solar system, and as applied to the large number of extrasolar planets, has been fuelled by the quest for *habitability* (Horneck and Rettberg 2007; Hanslmeier 2009). According to NASA's *Astrobiology Roadmap*:

> We must move beyond the circumstances of our own particular origins in order to develop a broader discipline, "Universal Biology." Although this discipline will benefit from an understanding of the origins and limits of terrestrial life, it also requires that we define the environmental conditions and the chemical

> structures and processes that could support life on other habitable planets. Thus we need to exploit universal laws of physics and chemistry to understand polymer formation, self-organization processes, energy utilization, information transfer, and Darwinian evolution that might lead to the emergence of life in planetary environments other than Earth. ... Some conditions that support chemistry that is sufficiently rich to seed life might be detrimental to self-organization of biological structures. Conversely, conditions that promote the emergence of biological complexity might be unfavorable to organic chemistry. Thus an integrative, interdisciplinary approach is necessary to formulate the principles underlying universal biology. The perspectives gained from understanding these principles will markedly improve our ability to define habitability and recognize biosignatures (Des Marais et al. 2008, 720).

And subsequently:

> Viewing Earth's ecosystems in the context of astrobiology challenges us to consider how "resilient" life really is on a planetary scale, to develop mathematical representations of stabilizing feedbacks that permit the continuity of ecosystems in the face of rapidly changing physical conditions, and to understand the limits of these stabilizing feedbacks. Ideally this consideration will provide insight into the potential effects of environmental changes that are abrupt as well as those changes that unfold over time scales ranging from seasonal cycles to millions of years (Ibid., 727).

Therefore, probability of observing specific properties of an environment – *any* environment in the Galaxy! – are linked with the number of observers, whose density is a consequence of complex evolutionary processes which could not be reasonably accounted for in the conventional causal manner. While planetary (to simplify things from the start![6]) parameters are certain to play the determining role in the degree of habitability, they cannot be said to cause it in proximate sense. The relationship is similar to the one between fundamental physical constants like Planck constant, coupling constants of forces, or masses of quarks and leptons in the Standard Model on one side, and properties of the ^{12}C nucleus exhibiting famous fine-tuning discovered by Sir Fred Hoyle (Dunbar et al. 1953; Hoyle 1954; Barrow and Tipler 1986) on the other side. In theory, fine-tunings of energy levels in the ^{12}C nucleus are causally explained by properties (and, consequently, fine-tunings) of fundamental constants. If we accept ontological reductionism there is no other mystical factor there – the ^{12}C nucleus consists of quarks and gluons obeying dynamical laws prescribed by quantum physics and nothing else, so its properties, including "controversial" ones, are reducible to the properties of the constituents. In practice, however, the ^{12}C nucleus is too complex and too strongly-interacting quantum system to be understood in terms of these basic ingredients.[7] So we cannot track the fine-tuning in the ^{12}C nucleus down to the fine-tuning of fundamental constants. In practice we are treating the fine-tuning in ^{12}C as a distinct phenomenon.

[6] Not necessarily rejecting the possibility that there are different and more radical kinds of habitats (asteroids, small icy bodies, molecular clouds, etc.).

[7] For modern at-tempts to do so, see for example Freer and Fynbo 2014.

The analogy with other complex systems which are also prerequisites for life as we know it is obvious. In particular, habitable planets are complex outcomes of many lower-level factors of planetary sciences in a manner still more complicated than the ^{12}C nucleus being complex outcome of many lower-level factors of quantum/particle physics. So we again cannot hope to track down apparently fine-tuned high-level properties to a set of basic, low-level planetary properties such as the total mass, chemical composition, distance from the parent star, etc. Hence, the very concept of habitability is inextricably linked to the anthropic reasoning – it summarizes physical prerequisites for particular spacetime density of observers. It is unimportant here that we tend to take this concept as unnecessarily narrow – it is not really required that the life we expect to emerge in habitable locales be exactly like the terrestrial one, or even very similar. We speak about potential for habitability on Jupiter's moon Europa below the thick ice crust, although we can be pretty certain that any Europan forms of life will have to be very different from anything we have encountered on Earth. So, instead of being anthropocentric, as criticisms directed at astrobiology often charge, the notion of habitability actually represents a scientific attempt to liberate ourselves from anthropocentrism while retaining the advantages of scientific method as developed among terrestrial biologists.

In order to do so, understanding of preconditions leading to life, complex life and observership are necessary – and it is exactly the essence of anthropic reasoning such as contained in the works of Dicke or Bostrom or Barnes (and parallel with the one developed in cosmology and fundamental physics by Carter, Linde, Susskind, Tegmark, Hogan, Carroll, and others). An additional difficulty which makes the application of anthropic reasoning here more difficult lies in the feedback loops created by even very simple lifeforms, in which they influence their physical environment. This is in sharp contrast with the situation in fundamental physics and cosmology, where there is no feedback loop between life and observers on the one side and physical systems under study on the other (spatial and temporal scales are either too small or too large for observers to have meaningful impact). It is well-known, for example, that abiogenesis as currently understood could have only occurred in an anaerobic conditions; the present amount of oxygen in the atmosphere is solely a consequence of the evolution of photosynthetic life forms and their establishing dominance over methanogen and other contemporaries (e.g., Horneck and Rettberg 2007). In turn, the aerobic environment predominant on Earth since the "Great Oxygenation Event", roughly 2.3 billion years before present, enabled emergence of a huge number of complex organisms, some of which have been the cause of further feedback loops in the environment. While these feedbacks play crucial role in the classic Gaia hypothesis and are still very much in play as explanatory vehicles (a great example being recent study of Chopra and Lineweaver 2016), we need not endorse any such general theory in order to perceive how habitability is highly emergent property for whose explanation all naive and greedy reductionist schemes simply fail. In their stead, we need at least something akin to the anthropic reasoning.

A specific model is provided by the "rare Earth" hypothesis of Ward and Brownlee (2000). In a nutshell, it is a probabilistic argument suggesting that, while simple microbial life is probably ubiquitous throughout the Galaxy, complex biospheres, like the terrestrial one, are very rare due to the exceptional combination of many distinct requirements for high habitability. These requirements have become familiar to this day even in popular science accounts: **Circumstellar habitable zone** (a habitable planet needs to be in the very narrow interval of distances from the parent star in order to possess liquid water on surface), **"Rare Moon"** (having a large moon to stabilize the planetary axis is crucial for long-term climate stability), **"Rare Jupiter"** (having a giant planet ('Jupiter') at the right distance to deflect much of incoming cometary and asteroidal material enables a sufficiently low level of impact catastrophes), **"Rare nuclides"** (radioactive r-elements – especially ^{238}U and ^{232}Th – need to be present in the planetary interior in sufficient amounts to enable plate tectonics and the functioning of the carbon–silicate cycle), etc. There are other items on the list as well – in this sense, "rare Earth" hypothesis is an open theoretical system, since everyone is free to add items pertaining to a particular area of relevance and expertise. However, the general reasoning is that all these requirements are mostly independent and *a priori* unlikely, so that their combination is bound to be incredibly rare and probably unique in the Milky Way. Without entering into pros and cons of this important hypothesis (and there are many items on both sides of the ledger!), we need to understand that there is nothing contradictory between the conclusion that complex biospheres are extremely rare and the empirical observation that we are living in a complex biosphere. Contradiction is removed by – lo and behold! – the anthropic reasoning. It is the anthropic reasoning which accounts for the Bayesian probability shift: since we require complex biosphere to exist as observers with an extremely high probability,[8] we need to update our probabilistic beliefs accordingly.

In addition, the rare Earth paradigm assumes something about the *duration of time* necessary for the processes leading to complex biospheres to complete. This is the topic of another empirical success of the anthropic reasoning, namely Carter's relation, which predicts the probable length of time the Earth will remain a habitable planet from the number of improbable steps that occurred in the evolution of intelligent life on Earth (Carter 1983). If the total habitable lifetime of Earth, both past and future is denoted by T and the time necessary to evolve intelligent observers like us is t_e, Carter argues that it should hold that, where n is the number of "critical steps" (or "improbable steps" or "key transitions") in the evolutionary process. Clearly, a relation between these numbers is neither obvious nor trivial – without going into whether it is verified or falsified, *it certainly is an empirical claim*. Yet Carter's relation is a consequence of the traditional "weak" anthropic principle: self-selection from our nature as intelligent observers which evolved on a habitable planet.

[8] Not exactly necessity of unity probability, since we have to account for a minuscule probability of our being Boltzmann brains or our observations of complex biosphere around us being deceitful.

Finally, one might argue that the first anthropic argument in planetary sciences appeared centuries ago, in the work of French philosopher and naturalist Bernard Le Bouyier de Fontenelle (1657–1757), published in 1686, a single year before the great scientific revolution inaugurated by Newton's *Principia*.[9] It deals with habitability of *planet Earth*, and is contained in a single paragraph of his *Conversation on the Plurality of the Worlds*. It reads:[10]

> In the next places, the reason why the planes of their [comets'] motions are not in the plane of the ecliptic, or any of the planetary orbits, is extremely evident; for had this been the case, it would have been impossible for the Earth to be out of the way of the comets' tails. Nay, the possibility of an immediate encounter or shock of the body, of a comet would have been too frequent; and considering how great is the velocity of a comet at such a time, the collision of two such bodies must necessarily be destructive of each other; nor perhaps could the inhabitants of planets long survive frequent immersions in the tails of comets, as they would be liable to in such a situation. Not to mention anything of the irregularities and confusion that must happen in the motion of planets and comets, if their orbits were all disposed in the same plane (Fontenelle 1990 [1868], 466).

Thus, to the question: *why are (observed) orbits of comets highly inclined, in contradistinction to the coplanar planetary orbits?* Fontenelle offers a deceptively simple answer. We would not be here – to contemplate on the peculiarities of cometary trajectories – if these orbits were different (that is, similar to those of planets). In modern rendition, our existence in a stable environment selects some planetary+cometary configuration out of the entire set of all such configuration possible under the dynamical laws. Therefore, there is no reason to believe that the observed planetary+cometary configuration will be typical or average. This passage of Fontenelle had been published 8 years before celebrated Halley's suggestion of December 12, 1694, that comets might collide with planets:

> This is spoken to Astronomers: But, what might be the Consequences of so near an Appulse; or of a Contact; or, lastly, of a Shock of the Coelestial Bodies, (which is by no means impossible to come to pass) I leave to be discuss'd by the Studious of Physical Matters (published in Halley 1704, 24).

This famous idea has been followed up by such luminaries as Newton, Wright, Laplace, Lagrange, and others, in the vein of what is usually (and only partially justifiable) called "Biblical catastrophism" of the pre-Cuvier epoch.[11] Fontenelle wrote the passage 18 years before Newton wondered (in *Opticks*):

> [W]hence is it that planets move all one and the same way in orbs concentrick, while comets move all manner of ways in orbs very excentric... blind Fate could never make all the planets move one and the same way in orbs concentrick, some inconsiderable irregularities excepted, which may have risen from the mutual actions of comets and planets upon one another, and which will be apt to increase, till this system wants a reformation. Such a wonderful

[9] More details in Ćirković 2002.

[10] According to the 1990 translation by H. A. Hargreaves.

[11] For colourful pieces of its history, see Clube and Napier 1990.

> uniformity in the planetary system must be allowed the effect of choice (Newton 1730, 344, 378).

Thus Newton, as a promoter of the Design argument in natural philosophy, failed to understand the power of the Fontenelle's argument, and went deeper into a blind alley (from the modern point of view) of seeking the supernatural Design and/or supernatural regulating mechanism. The same tension between the apparent design and the explanatory "filtering" through various observation selection effects persists to this day, and is the source of innumerable debates and confusions.[12]

And the issue of perception of peculiarity of the inhabited subset of planets in the entire set is a legitimate target of physical inquiry – the one which is directly guided by anthropic reasoning. Again, like in the cosmological case, we have three possible options of explaining the peculiar (non-planar) nature of cometary orbits in our Solar system. The first is to deny the validity and meaningfulness of the question; this is the standard theistic answer which forbids further discussion. In the less direct and allegedly more philosophical manner, we could speak about this state-of-affairs as being a brute fact. The theistic answer can hardly be further discussed, and the explanatory nihilism/brute-factism seems irrational today, since we know that the Solar system forms a tiny part of a much larger whole. As to the origin and properties of this larger whole (the Galaxy) we do have different (and working!) explanations – it would be truly strange to expect that the property of minuscule planetary system like ours cannot, even in principle, be explained.[13]

From the other two options, one—causal—entails the idea that there is a law-like reason (presumably to be derived from the future "Theory of Everything" or some other high-level physical theory) for atypical or surprising structure of the early Solar system. In other words, an enormous amount of information necessary for description of the atypical initial conditions can be encoded in some new law(s) of nature and consequent law-like correlations of various matter and vacuum fields. This option is still viable for cosmology, but hardly for planetary cosmogony. *Cosmogonical* initial conditions are not privileged in any way over initial conditions for any other physical process; we do not seek an explanation for (say) the formation of chemical elements or the formation of the ozone layer or for origin of ice ages in a future unified field theory. We seek it much lower down on the epistemological ladder – and higher in terms of complexity of the explanans. Moreover, the very existence of such higher-level theory to which the explanatory work could be outsourced seems highly dubious, again in the light of the necessity for such theory to explain other planetary systems in the Galaxy as well. Some of them – and perhaps most of them – seem to be very different from the Solar System (e.g. Kepler-16b) (Doyle et al. 2011).

[12] For example of the curious mixture of teleological and physical thinking characteristic of the early modern era, see Halley 1726.

[13] I shall return to the issue of explanatory nihilism in the section 4 below.

The third—anthropic—option is to avoid giving a specific description through embedding those conditions into a sufficiently symmetric ("typical") background. Again, stated in terms of information, the same long description of what we perceive as atypical initial conditions arises—as so often in physics!—from the process of symmetry breaking. The overall description is simple enough, and may be reduced (in the extreme case) to a rule similar to "All possible combinations of initial conditions exist." That such a high degree of symmetry can indeed completely reproduce the situation in our particular cosmological domain becomes an immediate consequence (cf. Tegmark 1996; 2008; Collier 1996). Observation selection effect then accounts for features of the observed system required for the local observers to exist.

Let us now think of the application of this mode of thinking to the particular cosmogonical issue. We would similarly say "planetary systems with all possible configurations of planetary and cometary orbits exist" as a part of the larger whole (say our Galaxy). Now, we ask: are all such configurations compatible with our existence (on a planet!)? And the answer, intuitively clear even to Fontenelle, with his rudimentary understanding of preconditions for complex life and intelligence, seems to be negative. Planetary "irregularities and confusions" would have likely resulted in the absence of all observers, contrary to the empirical evidence. There is only a subset of all configurations leading to the emergence of us as intelligent observers, namely the one in which collisions between planets and smaller bodies (comets and asteroids) are not too frequent. Thus, our existence acts as an observational selection effect (cf. Bostrom 2002), or "filter" selecting those sites – in this case planetary systems – where configurations of cometary orbits are in some sense atypical. There is nothing inherently problematic in this methodological approach – or at least nothing *more* problematic than in explaining properties of stellar surveys by the Malmquist bias, or paleontological extinctions by the Signor-Lipps effect, etc. All of these are just manifestations of the observation selection-effects, which need to be studied and corrected for in the context of the anthropic reasoning.

3. Anthropic arguments in risk analysis

Another scientific field which by default deals with variable number of observers is risk analysis – in particular when applied to large catastrophic risks (Bostrom and Ćirković 2008). In the previous section, we have seen that one of the "rare Earth" requirements is that a habitable planet is not subjected to impact and supernova risks more severe than a particular threshold value. Let us now focus on Earth and consider big natural hazards: how do we actually measure the level of such *historical* catastrophic risks as impacts or supervolcanoes?[14] Usual procedure (e.g., Woo 1999) is to fit a distribution to a sufficiently big

[14] The risk from close supernovae/γ-ray bursts is somewhat different case, since these dramatic explosions will tend to leave few traces in the historical record. While in some cases it was considered an *advantage* (e.g., Schindewolf 1962), the anthropic selection effect I wish to discuss here does not apply to such processes (but see Bishop and Egli 2011).

database of the events of well-defined type in terms of both severity and time. From calculated moments of this distribution one can compute further values of interest, for instance the probability of an earthquake of magnitude larger than *S* hitting within next *T* years, or the probability a supereruption emitting more than *V* cubic kilometers of ash and dust over the same interval of time. The database is in most cases built up from *historical evidence* (or even indirectly, from the fossil record). Empirical predictions are, in general, moments of the distribution function of any particular risk.

The usual procedure does not take into account the tally of observers. Therefore, it is subject to a specific bias called the *anthropic shadow* in Ćirković, Sandberg, and Bostrom (2010): a part of the parameter space of large catastrophic events is not adequately sampled, due to its incompatibility (or weaker compatibility, so to speak) with our existence as observers at present. In other words, the usual procedure, which does not distinguish between the directions of time (the distribution function is assumed to be the same, or similar, or obtained by the same procedure in the future as in the past) fails to account for the selection effects. Consequently, it leads not only to *undersampling* of the total empirical dataset, but to *underestimates* of the predicted risks as well. The past record is unrealiable, due to our presence as observers in the present; and it is unreliable in the consistently misleading direction – it makes our environment seem safer than it really is. This is a quintessentially anthropic effect with a wide range of consequences, from risk analysis to epistemology of historical science to ethics and politics of risk.

Anthropic shadow is easiest to understand through a particular example of a specific risk whose magnitude we assess from historical data: terrestrial impact craters. Craters are consequences of impacts of small Solar System bodies (asteroids and comets) upon Earth's surface. Impacts produce physical effects ranging from local and small (most of meteorites, like the famous Murchison meteorite which fell in Australia in 1969), to regional and severely destructive (e.g., the event which created Barringer crater in Arizona about 50,000 years ago), to global and cataclysmic (e.g., the Chicxulub impact, 65 million years ago at the K/Pg boundary). There are 188 confirmed impact craters on Earth at the time of this writing (February 2016).[15] There is, obviously, a relationship between the size of the impactor and the size of the crater created, although the relationship is non-linear and complicated, needing to account for properties such as chemical composition of the impactor, incident angle, etc. (Hughes 2003). But if we do not bother with such complications, we can gauge the severity of an impact catastrophe by the size of an impact crater. So, in the long term, the list of known impact craters will give us the distributionfunctionof impact catastrophes – and by reconstructing it we can predict or retrodict the catastrophe of a given size in a given time interval; for example, we can answer the question how probable is a Chixhulub-size impact within next 1000 years from now. Of course, we need to deal with "classical" selection effects, such as erosion, incompleteness of our crater list, etc.

[15] See The Planetary and Space Science Centre 2011–2016.

However, this neglects another key selection effect: the fact that our existence as observers as present does impose limitations on the possible risk distribution functions. Not all distribution functions *compatible with all the historical evidence* are compatible with our existence as observers (and our number) at present. So, those historical trajectories are falsely taken into account – in essence, *subsumed into* – in constructing the distribution function. We need to unravel this biased sampling in order to get to the *real* (or *a posteriori*) distribution function, giving us the *real* magnitude of risk. Surprisingly enough, (as shown by Ćirković et al. 2010) the correction leads to underestimating the risk or *overconfidence*. In our example of impact risk, it is clear that there is no chance whatsoever to find a 100km-size crater younger than 1 million years, and that a probability of find such huge crater younger than 10 million years is minuscule. This is because such catastrophe would have caused the extinction of humankind, if it happened, or would have cut the evolutionary chain leading to humans if further in the past. On the other hand, the probability of such crater being formed in the *next* million years is certainly not zero, and the probability of its formation in the next 10 million years is sizeable (e.g., Grieve and Shoemaker 1994). But since the moments of the same distribution function must be the same in both past and the future, we can only infer that the real risk is greater than that calculated by the standard procedure.

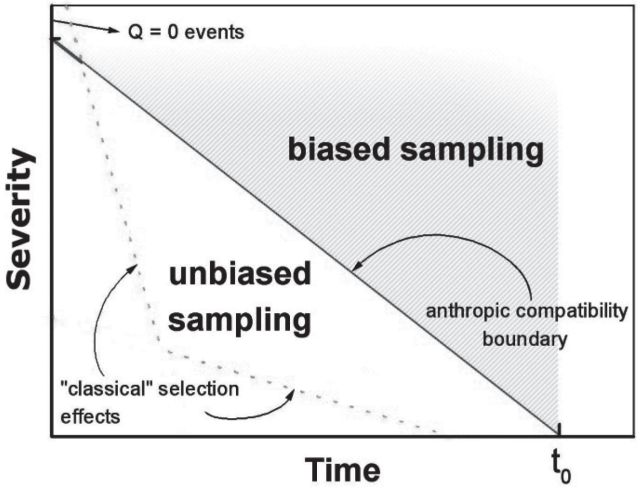
Q = 0 events
biased sampling
unbiased sampling
anthropic compatibility boundary
"classical" selection effects
Severity
Time
t_0

Figure 1: A sketch of the anthropic bias: we do not fairly sample the entire time-severity plane of large risks, only a region compatible with our existence at this particular epoch (the rest is in the "anthropic shadow" – shaded region, see text). The current epoch is denoted by t_0 and we count time from the formation of our planet (from Ćirković et al. 2010). Even if the amount of bias is small – that is,

> if evolution is convergent and robust against perturbations – this is still a legitimate effect which clearly demonstrates the importance of anthropic thinking in the field of risk analysis. Q = 0 events are completely destructive events which leave no survivors.

What happens as we increase – in a thought experiment if not actual historical record – the severity of the catastrophic event whose probability we wish to calculate? Catastrophic events exceeding some threshold severity eliminate all observers and all ecological conditions necessary for subsequent emergence of observers, and are hence unobservable. Some types of catastrophes may also make the existence of observers on a planet impossible in some subsequent interval, the size of which might be correlated with the magnitude of the catastrophe. Because of this bias, the events reflected in our historical record are not sampled from the full events space but rather from the part of the events space that lies beneath the "anthropic compatibility boundary" (illustrated in Figure 1). The part of the parameter space above the boundary lies in what can be called *anthropic shadow*: the observation selection effect implicit in conditioning on our present existence prevents us from sharply discerning magnitudes of extreme risks close (in both temporal and evolutionary terms) to us. This shadow is the source of bias which must be corrected when we seek to infer the objective chance distribution from the observed empirical distribution of events.

So much about anthropic selection effects in our *past*. What about the future? Global catastrophic risk studies uniformly indicate that the biggest threats for humanity's future now belong to anthropogenic risks, such as nuclear winter, anthropogenic global warming or misuse of biotechnology or nanotechnology. Overconfidence following neglect of the anthropic bias applies here as well – if we have a bias in calculating the distribution function of impact catastrophes from the past leading us to underestimating impact risk, this is not dissimilar from reasoning that since there was no nuclear war in the last cca. 60 years, the risk of its occurring is smaller than we would *a priori* expect. The lack of our prior historical experience does not diminish expected risk – it enhances it; while this may sound counterintuitive, it is a direct consequence of the anthropic reasoning.

This introduces overconfidence in our probabilistic estimates of the magnitude of risk above given severity, leading to possibly detrimental neglect of research and mitigation of particular possibly risky processes. Therefore, what is essentially a philosophical question about justification and discrimination between explanatory hypotheses becomes a very serious and very practical issue: *do we underestimate the risks we are unavoidably facing?* If the answer is affirmative, this leads straight to the situation in which methodological biases and neglects lead to increasing likelihood of harm for billions of humans and uncountable number of other living beings. (This can be generalized to other hypothetical intelligent communities in the Galaxy through another anthropic argument, as has been done by Tegmark and Bostrom 2005).

So, in this case the downplaying of anthropic reasoning might be not only bad science and bad philosophy, but might be ultimately *unethical* as well. Since at least some of the processes subject to anthropic shadow could lead to the ultimate harm (Persson and Savulescu 2008), only minor auxiliary assumptions are necessary to conclude that it is our moral duty to increase our understanding and, ultimately, resilience to such processes. If anti-anthropic dogma stands in the way of better understanding, then this shows not only cognitive bankruptcy of such dogma, but the moral bankruptcy as well.

4. Discussion

It is quite easy to realize that the anthropic reasoning is, in fact, applicable to all fields and problems where processes could lead, or indeed have historically led, to a significant change in the number of observers. While such fields and problems are still not numerous, they delineate an area much wider than what is usually circumscribed to cosmology and the "bottom" of fundamental physics. To those mentioned above, one may add many other contribution from various other fields; for instance, Reinganum's (1986) "financial proof" of the impossibility of time travel also belongs to this category, since it depends on observation selection against observers belonging to a particular reference class (those time travelling from the future to the present). The present discussion could be understood as a provocation and call for more similar examples from different scientific disciplines. This could help in defending the whole province of explanatory work from interference by mysticism and religion-inspired abuse.

Moreover, any research involving situations – possible or actual – with changing number of observers, are *necessarily a* subject to the anthropic bias. It is mandatory, not optional, to account and correct for this bias. Conversely, it is *bad science* not to do it, no matter what philosophical and methodological prejudices might be in play. In my view, it is also a case of *bad philosophy* to call for rejection of the question or explanatory nihilism *without previously investigating all explanatory options and finding them unacceptable*. Unfortunately, the nihilistic approach still has many adherents when it comes to explanation of very general phenomena of cosmology and fundamental physics (e.g., Callender 2004; Mosterín 2005); the idea is that the issues such as the initial entropy of the universe or the strength ratios of low-energy effective forces are so general that everything else, including entities used in a potential explanation, is contingent upon those fundamental "almost brute" facts. (The qualification "almost" pertains to the lip service given to possible future advances in physics, which is often not taken seriously at all. Thus Mosterín claims, for instance, that "the most desirable epistemic situation would be that the values of the fundamental constants... could one day be derived from some fundamental physical theory. In the mean time (and this mean time can last *forever*) they *have to be accepted as brute facts*" (Mosterín 2005, 457; emphasis M. M. Ć.). Barring the decision of some Central Committee of Science or a Supreme Philosophical Leader, I completely fail to see any reason whatsoever for such compulsion.) So we cannot formulate a causal explanations for those ultra-deep empirical facts about the physical universe, since we do not know in terms of what to formulate such an explanation. We should accept them as primitive happenstances and proceed further from that.

Clearly, the nihilistic strategy fails in clearly circumscribed fields such as planetary science or risk analysis considered above. It would be unacceptable – not to mention rude – to reject further discussion in these fields even in cases where due to historical loss of information we do not have insight in causal agents acting in distant past. While rudeness has not been, however, below some of the anthropic detractors' discourse – e.g., abovementioned numerical errors in the paper by Klee (2002) criticising anthropic theorists for laxity with numbers – it is obvious that those parts of the battleground are effectively lost. We cannot consciously avoid discussing the impact of habitability on the number of observers or selection effects downsizing huge physical risks we are facing. In the latter case, declining to engage in the discussion has adverse morale consequences as well. So, examples like those discussed in the present paper act to decrease our confidence in explanatory nihilism even in cosmology or string theory, quite independently of actually affirming the value of the anthropic reasoning.

We conclude with a subjective and speculative outline of what could be an interesting project in sociology and psychology of science. It is a telling phenomenon that the existence of wide applications of anthropic reasoning has been a "well-guarded secret" for quite some time – and that the prejudice that only relevant conclusions could be drawn from it in string theory and cosmology is so widespread. I suspect that a major factor in resistance to anthropic arguments and explanations is the fact that they go against received wisdom of two otherwise strongly opposed broad camps, which, for the lack of better words, I will call *naive reductionism* and *naive teleology.* In both cases, the qualification "naive" is central, since anthropic accounts – as discussed above – do not really undermine the role of reductionism and, to a limited extent, teleology in explanations. It is one of those prejudices which is strongly supported by proponents of two otherwise opposing views – and therefore it is much harder to reject. In the course of the Cold War, for instance, both sides have been in agreement, for very different reasons, of course, that USSR and its satellite dictatorships are countries of "real socialism". Many honest and well-intentioned social-democrats or non-Marxist socialists in the West have had a hard time in trying to dispel this notion, since it was promoted by both powerful propaganda apparati. The position of anthropic theorist is similarly uncomfortable, being attacked by both naive reductionists and naive teleologists; the metaphor of Scylla and Charybdis immediately comes to mind. Both naive positions strive to contain anthropic reasoning – if they cannot stamp it out completely – as a contagious heresy, which must not pollute "normal" fields of inquiry. Hence completely nonsensical repeated claims that the anthropic reasoning deals in tautologies, which cannot tell us anything we don't already know; quite to the contrary, those are *quite informative* tautologies, which tell us much about the world.

Naive reductionism, usually sprinkled with remnants of logical positivism, cannot ever come to terms with the idea than extremely complex structures such as intelligent observers may play a role in explanation on any level. Naive reductionists tend to reject or downplay any explanation which is not mechanist and atomic explanation and implicitly hold the dogma that there must be a unique underlying micro-level causal account for any observed macro-level phenomenon. Obviously, the reference to observers and their *historical experience* contained in the very heart of the anthropic reasoning is anathema to naive reductionists. It is very similar to suspicions and misunderstandings characterizing the response of many *secular* scientists and philosophers to neodarwinian evolutionary paradigm; in both cases I suspect deeply rooted suspicion toward using extremely complex and inevitably historical entities in explanatory accounts.[16] (And other modes of explanation, in particular geometric or topological ones, fare no better at hand of naive reductionists; this only confirms that naive reductionism should be fought against and overcome as much as possible.)

Similarly, naive teleologists are dismayed by the *disteleology* of anthropic reasoning. What could be a splendid opportunity for prolonged mystical musings on the "relationship of mind and the universe", "the role of mind [or Mind] in the overall scheme of things", and similar theatrical clichés, is suddenly discounted as the consequence of simple observation selection effects and could be quantitatively modelled in Bayesian terms! How terribly inconvenient. And not only creationists and similar backward scarecrows feel uncomfortable. While naive teleology is, at first glance, mostly absent from serious literature in at least last several decades and therefore seemingly presents no serious threat to science, one should remain highly vigilant, since anti-Copernican, teleological, and anthropocentric thinking is still powerfully present in popular culture (including many instance of popular science and philosophy), as well as in large segments of the media, politics, arts, and humanistics. Especially when coupled with more virulent strains of social constructivism, epistemic relativism and postmodernism, it still has the capacity to do great and hard-to-repair harm. But even in the relatively benign form exhibited, for instance, by Midgley (1985) or the anthropic detractors cited in the introduction such as Earman or Pagels or Mosterin, it can produce confusion and lead to closed-mindedness on many sensitive scientific issues, from animal sentience, to artificial intelligence, to SETI projects.

That the double fallacy of both naive reductionists and naive teleologians has some traction is testified, in a bizarre way, by perhaps the *least* naive of the naive teleologians of the 20th century, Sir Fred Hoyle, who in his popular book *The Intelligent Universe* proclaimed that "[t]he same nihilistic belief that no aspect of the Universe can be thought of as a consequence of purpose underlies both Darwinism and the anthropic principle" (Hoyle 1983, 220). His caricature of Darwinism amounts to what I have above labelled naive reductionism (specifically in biology). It sounds funny and outright bizarre today, but Hoyle

[16] An example of this is Fodor and Piattelli-Palmarini 2010.

seems to had been fully aware that there are altogether *three* positions, rather than two, and that the third one, which is the anthropic reasoning, is bound to be regarded with suspicion and hostility by the adepts of the other two.

The *expanded mandate* of the anthropic reasoning, as dealing with any field with variable number of observers, could easily be branched into other domains. Elaboration of the properties of observers requires multidisciplinary collaboration between mathematics and physics of complexity, evolutionary biology and astrobiology, neurosciences, risk analysis, and philosophy. There is a wide front along which progress could be made by better understanding and building quantitative models of physical, chemical, planetological, etc. preconditions for observership. Both work in artificial intelligence and in zoopsychology and in SETI studies should give us insight into properties of observers different from ourselves. Fields which we have on purpose avoided here, cosmology and fundamental physics, give us tantalizing glimpses of the multiverse, and the interest in physical properties of other universes and their habitability grows almost by day.[17] Prospects for further research are bright indeed. Diametrally opposed to the dark, scholastic, uninformative picture painted by the detractors, the anthropic reasoning is indeed liberating, future-oriented and unifying.

Acknowledgements. I wish to thank Jelena Dimitrijević, Nick Bostrom, Anders Sandberg, Branislav Vukotić, Slobodan Popović, Slobodan Perović, Ivana Kojadinović, Karl Schroeder, Aleksandar Obradović, Jelena Andrejić, Goran Milovanović, George Dvorsky, Zoran Knežević, Jacob Haqq-Misra, the late Robert Bradbury, and the late Branislav Šimpraga for many pleasant and useful discussions on the topics related to the subject matter of this study. This is an opportunity to thank KoBSON Consortium of Serbian libraries, as well as NASA Astrophysics Data System. This research has been supported by the Ministry of Education and Science of the Republic of Serbia through the project ON176021.

References

Adams, Fred C. 2008. "Stars in other Universes: Stellar Structure with Different Fundamental Constants." *Journal of Cosmology and Astroparticle Physics* 08(010): 1–28. doi: 10.1088/1475–7516/2008/08/010

Adams, Fred C., Katherine R. Coppessa, and Anthony M. Bloch. 2015. "Planets in other Universes: Habitability Constraints on Density Fluctuations and Galactic Structure." *Journal of Cosmology and Astroparticle Physics* 09(030): 1–27. doi: 10.1088/1475–7516/2015/09/030

Aguirre, Anthony, Matthew C. Johnson, and Assaf Shomer. 2007. "Towards Observable Signatures of other Bubble Universes." *Physical Review D* 74: 063509–1–17. doi: 10.1103/PhysRevD.76.063509

[17] Among rapidly growing literature on this topic, let us mention Harnik, Kribs, and Perez 2006; Aguirre, Johnson, and Shomer 2007; Adams 2008; Schellekens 2013; Adams, Coppessa, and Bloch 2015.

Barnes, Luke A. 2012. “The Fine-Tuning of the Universe for Intelligent Life.” *Publications of the Astronomical Society of Australia* 29: 529–564.

Barrow, John D., and Frank J. Tipler. 1986. *The Anthropic Cosmological Principle*. New York: Oxford University Press.

Bishop, Shawn, and Ramon Egli. 2011. “Discovery Prospects for a Supernova Signature of Biogenic Origin.” *Icarus* 212: 960–962.

Boltzmann, Ludwig. 1895. “On Certain Questions of the Theory of Gases.” *Nature* 51:413– 414. doi: 10.1038/051413b0

Bostrom, Nick. 2002. *Anthropic Bias: Observation Selection Effectsin Science and Philosophy*. New York: Routledge.

Bostrom, Nick, and Milan M. Ćirković, eds. 2008. *Global Catastrophic Risks*. Oxford: Oxford University Press.

Callender, Craig. 2004. “Measures, Explanation and the Past: Should ‘Special’ Initial Conditions Be Explained?” *British Journal for the Philosophy of Science* 55(2): 195–217. doi: 10.1093/bjps/55.2.195

Carroll, Sean M. 2006. “Is Our Universe Natural?” *Nature* 440: 1132–1136. doi: 10.1038/nature04804

Carter, Brandon. 1974. “Large Number Coincidences and the Anthropic Principle in Cosmology.” In *Confrontation of Cosmological Theories with Observational Data; Proceedings of the Symposium, Krakow, Poland, September 10–12, 1973*, edited by Malcolm S. Longair, 291–298. Dordrecht: D. Reidel Publishing Co.

Carter, Brandon. 1983. “The Anthropic Principle and Its Implications for Biological Evolution.” *Philos. Trans. R. Soc. London A* 310(1512): 347–363. doi: 10.1098/rsta.1983.0096

Carter, Brandon. 1993. “The Anthropic Selection Principle and the Ultra-Darwinian Synthesis.” In *The Anthropic Principle*, *Proceedings of the Second Venice Conference on Cosmology and Philosophy*, edited by by Francesco Bertola and Umberto Curi, 33–66. Cambridge: Cambridge University Press.

Ćirković, Milan M. 2002. “On the First Anthropic Argument in Astrobiology.” *Earth, Moon, and Planets* 91(4): 243–254. doi: 10.1023/A:1026266630823

Ćirković, Milan M. 2003. “The Thermodynamical Arrow of Time: Reinterpreting the Boltzmann-Schuetz Argument.” *Foundations of Physics* 33(3): 467–490. doi: 10.1023/A:1023715732166

Ćirković, Milan M. 2004. “The Anthropic Principle and the Duration of the Cosmological Past.” *Astronomical and Astrophysical Transactions* 23(6): 567–597. doi: 10.1080/10556790412331335327

Ćirković, Milan M. 2006. “Too Early? On the Apparent Conflict of Astrobiology and Cosmology.” *Biology and Philosophy* 21(3): 369–379. doi: 10.1007/s10539–005–8305–2

Ćirković, Milan M., Anders Sandberg, and Nick Bostrom. 2010. “Anthropic Shadow: Observation Selection Effects and Human Extinction Risks.” *Risk Analysis* 30(10): 1495–1506. doi: 10.1111/j.1539–6924.2010.01460.x

Chopra, Aditya, and Charles H. Lineweaver. 2016. "The Case for a Gaian Bottleneck: The Biology of Habitability." *Astrobiology* 16(1): 7–22. doi: 10.1089/ast.2015.1387

Chyba, Christopher F., and Kevin P. Hand. 2005. "Astrobiology: The Study of the Living Universe." *Annu. Rev. Astron. Astrophys* 43: 31–74. doi: 10.1146/annurev.astro.43.051804.102202

Clube, Victor, and Bill Napier. 1990. *The Cosmic Winter*. Oxford: Basil Blackwell Ltd.

Collier, John D. 1990. "Intrinsic Information." In *Information, Language and Cognition: Vancouver Studies in Cognitive Science*, edited by Philip P. Hanson, 390–409. Oxford: Oxford University Press.

Des Marais, David J., Joseph A. Nuth III, Louis J. Allamandola, Alan P. Boss, Jack D. Farmer, Tori M. Hoehler, Bruce M. Jakosky, Victoria S. Meadows, Andrew Pohorille, Bruce Runnegar, and Alfred M. Spormann. 2008. "The NASA Astrobiology Roadmap." *Astrobiology* 8(4): 715–730. doi: 10.1089/ast.2008.0819

Dick, Steven J. 2003. "Cultural Evolution, the Postbiological Universe and SETI." *Int. J. Astrobiology* 2(1): 65–74. doi: 10.1017/S147355040300137X

Dicke, Robert H. 1961. "Dirac's Cosmology and Mach's Principle." *Nature* 192: 440–441. doi: 10.1038/192440a0

Dirac, Paul A. M. 1937. "The Cosmological Constants." *Nature*139: 323–323. doi: 10.1038/139323a0

Dirac, Paul A. M. 1961. "Dirac replies [to Dicke's letter]." *Nature* 192: 441–441. doi: 10.1038/192441a0

Doyle, Laurance R., Joshua A. Carter, Daniel C. Fabrycky, Robert W. Slawson, Steve B. Howell, Joshua N. Winn, Jerome A. Orosz, Andrej Prˇsa, William F. Welsh, Samuel N. Quinn, David Latham, Guillermo Torres, Lars A. Buchhave, Geoffrey W. Marcy, Jonathan J. Fortney, Avi Shporer, Eric B. Ford, Jack J. Lissauer, Darin Ragozzine, Michael Rucker, Natalie Batalha, Jon M. Jenkins, William J. Borucki, David Koch, Christopher K. Middour, Jennifer R. Hall, Sean McCauliff, Michael N. Fanelli, Elisa V. Quintana, Matthew J. Holman, Douglas A. Caldwell, Martin Still, Robert P. Stefanik, Warren R. Brown, Gilbert A. Esquerdo, Sumin Tang, Gabor Furesz, John C. Geary, Perry Berlind, Michael L. Calkins, Donald R. Short, Jason H. Steffen, Dimitar Sasselov, Edward W. Dunham, William D. Cochran, Alan Boss, Michael R. Haas, Derek Buzasi, Debra Fischer. 2011. "Kepler-16: A Transiting Circumbinary Planet." *Science* 333(6049): 1602–1606. doi: 10.1126/science.1210923

Dunbar, David Noel, Ralph E. Pixley, W. A. Wenzel, and Ward Whaling. 1953. "The 7.68-Mev State in C^{12}." *Physical Review* 92(3): 649–650. doi: 10.1103/PhysRev.92.649

Earman, John. 1987. "The SAP also Rises: A Critical Examination of the Anthropic Principle." *American Philosophical Quarterly* 24(4) 307–317.

Ellis, George F. R. 2011. "Editorial Note to: Brandon Carter, Large Number Coincidences and the Anthropic Principle in Cosmology." *General Relativity and Gravitation* 43(11): 3213–3223. doi: 10.1007/s10714–011–1257–8

Fodor, Jerry, and Massimo Piattelli-Palmarini. 2010. *What Darwin Got Wrong*. New York:Farrar, Straus, and Giroux.

Fontenelle, Bernard le Bovier de. 1990 [1686]. *Conversation on the Plurality of the Worlds*-Translated by H. A. Hargreaves. Berkley: University of California Press.

Freer, Martin, and Hans Otto Uldall Fynbo. 2014. "The Hoyle State in ^{12}C." *Progress in Particle and Nuclear Physics* 78: 1–23. doi: 10.1016/j.ppnp.2014.06.001

Grieve, Richard A. F., and Eugene M. Shoemaker. 1994. "The Record of Past Impacts on Earth." In *Hazards Due to Comets and Asteroids*, edited by Tom Gehrels, 417–464. Tucson: University of Arizona Press.

Halley, Edmond. 1704. "Astronomiae Cometicae Synopsis, Autore Edmundo Halleio apud Oxonienses. Geometriae Professore Saviliano, & Reg. Soc. S.." *Philosophical Transactions of the Royal Society ofLondon*24(289–304): 1882–1899. doi: 10.1098/rstl.1704.0064

Halley, Edmond. 1726. "Some Considerations about the Cause of the Universal Deluge." *Philosophical Transactions of the Royal Society ofLondon*33(1724–1725): 118–123. doi: 10.1098/rstl.1724.0023

Hanslmeier, Arnold. 2009. *Habitability and Cosmic Catastrophes*. Berlin: Springer.

Harnik, Roni, Graham D. Kribs, and Gilad Perez. 2006. "A Universe without Weak Interactions." *Physical Review D* 74(3):035006–1–15. doi: 10.1103/PhysRevD.74.035006

Horneck, Gerda, and Petra Rettberg, eds. 2007. *Complete Course in Astrobiology*. Weinheim: Wiley-VCH.

Hoyle, Fred. 1954. "On Nuclear Reactions Occuring in Very Hot Stars. I. the Synthesis of Elements from Carbon to Nickel." *Astrophysical Journal Supplement* 1: 121–146. doi: 10.1086/190005

Hoyle, Fred. 1983. *The Intelligent Universe*. London: Michael Joseph Limited.

Hughes, David W. 2003. "The Approximate Ratios between the Diameters of Terrestrial Impact Craters and the Causative Incident Asteroids." *Monthly Notice of the Royal Astronomical Society* 338(4): 999–1003. doi: 10.1046/j.1365–8711.2003.06157.x

Gould, Stephen J. 1987. *The Flamingo's Smile:Reflections in Natural History*.New York: W. W. Norton & Company.

Klee, Robert. 2002. "The Revenge of Pythagoras: How a Mathematical Sharp Practice Undermines the Contemporary Design Argument in Astrophysical Cosmology." *Brit. J. Phil. Sci.* 53(3): 331–354. doi: 10.1093/bjps/53.3.331

Leslie, John A. 1989. *Universes*. Oxford:Routledge.

Maddox, John. 1984. "New Twist for the Anthropic Principle." *Nature* 307(5950): 409. doi: 10.1038/307409a0

Manson, Neil A. 2000. "There Is No Adequate Definition of 'Fine-tuned for Life.'" *Inquiry* 43(3): 341–351. doi: 10.1080/002017400414890

Midgley, Mary. 1985. *Evolution as a Religion: Strange Hopes and Stranger Fears*. London: Routledge.

Mosterín, Jesús. 2000. "The Anthropic Principle in Cosmology: A Critical Review." *Acta Institutionis Philosophiae et Aestheticae* 18: 111–139.

Mosterín, Jesús. 2005. "Anthropic Explanations in Cosmology." In *Proceedings of the 12th International Congress of Logic, Methodology and Philosophy of Science*, edited by Petr Hajek, Luis Valdés-Villanueva, and Dag Westerståhl, 441–471. Amsterdam: North-Holland Publishing.

Newton, Isaac. 1730. *Opticks*.London: William Innys.

Pagels, Heinz R. 1998. "A Cozy Cosmology." In *Modern Cosmology & Philosophy*, edited by John Leslie, 180–186.Amherst: Prometheus Books.

Persson, Ingmar, and Julian Savulescu. 2008. "The Perils of Cognitive Enhancement and the Urgent Imperative to Enhance the Moral Character of Humanity." *Journal of Applied Philosophy* 25(3): 162–177. doi: 10.1111/j.1468–5930.2008.00410.x

Reinganum, Marc R. 1986. "Is Time Travel Impossible? A Financial Proof." *The Journal of Portfolio Management* 13(1): 10–12. doi: 10.3905/jpm.1986.10

Schellekens, A. N. 2013. "Life at the Interface of Particle Physics and String Theory." *Rev. Mod. Phys.* 85: 1491–1540. doi: 10.1103/RevModPhys.85.1491

Schindewolf, Otto. 1962. "Neokatastrophismus?" *Deutsch Geologische Gesellschaft Zeitschrift Jahrgang* 114: 430–445.

Smolin, Lee. 2007. "Scientific Alternatives to the Anthropic Principle." In *Universe or Multiverse?*, edited by Bernard Carr, 323–366. Cambridge: Cambridge University Press.

Steckline, Vincent S. 1983. "Zermelo, Boltzmann, and the Recurrence Paradox." *Am. J. Phys.* 51(10): 894–897. doi: 10.1119/1.13373

Susskind, Leonard. 2006. *The Cosmic Landscape: String Theory and the Illusion of Intelligent Design*. New York: Back Bay Books.

Tegmark, Max. 1996. "Does the Universe in Fact Contain almost no Information?" *Found. Phys. Lett.* 9: 25–42. doi: 10.1007/BF02186207

Tegmark, Max. 2008. "The Mathematical Universe." *Foundations of Physics* 38: 101–150. doi: 10.1007/s10701-007–9186–9

Tegmark, Max, and Nick Bostrom. 2005. "Astrophysics: Is a Doomsday Catastrophe Likely?" *Nature* 438(7069): 754. doi: 10.1038/438754b

Walker, Mark A., and Milan M. Ćirković. 2006. "Anthropic Reasoning, Naturalism and the Contemporary Design Argument." *International Studies in the Philosophy of Science* 20(3): 285–307. doi: 10.1080/02698590600960945

Wiener, Norbert. 1961. *Cybernetics*.New York: John Wiley and Sons.

Wilson, Patrick A. 1994. "Carter on Anthropic Principle Predictions." *Brit. J. Phil. Sci.* 45(1): 241–253. doi: 10.1093/bjps/45.1.241

Woo, Gordon. 1999. *The Mathematics of Natural Catastrophes*. London: Imperial College Press.

Online sources

Motl, Luboš. 2004–2017. “The Reference Frame.” *Luboš Motl’s Blog*. http://motls.blogspot.rs/

The Planetary and Space Science Centre. 2011–2016. “Earth Impact Database.” *The Planetary and Space Science Centre*. http://www.passc.net/EarthImpactDatabase/index.html

Woit, Peter. 2004–2016. “Not Even Wrong.” *Peter Woit's Blog*. http://www.math.columbia.edu/~woit/wordpress

ARTICLES

Vladimír Marko
FiF, UK

Original Scientific Paper
UDK 167/168:123.2

DETERMINISMS*

Abstract: *Determinism is usually understood as a commonly clear and obvious thesis. In the most of the actual literature a character of determinism is rarely enough explicitly underlined and we believe that it is the reason why common uses of the term often leads to inconsistencies and present a source of misunderstandings of different sorts. Here we will try to show that that there are many forms of determinism; that the concept of determinism has a composite character; and that conceptions of determinism can be mutually discriminated and organized according to particular elements they are consisting of by applying the procedure of classification.*

Keywords: *classification, definition, determinism, logical determinism, scientific determinism.*

Introduction

Our impression is that determinism is usually understood as a commonly clear and obvious thesis. In the most of the actual literature a character of determinism is not enough explicitly underlined and we believe that this is the reason why a common use of the term often leads to inconsistencies and misunderstandings of different sorts.

In the article we put aside questions concerning soundness or deficiencies in philosophical conceptions we are dealing with here. We are interested here in some basic conceptual shapes that are present in different kinds of determinism and in the question: is it possible to sort them in some interrelated order for purposes of their better differentiation? Our main thesis here is essentially threefold: that there are *many forms* of determinism; that the concept of determinism has *composite character*; and that conceptions of determinism can be *mutually discriminated and organized* according to particular elements they are consisting of.

[*] This article is result of work supported by Slovak Research and Development Agency grant No. APVV-0149–12

The basic step in our attempt and probably the most important one, is to show a presence of different ways of understanding the philosophical concept of determinism. For this reason, we will take as our primarily concern to present some collection of approaches as an evidence that could be enough persuasive for our claim. Lists that will be presented here have no ambition to be exhaustive compilation. We see our mission more in drawing attention to the problem and to sketch some of the ways out than to give full account of it. Extensive list able to compile the most variations of determinism would be the task that requires more space and different approach (that can rely on far better developed criteria for comparative or historical analysis of particular cases). Since there are no currently well formed and precise criteria for such mission our exposition has to be understood as a preliminary attempt that probably could lead toward such more elaborated work.

Certainly, it is possible to imagine more such lists that could include a different choice of persons and their conceptions. However, we hope that, even in this form, it will highlight our basic observations and be of help in transparency of our intention. Opinions reflected here are taken from different sides either in respect to the philosophical orientations of authors or in respect to the problems that are processed by some theories. We believe that those amended in the list will be sufficiently transparent to show the idea, that there are many forms of determinism – actually, that many among them present conceptually quite different claims.

An important question is how we are able to highlight differences among forms of determinism? We believe that to point up the differences it will be enough to identify these forms in regards to components they are consisting of. So we decided to attach some labels to components of these forms and to appropriately emphasize them. Even more such components can be found, we will here devote place only to some of them that are seen as crucial. We will take into account those that are dominant in respect to assessment of their role in particular composite forms of determinism or at least with respect to frequency of their common use. In some cases and to some extent these forms are evidently mutually related while in other cases their properties not quite obviously differ. However, we believe that it could be possible to expose them for the purpose of their appropriate understanding.

There is another question we are interested in: If there are more deterministic conceptions such that they are related according to some of their components is it possible to systematically classify them? By accentuate differences among them and by labeling them according to their 'most characteristic proprieties' we could finally obtain some typology of forms of determinism. However, we believe that it is possible to go further in this ambition of sorting. Typology itself will give us divided forms of determinism but sorting according to types will be result that does not give us relevant information about familiar relations among different sorted forms, how they are interconnected and to what extant. If there are some common particles in different forms of determinism then they could be organized by help of classification. This kind of sorting can be more informative in respect to the nature of any member included in it. For this reason we prefer this last option as more fruitful than typology approach since, by this way, mutual dependence of different forms of determinism will stay distinguished, evident and more useful for detecting a character of particular philosophical theories and their conceptual basis.

1.

The most modern authors, when they want to label certain philosophical conception as deterministic, do not always feel need to additionally explicate what they mean by determinism. They simply take it as something clear and granted. However, meaning of the term varies – not only during the past but even nowadays. Besides, its actual use seems to tacitly assume different theoretical backgrounds, regardless of whether we have in mind philosophical conceptions or standard scientific practice.

The term has the Latin origin. In Roman authors we can find *determino* or *determinatio* with a following meaning: *to enclose within boundaries*, *to bound*; *to limit*, *to prescribe*, *to determine*; *to fix*, *to settle*. Livy uses it as a technical term in describing augur's procedure of dividing parts of heaven into regions (*determinavit regiones*) and for marking their boundaries [*Ab urbe condita libri*, i, 18, 7, 32]. Almost the similar example is in Gellius (Gellius 1927, 13, 14). In Cicero, "the conclusion [i.e. *peroration*] is the end and terminating of the whole speech (*determinatio totius orationis*)" (Cicero 1949, 1, 52, 98). A Greek equivalent of Latin *definire*, *determinare* would be ἀφωρισμένης. It has been used in approximately the same manner as in later Latin authors.

However, this terms during the ancient and medieval times nowhere has it usual nowadays sense. We does not encounter these terms in the contemporary philosophical debates related to the themes on determinism. Beside the fact that determinism, as philosophical conception, never lacked its advocates during history, especially the ancient history, sole term determinism comes to us from some later times.

2.

OED (*Oxford English Dictionary*) gives us the following formulation that has to cover sense of the term:

> "The philosophical doctrine that human action is not free but necessarily determined by motives, which are regarded as external forces acting upon the will."

This formulation is evidently complex and we will engage here in its details but it enough illustrate that determinism is here taken as a subject interlaced with agency. This formulation in some sense corresponds to words of W. Thomson (Lord Kelvin) from his *Oxford Essays*:

> “The theory of Determinism, in which the will is regarded as determined or swayed to a particular course by external inducements and formed habits, so that the consciousness of freedom rests chiefly upon an oblivion of the antecedents to our choice” (Thomson 1855, 181).

OED situates first occurrence of the term in year 1846 when editor of Thomas Reid’s collection, Sir William Hamilton, wrote in a brief footnote:

> “There are two schemes of Necessity—the Necessitation by efficient—the Necessitation by final causes. The former is brute or blind Fate; the latter rational Determinism” (Hamilton 1846, 87n.†).

Hamilton here joins ‘rational determinism’ with ‘final causes’ while ‘necessitation by efficient’ is characterized as brute fate (let we say, fatalism).[1]

The history of the term is a little bit older. The term actually appears a few decades earlier then OED notes, following its English sources.

By leafing Krug’s *Allgemeines Handworterbuch* we can find terms *Determinismus (Bestimmung*, *Predeterminismus*) and *die Deterministen* (Krug 1827, i, 500–501). Also, here is a note on the first appearance of these terms (Krug 1829, 100): Christian Wilhelm Snell used them in commenting Kant’s moral themes, in his brochure *Über Determinismus und moralische Freiheit* (Snell 1789). At several other places in *Allgemeines Handworterbuch*, determinism is used with a sense of ‘a philosophical necessity’. These lines refer to English sources, related to Joseph Priestley’s concept of ‘determination’ (a correspondence with John Palmer is quoted as a source; cf. Priestley 1779; 1780). A year after Snell (in 1790), Carl Friedrich Bahrdt also reflects determinism as theoretical concept (Bahrdt 1790, 291). Soon after, the term appears in Kant’s treatises on religion (Kant 1793). In a footnote, Kant considers determinism in a context of agency and person’s determination by external forces and read it as *predeterminism*, at the same time rejecting it as an ‘illusion’.[2]

Herbart uses the word once for the first time at the end of his and several times later (Herbart 1842). He claims that determinism is prerequisite for action – ‘Determinismus ist Voraussetzung des Handelns’ (Herbart 1843, 147, 152). Hegel [1816: ii, 206; 236] already uses the term as standard philosophical notion (in the context of mechanical processes and also religion and freedom) (Hegel 1816, ii, 206, 236). Extensive list of using the term in German can be found (with

[1] Mill, for example (during approximately the same period, even nowhere directly mentions determinism) claims something different from Hamilton’s option when notes that “[i]f the whole prior state of the universe could occur again, it would again be followed by the present state” (Mill 1843, Bk. III, Chap. VII, §1).

[2] “Die, welche diese unerforschliche Eigenschaft als ganz begreiflich vorspiegeln, machen durch das Wort Determinismus (dem Satze der Bestimmung der Willkür durch innere hinreichende Gründe) ein Blendwerk, gleich als ob die Schwierigkeit darin bestände, diesen mit der Freiheit zu vereinigen, woran doch niemand denkt; sondern: wie der Prädeterminismus, nach welchem willkürliche Handlungen als Begebenheiten ihre bestimmenden Gründe in der vorhergehenden Zeit haben (die mit dem, was sie in sich hält, nicht mehr in unserer Gewalt ist), mit der Freiheit, nach welcher die Handlung sowohl als ihr Gegenteil in dem Augenblicke des Geschehens in der Gewalt des Subjekts sein muß, zusammen bestehen könne: das ist’s, was man einsehen will und nie einsehen wird” (Kant 1793, 58A).

minor shortcomings) in *Deutsches Fremdwörterbuch* (Schulz, Brasler, Strauss 1999, 442–3).

On the basis of 19th century OED formulations as well as on the basis of earlier German texts, it seems that we need to make distinction between determinism as *a term* and determinism as *a philosophical conception*. As we saw, not so long ago, the term determinism refers to the conception in a good part far from its modern sense. By following only historical appearances of the term we are not always on the path that could signify some unique philosophical conception or at least such corresponding to the modern sense of determinism. Besides, even in that age the term determinism is not always followed by representation of unique philosophical conception in its background.

3.

Let we continue here with recalling some historical background of this notion and its use. Cassirer seems to be the first who points to the term by reflecting discrepancies in its conceptual background. The term is, up to the second part of 19th century, regularly used in the context of free will and its determination by antecedent circumstances, usually seen as 'external causes' that determine agent decisions or as 'causa finalis'. Cassirer is aware of it and (in the opening pages of his *Determinism and Indeterminism in Modern Physics*, some kind of a chronicle of the epoch) comes to the conclusion that *the genuine meaning* of the term has to be searched on the other side (Cassirer 1956). He dates rebirth of determinism to 1872, to a year of public speech of Emil du Bois-Reymond on the limits of knowledge of nature (du Bois-Reymond 1886, 107). Why this lecture seems to be important? According to Cassirer, in 19th century accounts on determinism there is a gap in continuity of essential meaning of this notion. Du Bois-Reymond is a person who reflects Laplacean roots of the notion and who tries to restore a genuine philosophical conception of determinism, in a sense of completes causal physical determinism. In fact, du Bois-Reymond is simply refraining words from the key passage of *Essai philosophique sur les probabilités*. Let we only reminds here that Laplace's determinism, based on the principle of universal causal concatenation, was inspired by Leibniz principle of sufficient reason. This is the famous place from Laplace's book, where he writes:

> "We may regard the present state of the universe as the effect of its past and the cause of its future. An intellect which at a certain moment would know all forces that set nature in motion, and all positions of all items of which nature is composed, if this intellect were also vast enough to submit these data to analysis, it would embrace in a single formula the movements of the greatest bodies of the universe and those of the tiniest atom; for such an intellect nothing would be uncertain and the future just like the past would be present before its eyes" (Laplace 1902, 4).

Let us try for a moment to resume in brief these words and to see what could be essential to Laplacean determinism. This form of determinism relies on identification of *causation* and *lawfulness* with *determinism*. Laplace wishes to say that *predictability* (*p*) (or better to say, *calculability*) has to be, at least in principle, grounded on the following postulation. It assumes existence of some form of *intellect* (*i*) equipped with such capacities that enable obtaining *all relevant data* for analysis (*d*), where data are consisting of information about *all forces* (*l*) and *all states* (position of all items at time *t*) of *the system* (*s*) and such that an intellect could be able to cover all these data by *a single formula* (*f*). In short, predictability (*p*) could be read as *result* of *ability* of applying a unique function (calculability) over the all relevant data. He mentions following conditions for predictability: <*i, d, f*>, where data (*d*) here consist of subset <*l, s, t*>. It is clear that Laplace determinism is the philosophical conception compounded from more different elements. Here we have a system; system is governed by causation; causations proceeds according to laws; we have several exceptional abilities (to obtain relevant data, to analyze them, to calculate them by a function, to predict the future), data are consisted of laws (!); laws are understood as active forces with abilities to cause occurrences; etc.

Determinism, as *the philosophical conception* Cassirer has in mind, never completely ceased to exist even some other interpretations of *the term* starts to be more dominant and more influenced. *Cassirer's thesis*, concerning du Bois-Reymond (about exact date of breaking point and of returning to the roots of some genuine determinism) is not quite reliable. The fact is that Ernst von Brücke and Emil du Bois-Reymond are advocating this conception in 1842, almost thirty years earlier than Cassirer situates a breaking point for determinism. Soon after, in 1847, they will be joined by Hermann von Helmholtz and Carl Ludwig. As Hacking shows in his article on 19th century standpoints on determinism,[3] these four has immense impact on the later authors, either for or against the thesis. Anecdote with Cassirer's assertion testifies enough that the term determinism concealed many different philosophical positions, whether those 'genuine' or of another kind. In any case, in the background of 19th century use of the notion there were different philosophical conceptions.

Cassirer prefers one among options and willing to find difference between new, 'critical', and old, 'metaphysical' determinism. The first is based on belief that causal relations and laws are mental in their origin – their source is in our experience. Natural laws are not the domain of objective things, as 'metaphysical' determinism believes, but about cognitions and their ordering. In that sense, causal relation is necessarily on epistemological platform (cf. Cassirer 1956, 114).

Toward the end of 19th century difference in approaches to determinism can be observed in another author. When he tries to demarcate some of deterministic standpoints, William James, motivated to find place for our free will, gives the next descriptions. There is *the old determinism*, claiming that:

> "...parts of the universe already laid down absolutely appoint and decree what the other parts shall be... Any other future complement than the one fixed from eternity is impossible. The whole is in each and every part, and welds it with the rest into an absolute unity, an iron block, in which there can be no equivocation or shadow of turning" (James 1907,150).

[3] For the story about rising of 19th century determinism cf. Hacking 1983.

This ‘old determinism’ he labels as ‘hard determinism’. ‘Hard determinism’ is one that doesn’t shrink from such words as fatality, bondage of the will, necessitation, and the like. However, from the other side, there is an alternative to this old determinism, it is *determinism of nowadays*, i.e. ‘soft determinism’:

> “Nowadays, we have a soft determinism which abhors harsh words, and, repudiating fatality, necessity, and even predetermination, says that its real name is freedom; for freedom is only necessity understood, and bondage to the highest is identical with true freedom” (Ibid.,149).

4.

When contemplates William James’ “The Dilemma of Determinism”, S. Langer conjoins fatalism and determinism, what is not unusual in philosophical practice (Langer 1936, 474). She relates forms of scientific determinism and fatalism and claims that there is a strong connection between these two. Fatalism is usually seen as an outcome of some kind of the full-fledged determinism. Determinism is a useful scientific conception based on assumption that every event has immediate cause, what is a tenable thesis for scientific purposes. Problem arises when this thesis is connected with a thesis of predictability. This very thesis, derived from above place from Laplace, is seen as a form of *the scientific fatalism*. The modern scientific fatalism is “the assumption that there is a theoretically knowable collection of causes for any act”. The thesis is derivable, according to her, from determinism that includes false assumption (given through illustration of Laplace’s demon) about ability to obtain knowledge about ‘total state of the universe’. The last assumption is credited by Russell and Whitehead (1910, 40) as ‘illegitimate totality’ since “a whole cannot be theoretically constructed” and for these reason, such doctrine of determinism, in its philosophic form, is “a modern version of belief in Fate” (Langer 1936, 478). So, legitimate scientific conception of determinism, the scientific determinism, which forms the ground of everyday scientific practice, by adding the metaphysical thesis of unrestricted predictability, leads to *the scientific fatalism*. Langer wishes to make demarcation between predictability and existence of immediate causes since existence of immediate causes does not directly imply predictability. ‘The scientific fatalism’ is assumption that there is a theoretically knowable collection of causes for any act (Ibid.). However, even ‘pure’ determinism and fatalism commonly claim causal connection of the past and future, so that the future can be predicted from the past, they do not entail predictability of the future, for causality does not necessary implies predictability. Besides, even in the case of completely causal universe, unpredictability of human agency brakes down this contention (Ibid., 472), since human agency is not subject of knowability.

Bunge respect Langer’s considerations. He also sees the idea ‘that causality is fatalistic’ as a wrong take while *the scientific determinism* presents as something different from *the fatalistic determinism* (even ‘incompatible’ with it). However, his view of fatalism, causality and determinism differs slightly from that of Langer. While causal determinism is rational theory “offering the means for knowing, predicting, and consequently changing the course of events", fatalism is based of assumption that there is some lawless supernatural Fate that drives unknowable and inescapable Destiny. According to him, there is no fatalism without *fatum* – the power that is above the law, one that installs unconditional necessity and directs the course of events. Causality, on the other side, need not to assume any such transcendental or supernatural agency. Causality does not entail inevitability: some causes can, for example, interfere with one another; the background or hidden causes and conditions may be present; the human conscious may intervene; and so on. Bunge inclines to a conception known as agent-causation while presence of some of elements listed can result in different outcomes (what he interprets as a source of probability). So, 'general determinism' has not to be seen as something that pays unconditionally. It is enabling us to use the knowledge of laws with a purpose to change or modify the course of events and it also leaves a room for chance and freedom. Besides, Bunge firmly believes that statistical laws completely excludes determinism and they are incompatible with it, since they are based not on causal principles but on probability and generalized correlations obtained from data. He believes that "statistical law and probability destroys determinism" (Bunge 1959, 101–102).

5.

We see that *calculability*, *predictability* and *determinism* are usually covered by conception that circulates, during 20th century, under the name of *the scientific determinism*. K. Popper, who himself prefers to interpret determinism as an epistemological thesis, in his *Open Universe* summing up the doctrine of scientific determinism ("the doctrine much stronger than common sense") and considers it as a claim which "most physicians would have agreed at least prior to 1927" (Popper 1982, xx).[4] This doctrine states that:

> "the structure of the world is such that *any event can be rationally predicted, with any desired degree of precision, if we are given a sufficiently precise description of past events, together with all the laws of nature*" (Ibid., 1–2).

According to Popper, the idea of scientific determinism has its roots in 'religious determinism' and seems to be "a kind of translation of religious determinism into naturalistic and rationalistic terms" (Ibid., 6). On the other side, he placed *the metaphysical doctrine of determinism*. This one simply asserts that:

> "all events in this world are fixed, or unalterable, or predetermined. It does not assert that they are known to anybody, or predictable by scientific means. But it asserts that the future is as little changeable as is the past. Everybody knows what we mean when we say that the past cannot be changed. It is in precisely the same sense that the future cannot be changed, according to metaphysical determinism" (Ibid., 7).

[4] Here he has in mind the date of *Fifth Solvay International Conference on Electrons and Photons*.

'Metaphysical' determinism differs from 'scientific' determinism. It is entailed by both religious and 'scientific' determinism. However, metaphysical determinism (as well as 'metaphysical' indeterminism) is *not testable* since its lack of empirical content. In respect to testability, another difference that Popper makes is between a *weak* version of 'scientific' determinism and its *stronger* form (Ibid., 36*ff*).

'The weak' version supposes predictability of the state of any future instant of time of any close physical system ("even from within") "by deducing the prediction from theories in conjunction with initial conditions" (i.e. with conceivable initial conditions). Theories here play role of instruments of describing the world, asserting that the world has certain properties. However, this does not mean that, if the theory that describes certain properties of the world is true, that at the same time all what could be deduced from the theory has to have corresponding property of the world. This last would be position of 'the stronger' kind of determinism, marked by Popper as false, that will suppose predictability of "any given state, *whether or not the system in question will ever be in this state*."

It has to add, that in the question of relation between causality and determinism Popper is not always consistent. In part of his book, it seems to identify causation and determinism (Ibid., 149), while at some other parts asserts them as different (Ibid., 4, 19, 23).

Even predictability is a form of testability of scientific theories Popper criticizes metaphysical form of determinism and the stronger form of determinism. However, there are more critics of formulation of determinism in a form of predictability.[5] Predictability, just one of proprieties of (Laplacean) determinism, is an epistemological concept while determinism should be analyzed as an *ontic* or *physical thesis* and for this reason it is necessary to distinguish determinism in *a proper sense* from determinism related to ability of making predictions. Suppes brings to mind examples of three body problem and Turing machine: both examples are *par excellence* illustrations of deterministic systems. It is known that there is no algorithm (that would support ability of prediction) in determining whether an arbitrary Turing machine in an arbitrary configuration will ever halt (Suppes 1993, 245–246). So he insists on *conceptual separation of two notions*: predictability and determinism. We have good reasons to interpret some systems as 'deterministic' even we are not always able to test it by means of predictability and by recalling mental aspects of the predictability thesis as component of determinism.

6.

By beginning of 20th century the debates on determinism from the second half of 19th century continue. Russell joins the discussion on elucidating the notion of determinism in his well-known lecture on the notion of cause. He tries

[5] For example, Earman 1986, 9–10; Suppes 1993; Kellert 1993; Stone 1989.

to make transparent interconnection among several traditional philosophical notions. A source of philosophical misapprehensions is obscurity of these notions. The notion of determinism has to be demystified by showing its real nature – it has to be considered rather as *a functional relation*:

> "A system is said to be 'deterministic' when, giving certain data, $e_1, e_2, \ldots, e_n$ at times t_1 $t_2,\ldots, t_n$ respectively [*s.c.* 'determinants'], concerning this system, if E_t, is the state of the system at any time *t*, there is a functional relation of the form $E_1 = f(e_1, t_1, e_2, t_2, \ldots, e_n, t_n)$.
>
> The system will be 'deterministic throughout the given period' if *t*, in the above formula, may be any time within that period, though outside that period the formula may be no longer true. If the universe, as a whole, is such a system, determinism is true of the universe; if not, not" (Russell 1917, 199).
>
> "Determinism in regard to the will ... Whether this doctrine is true or false, is a mere question of fact; no *a priori* considerations (...) can exist on either side" (Ibid., 205). "We were unable to find any *a priori* category involved: the existence of scientific laws appeared as a purely empirical fact, not necessarily universal, except in a trivial and scientifically useless form" (Ibid., 208).

For these reasons Russell insists on revision of notions of cause and necessity – two fundamental tools and backbones of the former science – since "there is no *a priori* category of causality" (but merely certain observed uniformities, (Ibid., 205)), the notion of necessity is "a confused notion not legitimately deducible from determinism" (Ibid., 207) and it has to be perceived simply as logical necessity driven by constitutive *determinants* as arguments of a necessary propositional function.

As we can see, Russell's definition is not only about determinism but it is in some sense about pairing determinism with ability of making predictions. Russell, though leaves the notions of cause and causality, his formulation leaves room for conjoining determinism and predictability: system is deterministic exactly if its previous states *determine* its later states in the exact sense in which the arguments of a function *determine* its values.

There is an important suggestion of Russell (which concerns validity of formula outside the period covered by formula) that reflects 'the principle of the irrelevance of time'. Our laws are not *a priori* principles that are applicable to the future besides the fact that they hitherto hold for the past facts. Our formulas are only 'methodological precepts' not 'real laws of Nature' that stands absolutely in respect to the time. It has to add, that Russell is not completely satisfied with his formulation of determinism. There are several reasons. Any set of data points, that are describable by some function, are in fact describable in other ways by infinitely many functions. Further, in the dynamic systems, the past state of a system to which our formula was hitherto applicable could be different in the future and our simplest way to cover facts would no more be the same as it was. Also, the way our system was described hitherto could be transformed to some advance form that will no more necessarily involve the same formula. For this reason we have to bear in mind 'the principle of the irrelevance of time', "that the time must no enter explicitly into our formulae".[6]

Russell's attempt to revise the meaning of determinism seriously shaken the traditional image of science in scientific community. Traditional representation of determinism, by unlinking causes and natural laws from it, now results in the logical form of deterministic necessity.

7.

Even this title is not yet mentioned, Russell observations would be first marks of a rising conception later named by Schlick as *the logical determinism*. M. Schlick repeats Russell's position by following words:

> "Let us see how the scientist uses the word determination—then we shall find out what he means by it. When he says that the state E at the time t_1 is determined by the state C at the time t_0, he means that his differential equations (his Laws) enable him to calculate E, if C and the boundary conditions are known to him. Determination therefore means Possibility of Calculation, *and nothing else*" (Schlick 1932, 114).

His understanding of the natural laws and necessity corresponds to that of Russell. 'The natural law' of science, however, "is not a prescription as to how something should behave, but a formula, a description of how something does in fact behave" (Schlick 1939, 147). The natural laws are just descriptions and they have no force that would push things to move according to their prescriptions. The laws of planetary motion, for example, do not force the planets to move as they do, but only describe their actual motion.

Necessity of logical determinism is not necessity of *the causal nomological determinism*. It is necessity of *functional determination* that enables us, at the basis of determinants and covering function, to calculate (or. better to say, *to infer*) necessary relational dependences among determinants in respect to the function.

Both Russell and Schlick formulations shares one of crucial assumption, that *determinism* is firmly linked with *predictability* (and converse, the ability to make *retrodictions*). Schlick's 'possibility of calculation' corresponds with Laplace's condition for making predictions (though Laplace had in mind a singular function that could be able to cover complete universe). Here, in some sense, there is some sort of comparability among 'calculability', 'predictability' and 'to be determined', between epistemic aspects concerning sequence of state

[6] "In fact we might interpret the 'uniformity of nature' as meaning just this, that no scientific law involves the time as an argument, unless, of course, it is given in an integrated form, in which case *lapse* of time, though not absolute time, may appear in our formulas" (Russell 1917, 205). Extension of Russell's formula in respect to determinism in dynamical or evolutive systems is given in van Fraassen (1989, 254). Russell's function has to be extended to cover all possible trajectories of the system, *i.e.* to encompass changes in successive states of the system.

of affairs (or knowledge about it) and the relational connection with another sequence in different moment that is related to some previous sequence. If one state of affairs is determined, in above functional sense, there is a place for this state to be predicted or to be calculated in advance in respect to the knowledge about its previous states and function that covers all its consequent states.

Schlick's calculability (predictability) is a form of deducibility. It represents a standard understanding about what the logical determinism is – one state is propositionally connected with another state by inferential power. However, logical necessity has to be distinguished from physical necessity and causation: "what is called causal necessity is absolutely different from logical necessity" and "former philosophers so frequently made the mistake of confusing the two and believing that the effect could be logically inferred from the cause" (Schlick 1932, 108). Relationship between logical principles and reality Schlick titles as 'a problem of logical determinism'.[7] He located it in Aristotle's believing:

> "that the Principle of the Excluded Middle could not be applied to future events unless we assume the truth of Determinism."

Having probably Łukasiewicz in mind, Schlick adds that "there are even modern logicians who follow him in this" (Ibid., 115).

8.

Łukasiewicz's formulation of determinism (given more than one decade earlier than that of Schlick), is the following: "By determinism I understand the belief that if *A* is *b* at instant *t* it is true at any instant earlier than *t* that *A* is *b* at instant *t*" (Łukasiewicz 1990, 113). The outcome of this formulation, according to him, would be that the future has to be treated at the same way as the past and that it differs from the past 'only in so far as it has not yet come to pass'. Everything is fixed in advance. The way out from determinism consists in taking seriously suggestion to abandon this believe that leads to conception of eternal truth and to absence of free will.

Łukasiewicz offers two arguments against determinism. One is based on 'the logical principles' while another is based on 'the principle of causality' (and he uses it also in his interpretation of stoical conception of determinism). We will not discuss here in details his attempt to prove determinism on the grounds of propositional calculus as bivalent logical system. We wish only to emphasize his position that *bivalent nature of propositional calculus leads to determinism*. As it is known, the proof relies on identifying two principles: the principle of bivalence and law of excluded middle. Even this proof appears to be valid logically, on the basis of propositional calculus, it has to be abandon for other reasons. The critical moment of argumentation against determinism he summarizes in the following comment: "Although this solution appears to be logically valid, I do

[7] On the other place he formulates it as *the paradox*: "Aristoteles zum Opfer gefallen ist und das noch gegenwärtig Verwirrung stiftet. Es ist das Paradoxon des sog. 'logischen Determinismus'" (Schlick 1931, 159).

not regard it as entirely satisfactory, for it does not satisfy all my intuitions" (Ibid., 124). Attitude against determinism "finds its justification both in life and in colloquial speech" [1990:125]. The principle of bivalence is not applicable to future oriented propositions, to not yet existing things that we use to designate as future and possible. Such propositions have no equal 'real correlates' as those propositions oriented to presence and past. The third, 'neutral', value would be more appropriate to future contingents and they "ontologically have possibility as their correlate".

To sum it up, on the logical basis determinism is consistent conception with its logically valid consequences but chosen logical base is unacceptable in respect to our common intuitions. In such its form, with bivalent assumption embodied, it is not only inappropriate for dealing with future contingents but also has unintuitive consequences in respect to human agency.

Waismann prefers for logical determinism more expressive term, *the logical Predestination,* since, according to this conception, it seems "that indeed the entire future is somehow fixed, logically preordained" (Waismann 1959, 352). Jordan (Jordan 1963, 18), following Waismann, interprets Łukasiewicz's formulation of logical determinism as *the semantic formulation of strict determinism* ("where the strict causal determinism implicitly assumes that an unending sequence of events has no limit") (Ibid., 23). Principle of causality is not necessary outcome of the principle of bivalence but it gives a firm link to real correlates that secure necessary truth of future propositions and at the same time, justify the thesis of eternal truth. In that sense, 'the strict determinism' is the outcome of (a) *the principle of bivalence*, in combination with two additional assumptions: (b) *the correspondence theory of truth* and (c) *the timelessness or absolute character of truth* (Ibid., 1). According to Jordan's representation, 'the strict determinism' occupies the following relative place in the transitive chain of principal dependence: "If the principle of bivalence entails strict determinism and strict determinism entails fatalism, the principle of bivalence entails fatalism" (Ibid., 3). In the same spirit, Wołenski (1996) recently interprets the logical determinism as *the radical determinism.*

9.

Above transitive order suggested by Jordan, during the time, proceeds toward representation of the logical determinism under the standard name of *logical fatalism*. Discussions on Aristotle's the sea-battle example and the future contingent propositions support anchoring of this tradition. Ryle's lecture "It Was to Be" (1953) or Taylor's articles and wide discussion that follows it during the sixties,[8] Ayer's (1963) and Dummett's (1964) texts of fatalism, are among many that certainly contribute to this custom. *Logical necessity* starts to be more frequently interpreted as one that leads to *inevitability*. Even alert to confusion between the logical determinism and fatalism is given yet in the late fifties by

[8] Articles from this polemics on Taylor's article are now collected in the book devoted to D. F. Wallace (cf. Wallace 2011).

Bradley, the tradition of interpreting logical determinism as fatalism (or at least a kind of fatalism) continues and still is present in many modern approaches, especially in those dealing with question of the logical status of future contingents.

Bradley (1959) in his article restates some of Schlick earlier warnings that logical necessity need to be discriminated from causal necessity and also, that truth of logical propositions and their relations have different character from truth of empirical evidence. He criticizes usual assumption that logical determinism *implies* (logical) fatalism. It is not true since *what is timeless* and *what is empirical* are different claims. The failure in this inference consists in *ascribing logical necessity* to *causal necessity* and *causal necessity* to *fatalism*. We can accept as valid that if '*x is causally determined*' it implies '*x is logically determinate*'. However, '*x is logically determinate*' not implies '*x is causally determined*'. There is no equivalence between two claims, one concerning causality and the other concerning logical necessity. Three logical principles we can find in Aristotle's discussion about the sea-battle – *the law of identity*, *the law of noncontradiction* and *the law of exuded middle* – that form kernel of logical determinism are not enough strong basis for projection of logical necessity to causal necessity or (actual) necessity of the future truths.

Let we remind that Aristotle in *de Interpretatione* (Ackrill 1963) conclusion of his opponents – that things happen of necessity – reaches apparently by reference to the premise, that of two contradictory predictions: 'one is true' (Ibid., 18b7); 'one is earlier true' (Ibid., 18b10); 'one has always been true' (Ibid., 18b10–11); 'one has been true for the whole of time' (Ibid., 19al-2). It is evident that determinist, to whom Aristotle replies, makes no explicit appeal to either causality or laws. He reckons on only logical matters. In additional inexplicit principles that Aristotle assumes (the asymmetry of time, the conservation of the past and the time direction, from left to right) it is hard to find some that leads toward the causes and causal necessity.

10.

The term (*logical*) *fatalism* – formulated across the symmetry of time and reduction of all possible worlds to the actual one – completely replaces former term the (*logical*) *determinism*. In his 'standard' argument for (logical) fatalism Taylor (1962) nowhere recalls determinism, logical or any other. It is interesting that Taylor, in the first version of his argument for fatalism, nowhere mentions laws. He only stresses the presence of causes. Latter, he declares opinion that fatalist is in fact the determinist – but such that has a certain attitude. Demarcation between *fatalism* (that claims only, in some its essential form, that future is unavoidable) and *determinism* (that lays on the causal assumption) principally seems to be superfluous. Fatalism as claim that certain events are going to happen *no matter what* and *regardless of causes* is, for him, 'enormously contrived' – "it would be hard to find in the whole history of thought a single fatalist, on that conception of it" (Taylor 1974, 55). Fatalistic claim about *unavoidability* and deterministic claim of *truth* and *necessity* coincides and are different only in regards to perspective. In the same manner as Taylor, S. Cahn identifies fatalism with a thesis that:

> "the laws of logic alone suffice to prove that no man has free will, suffice to prove that the only actions which a man can perform are the actions which he does, in fact, perform, and suffice to prove that a man can bring about only those events which do, in fact, occur and can prevent only those events which do not, in fact, occur" (Cahn 1967, 8).

This attempt is fully present today. Similar formulation of fatalism is supported by many authors. According to van Inwagen fatalism is claim that:

> "the thesis that it is a logical or conceptual truth that no one is able to act otherwise than he in fact does; that the very idea of an agent to whom alternative courses of action are open is self-contradictory" (van Inwagen 1983, 23).

Similarly, Horwich sets fatalism by these words:

> "What was true in the past logically determines what will be true in the future; therefore, since the past is over and done with and beyond our control, the future must also be beyond our control; consequently, there is no point in worrying, planning and taking pains to influence what will happen" (Horwich 1988, 29).

For J. M. Fischer fatalism is: "the doctrine that it is a logical or conceptual truth that no person is ever free to do otherwise" (Fischer 1989, 8).

11.

Taylor is only partly right when he says that "it would be hard to find in the whole history of thought a single fatalist", one who would make difference between unavoidability and necessity of universal causation. For example, during the ancient times we can find a wide range of such conceptions where *Fate* is conceptually treated different than *Necessity*. Some examples we are able to find, among others, in Cicero's *de fato* and *de divinatione*. If we wish to state some common features of different ancient sorts of fatalism we will need to represent it accord to some topological points. In the cases of so-called event-fatalism, future events are presented there as unavoidable in respect to either *time* or *place* or some *mean* (*the way* of its realization) or some *kind* of event (as it is some necessary realization of disposition, etc...) (Marko 2011a; 2011b). In some other cases, it goes only on correlation between a sign and thing signed, like in Stoical example of predictive sentence '*If Fabius was born during the Dogstar he will not die at the sea*,' where relation between antecedent state and consequent state has to be interpreted not by classical propositional implication but as some sort of *connectedness* or rather as relevant connection (or sort of responsibility relation). Since fatalism it is not always about fixed point in time, in many cases it is not always connected with examination of causes, laws, logical laws, etc. Many of these conceptions not even set aside possibility of agency, like in the case of *the conditional fate*. What ancient fatalisms have in common can be summarized by claim that it is about truth in advance – that once in the past it was true that *at least one kind of entity* (*event, occurrence*, *disposition* or *truth of proposition*) inevitable *will be actualized* (by this or that way).[9] Ancient cases of fatalism are only illustrations of treating inevitability of future without taking into account causes or laws of nature and also, in many cases, without help of supernatural forces that drives its realization.

Some forms of ancient fatalisms correspond with, for example, Earman's *naturalistic fatalism* – an event will occurs in every physically possible world, 'no matter what happened' – "for instance, that the laws of biology dictate that I am naturalistically fated to die". But in this case there is no basis for claim that this event rely on deterministic assumption: "Naturalistic fatalism in this sense neither entails nor is entailed by determinism" (Earman 1986, 18).

Logical determinism (at least in Bradley's sense) and logical fatalism (in a sense of Taylor and Cahn) seems that conceptually correspond. However, logical determinism or logical fatalism are theses that do not necessarily correspond with all forms of fatalism. Also, some particular forms of fatalism cannot simply be identified with determinism based on the principle of universal causation, as Taylor used to suppose.

Furthermore, difference between *the theological fatalism* (determinism) and *logical fatalism* Haack (1974) presents as an upgrading of argument for the logical fatalism with addition of proposition(s) with theological content (as for example 'The God is omniscient' or 'The God is omnipotent' and so on), that is formally inessential for the proof of logical part of the argument. Since the logical premisses are independent of theological this additional premiss has no role in the argument except as redundant detour from the logical character of the argument for (logical) fatalism.

12.

Several modern arguments for incompatibilism rely on explicit deterministic assumptions: for example, *The Direct Argument* and *The Consequence Argument* (van Inwagen 1983).[10] We will not here deal with these arguments and how they are inferred and defended. We are interesting only in character of its deterministic bases. Let we just see how determinism is presented in van Inwagen exposition (Ibid., 184–8).

Van Inwagen starts from a simple assumption that the past determines unique future and understands it as the thesis that there is at any instant exactly one physically possible future. According to him, determinism as a thesis

[9] It is interesting, that these approaches to fatalism are not endemic cases of ancient times. The term fatalism in above sense can be very often recognized in a current medical practice and literature devoted to analyses of patient attitudes toward hope in outcomes of treatment of their illness.

[10] Among rare exception in modern arguments for incompatibilism is, for example, Frankfurt (1969) with his cases against 'principle of alternative possibilities' where are explicitly mentioned neither causes nor laws (although conditional connector 'because' is used).

about propositions is necessary to distinguish from determinism based on the principle of universal causation. He does not feel obliged to accept the principle of universal causation and doubts that this principle even entails determinism or that determinism entails causation (place for *immanent causation* could be retain besides the fact that in complex physical events it is open question how and does causation can be distinguished). However, laws presents a firm constrain that limits our abilities. Laws are propositions that are simply *de dicto* true and they are defined as

> "any set of worlds that has as a subset the set of all worlds in which the laws of nature are the same as those of the actual world, or, as we might say, are *nomologically congruent* with the actual world."

Determinism is presented as conjunction of these two theses:

> "For every instant of time, there is a proposition that expresses the state of the world at that instant;
>
> If *p* and *q* are any propositions that express the state of the world at some instants, then the conjunction of *p* with the laws of nature entails *q*" (Ibid., 65).[11]

In respect to human agency (P) and in respect to *inability* to either change the laws of nature or the past truths, determinism is consisting in antecedental conjunction of *the past truths* (Po) and *the laws of nature* (L) and agency is conditioned by this conjunction [*i.e.* $\Upsilon((Po \& L) \rightarrow P)$].

Along this conception, claiming that human agency is determined by the past truths and the natural laws, we could find also a wide range of approaches accepting similar basis for determinism but, now in compatibilistic manner, that will allow *agent-causation* option as an intercessor link that keeps on deterministic chain. Here, the notion of causality has dominant weight and forms a central layer in these approaches. Some of compatibilists, that continues 'soft determinism' interpretation of James, will often hold both *causal determinism* and *logical determinism* to be true while others will hesitate to fully accept either the first or the second.

Frequently, contrary to formulations given above by Schlick, the laws of nature used to be understood and qualified as causes. In the recent book of Maudlin (2007, 1) we can find a thesis, widely accepted in scientific practice, that since laws are explanatory engine of occurrences in physical world they can be in some sense interpreted as responsible for occurrences: "laws of nature stand in no need of 'philosophical analysis'; they ought to be posited as ontological bedrock."[12] When we are using the laws of nature we are not analyzing or reducing one set of terms into another but we are starting from a point that these *are* actually laws. The laws cannot be reduced to other, more primitive, notions, they *are* basic ontological notions since "our world is governed by laws". In this sense, as Hoefer (2010) summarizes this position, *laws are causes* "that makes things happen in a certain ways."

[11] *Cf.* Ibid., 58; and van Inwagen 2004, 344.

[12] This position is usually called the *ontic conception* of scientific explanation according to Salmon 1998, 54.

Do we need laws if we wish to advocate determinism? Could we represent determinism without laws? It depends on how we are interpreting laws. Some interpretations of laws of nature not necessary count on the notion of cause. For example, Nagel syntactical formulation of laws of nature interprets them according to their logical function. Laws, relative to class of proprieties of some isolated system together with given the state of the system in one time *logically determine* a unique state of the system for any other time (Nagel 1999, 281). Laws are theoretical notions. According to him,

> "a theory is deterministic if, and only if, given its state variables for some initial period, the theory logically determines a unique set of values for those variables for any other period" (Ibid., 292).

If we wish to see the relation between two states as causally connected and to assume a causal version of determinism this step will pull us toward *the ontological determinism*. For this reason, Nagel insists, causality should be kept apart from determinism if we wish to escape ontological determinism. Some authors, however, prefer to retain causality, even 'a pure theoretical notion', as useful concept that has some indubitable explanatory advantages (Tooley 1987, ch. 11).

Cartwright also thinks that there are reasons to leave a notion of causation. It could be abandon in favor of theoretically more fruitful notions like capacities and structures that could be stronger scientific tools for explaining the events and for making predictions. These more appropriate notions could also replace laws and their role in science: "Capacities will do more for us at a smaller metaphysical price" (Cartwright 1989, 8). For her, all these philosophical notions (like 'universal determinism', 'law' and 'causality') are outcomes of the idea of 'nomological machine' – what is simple title for a way of organizing knowledge, "a way of categorising and understanding what happens in the world" (Cartwright 1999, 57). Science will survive without these notions.

13.

Some formulations of determinism are based on quite vaguely and controversially formulations in terms of 'event', causation', 'laws of nature' or 'prediction'. It is not always clear what the genuine characteristics of these formulations are and, besides, whether they are related to some theories or they include a wider metaphysical background. Furthermore, some of the elements used in definition (for example, 'predictability') are epistemologically oriented – like in Laplace's case – and related to scope of *abilities* (of 'intelligence' to obtain knowledge about the system in question).

More precise definition (one that partly could overrun above deficiencies) comes from Montague (in 1974, later slightly reformulated by Earman in 1986). Montague develops earlier formulations of Nagel (1953) and his idea is that determinism can be seen as propriety of a theory. Briefly, theory is interpreted by a formal semantics approach and is associated with a *class of models*. *Objects* of theory are represented as ‘systems’, *properties* are ‘states’ while *regularities* could be represented as a function ascribing a value to some point on *t* axis (and could be interpreted as ‘the laws of the theory’). A theory *T* is deterministic if any of at least two histories (*S* and *S*’) that realize (satisfy – *Rl*)[13] theory *T* and which are identical at a given time t_0 are identical at all times *t*. Theory *T* is deterministic *if and only if* all *models* of the theory that agree on the state of the world at one time [*state of S at t* – $st_S(t)$], also agree at certain other times.

Let we suppose that *S* is history, where $S = <D_1,\ldots, D_n>$ and *D* is one argument function defined at least for all real numbers *R*, and that *state of S at t* is defined as $st_S(t) = <D_1(t),\ldots, D_n(t)>$. Then,

a theory is *historically determined:*	a theory is *futuristically determined:*
If $S, S' \in Rl(T)$, t_0, $t \in R$, $t_0 < t$, and $st_S(t_0) = st_{S'}(t_0)$ *then* $st_S(t) = st_{S'}(t)$;	*If* $S, S' \in Rl(T)$, t_0, $t \in R$, $t < t_0$, and $st_S(t_0) = st_{S'}(t_0)$ *then* $st_S(t) = st^{S'}(t)$.

A theory is *deterministic* if it is both *historically* determined and *futuristically* determined, that is:

$$\textit{If } S, S' \in Rl(T), t_0, t \in R, \text{ and } st_S(t_0) = st_{S'}(t_0) \textit{ then } st_S(t) = st_{S'}(t).$$

Earman re-reads above Montague’s formulations of the basis of *deterministic theory* in terms of *physically possible worlds*. The determinism is here allowed by *the theory* by quantifying over all the physically possible worlds. Earman’s modification of Montague enables additional alternative approaches, where determinism can be interpreted not only as propriety of *the theory* alone, but also either as propriety of *the set of laws* or as propriety of *the world* or through given *the actual state of the universe* (where the history is settled by the laws even they do not determine future state of the universe) and so on.

14.

Russell’s above noted alert, related to applicability of formula outside the stabile period (where system is in process of change), today is analyzed as a wider question concerning ability to apply determinism to linear and non-linear (chaotic) dynamic systems. Montague’s approach to determinism, through its characterization as *deterministic theory* and *deterministic system*, enables some improvement in respect to Russell’s earlier forbearing. Determinism can be

[13] A formula φ of L is *realized by* a history *S* just in case there is a standard model *M* of language *L* such that *S* is *partial model corresponding to M (S = Pm(M))* and φ is true in *M*. *S realizes a class of formulas* or *theory K* (in symbols, *S* ∈ *Rl(K))* if there is a single standard model M such that *S* = *Pm(M)* and K holds in M.

compatible with description of the system that endures linear change. Linearity, as a product of differential equations, interpreted as 'additive', works well with dynamic systems seen within continual change. With linear equations, the 'state' of the system in one moment can determine the 'state' of the system at a later (or earlier) moment. It is possible by incrementally changing the variables: by adding some smaller values at lower level for the purpose to obtain higher level values of a final solution that covers the whole evolutive period.

Let we remind to above Russell's suggestion concerning 'the principle of the irrelevance of time'. His approach is suggest attempt to define determinism of changing system in terms of *actual* trajectories alone, not those *possible* (that could be infinite in number and has to be avoided). According to Montague, this approach will abolish deterministic condition that given state is always followed by the same history of state transition. Taking determinism as a modal notion van Fraassen tries to refine Russell's formulation, taking into account not only actual but possible trajectories (van Fraassen 1989, 254). The system is deterministic if two possible worlds have the same history of state transitions:

> "If u and v are possible histories, and $u(t) = v(t')$ then for all positive numbers b, $u(t + b) = v(t' + b)$."

Stone (1989) and Kellert (1993), by analysis of Laplacean determinism, attempted to notice and extract key properties of determinism that would be necessary and sufficient condition for determinism:

> "(a) there exists an algorithm which relates a state of the system at any given time to a state at any other time, and the algorithm is not probabilistic;
>
> (b) the system is such that a given state is always followed by the same history of state transitions;
>
> (c) any state of the system can be described with arbitrarily small (nonzero) error" (Stone 1989, 125).

According to them, determinism is necessary condition for predictability *but not vice versa*. Stone, Clark (1989) and Kellert, extend deterministic interpretation from *linear* to *non-linear systems* (systems usually interpreted as not fully stable or not transforming continually because they are affected by occasional 'jumps'). These systems are deterministic though predictable not *locally* but only *globally*. Their feature is that, even they behave chaotically, periodically they jump into some patterned (deterministic) behavior: even these systems have infinite possibility for movements, they are oscillating inside some steady and predictable macro patterns. Chaotic behavior of the system is due rather to epistemic reasons (or to lack of Laplacean 'demonic' abilities of observer) in respect to computability and to inability to give precise initial conditions. Determinism is here accepted as explanatory tool because some aspects of the system's evolution are coverable (not statistically or probabilistically, but) by strictly deterministic differential equations, enabling 'predictability-of-higher-order-characteristics' in respect to certain deterministic aspects of the system (related, for example, to its qualitative or topological character; (Kellert 1993).

Deterministic proprieties also could be analyzed in the scope of quantum theory, including there quantum field theory. Some newer results support the thesis that quantum theory also can be interpreted as deterministic and that this interpretation could be entirely coherent (Butterfield 2005).

15.

Up to now, we tried to briefly expose some persuasive picture of factual state in modern philosophy. Determinism is not a unique philosophical conception. There is no any representative, distinct and consensually approved formulation of determinism. More such conceptions are currently in circulation and we seen that they are different not only for the simple reason. To interpret some of these conceptions consistently it is not enough to call it merely by its common name without some additional designation. Under its common name we are usually able to find that particular segments of these conceptions can be in conflict or are mutually exclusive. To draw distinction among conceptions that we usually put in the same basket under the name of determinism, we have to stress on some necessary differences, at least by roughly highlighting their partially distinct proprieties. Our estimation is that the classification of different forms of determinisms (ranged from *ontic* to *semantic*) could be solution to keep the visibility of these differences.

Our suggestion is that the classification of different forms of determinism needs to be formed from bottom up and so we have to find a candidate for minimal common denominator of these different forms. As it seems, the basic level could rest only on notion of 'the functional determination'. Such layer will be equipped with 'order' of variables but without excessive additional features such as time-direction, as it is usually supposed by introducing indexed entities like '*state* plus *time* (of occurrence)'. This option could retain order of entities and guarantee that entities are sorted according some covering linear function. Such deterministic kernel, equipped only with *transitivity* and *continuity* (*i.e. ceaselessness*), later can ensure upgrading toward other forms of determinism that are currently in circulation among philosophers. This minimal essential form as a kernel could be compared with McTaggart's (1908, 462) idea of a 'flat' series of time or C-series. The basic kernel of determinism needs not to be understood as the determinism itself but only as a condition that enables different forms of determinism to be developed as upgrades on this basic level.

If to this grounding block we put another conceptual brick, consisting in, for example, *the universal principle of causation*, we will obtain *the causal determinism*. Further, by adding to this new composition of causal determinism another brick consisted of 'the laws of nature', new block results in *the nomological causal determinism*. Starting from the basic level again, by adding to the minimal deterministic kernel another layer, consisting of so-called 'Aristotelian laws of thought' (Principle of non-contradiction, Principle of Excluded the Middle, Principle of Bivalence, Principle of Identity), it leads to *the logical determinism*, while by adding to this composition *the principle of correspondence* (to the so called 'real correlates') we will obtain the form of *the metaphysical* version of *logical determinism* that Łukasiewicz had in mind in his critics of determinism.

Other forms of determinism can be seen as composed in the similar manner on the basic layer that retains only minimal features of the incessantness and transitive chain of ordering as a kernel.

Another our suggestion here is concerned with *a minimal formulation of fatalism*, such that would retain essential propriety the most of fatalisms have in common – *inevitability*. In that sense, minimal conditions for determinism and fatalism obviously differs. Inevitability not necessary crosses the minimal deterministic kernel. The crossing of two kernels is possible by some further layer additions. This would require one or more layer assumptions connected with either causal, logical or any other propriety of added layer. Upgrading the minimal kernel for fatalism to traditional forms, we mentioned above, that not shares some of deterministic proprieties, are possible by the same procedure of upgrading this basic sense of fatalism. If the minimal level for fatalism we seek on any more complex level, some known distinctions among fatalisms will be lost. However, in so-called logical form of fatalism usually we find some same layers that form the logical determinism where, additionally, *the logical necessity* is interpreted as *the logical inevitability*. In that sense, the term logical fatalism, as an extension of logical determinism, would be appropriate.

In both cases, either regarding determinism or fatalism, we are able to add some other features as a building block that will shape appropriate intended conception: temporal direction, causality, logical or physical proprieties, laws (laws of nature, statistical laws, probabilistic laws, etc.). Adding any of these distinct additions results in different philosophical conception and many of them are mentioned above.

Numerous composite combinations of elemental notions in a like manner, which leads to different philosophical conceptions, seem to be possible. Our suggestion here is that these combinations grown up on and are combined from the more elemental layers that need to be further investigated as the composite particles of complex conceptual structures. Every of these combinations, in whatever manner could resemble to some other compound of deterministic sort, have different meaning and leads to different philosophical standpoints.

Our primary interest here is to draw attention to various philosophical approaches to determinism and to sketch a way out for their explicit and transparent presentation but there is an open question, for some further discussion: is it possible to form exhaustive list of determinisms or fatalisms and their components?

16.

Determinisms we have outlined so far are composite in their character. Reducing them to their more elemental building blocks would be of help both for better understanding these composite conceptual structures as well as for better understanding of theories grounded on such compositions.

Could proposed classification of this sort be useful and philosophically relevant? Our position is that it can make more precise meaning of wide range of theories usually put under the same roof. Applying the proposed method it is possible to show how lot of misunderstandings could be escaped by appropriate bearing the concept of determinism and that heedless use of the notion easily leads to oversimplified results. Here we will take just one illustrative example from domain favored to scholars involved in debates on determinism.

Aristotle, in Ch. 9 of his *De Interpretatione*, seems to wish to acquaint us with his deterministic opponent. He introduces several illustrations and, among others noted there, well-known the sea-battle example, frequently studied by scholars. Commentators usually presume that mysterious Aristotle's opponent has to be some Megarian philosopher. They are simply taking over some features from proposed picture for making a profile of so-called hard exponent of determinism, *i.e.* the *(logical)* fatalism. This qualification without doubt corresponds with another example given, about truth of the future statements. Aristotle brings it into debate when imputed to his opponent claim that 10.000 years ago "it has been true that it will be true". But if to this deterministic picture we add 'the lazy argument example', the things now changes. With inclusion of this another example, the outcome would be that either Aristotle had no enough clear and refined picture about character of determinism he struggles with or his opponent is inconsistent. The most commentators of the issue, the ancient as well as modern, guided by spirit of loyalty, usually pass over this difficulty without its necessary elucidation.

Conflicted conceptual position can be exceeded only by an alternative interpretation of Aristotle text that will respect the fact that he fights there with several philosophical conceptions about the truth of future contingents. In respect to a way Aristotle exposes his problem either there are some inconsistencies in picturing his opponent *or* there is *not only one* but *at least two rivals* Aristotle is faced with. In that sense, the most acceptable exposition of the text would be, that even Aristotle is consistently faced here with *one problem* – the problem of truth of future contingencies – he fights with *two mutually different opponents*, one deterministic in orientation and another who is representative of fatalism. However, it is necessary to add here that Aristotle's fatalist is not an exponent of fatalism that rests on some deterministic layers, in the above sense of the logical fatalism. His fatalist there advocates inevitability of some future event but at the same time he allows empty space for intermediate agency maneuver period, that is, for (some restricted) free choice between alternative possibilities. This fatalism not rests on deterministic layers neither on deterministic kernel in the above sketched sense. Two conceptual profiles introduced by Aristotle, one deterministic and another fatalistic, are obviously mutually in the conflict.

Such interpretation, in some aspects, changes usual and traditional picture of this well-known text. If Aristotle polemicizes with several rivals supporting different mutually conceptions, then his intention could be understood rather as the general attack to those (who *at all* or at least *partially*) claim the actual truth of statements related to future contingents. We believe that this is its more probable version (in accordance with the principle of charity) than believing that he wastes his time by confronting with a strikingly inconsistent and unconvincing opponent.

References

Ackrill, J. L. 1963. *Aristotle Categories and De Interpretatione*. Translated with Notes by J. L. Ackrill. Oxford: Claredon Press.

Ayer, Alfred J. 1963. "Fatalism." In *The Concept of a Person and Other Essays*, 235–268. London and Basingstoke: The MacMillan Press Ltd.

Bahrdt, Carl Friedrich. 1790. *Leben und Thaten des weiland hochwürdigen Pastor Rindvigius: Ans Licht gestellt von Kasimir Renatus Denarée*, vol. 2. Verlag: Friedrich.

Bradley, R. D. 1959. "Must the Future Be What It Is Going To Be." *Mind* 68(270): 193–208. doi: 10.1093/mind/LXVIII.270.193

Bunge, Mario. 1959. *Causality*. Cambridge, Mass.: Harvard University Press.

Butterfield, Jeremy. 2005. "Determinism and Indeterminism." *Routledge Encyclopedia of Philosophy*, Volume 3. London: Routledge. doi:10.4324/9780415249126-Q025–1

Cahn, Steven M. 1967. *Fate, Logic, and Time*. New Haven: Yale University Press.

Cartwright, Nancy. 1989. *Nature's Capacities and Their Measurement*. Oxford: Clarendon Press.

Cartwright, Nancy. 1999. *The Dappled World: A Study of the Boundaries of Science*. Cambridge: Cambridge University Press.

Cassirer, Ernst. 1936. *Determinismus und Indeterminismus in der modernen Physik. Historische und systematische Studien zum Kausalproblem*. Göteborg: Elanders Boktryckeri Aktiebolag.

Cassirer, Ernst. 1956. *Determinism and Indeterminism in Modern Physics: Historical and Systematic Studies of the Problem of Causality*. New. Haven: Yale University Press.

Cicero, Marcus Tullius. 1949. *On Invention. The Best Kind of Orator. Topics*. Translated by H. M. Hubbell. Cambridge: Harvard University Press.

Clark, Peter. 1989, "Determinism, Probability and Randomness in Classical Statistical Physics." In *Imre Lakatos and Theories of Scientific Change*, edited by Kostas Gavroglu, Yorgos Goudaroulis, and Pantelis Nicolacopoulos, 95–110. Dordrecht, Boston, London: Kluwer Academic Publishers.

du Bois-Reymond, Emil. 1886. *Über die Grenzen des Naturerkennens* (Versammlung Deutscher Naturforscher und Aerzte, 14 August 1872) *Reden von Emil Du Bois Reymond*. Leipzig: Verlag von Veit & Co.

Dummett, Michael. 1964. "Bringing About the Past." *Philosophical Review* 73(3): 338–359. doi: 10.2307/2183661

Earman, John. 1986. *A Primer on Determinism*. Dordrecht: D. Reidel Publishing Company.

Fischer, John M. 1989. *God, Foreknowledge, and Freedom.* Stanford: Stanford University Press.

Frankfurt, Harry J. 1969. "Alternate Possibilities and Moral Responsibility." *Journal of Philosophy* 66(23): 829–839.

Gellius, Aulus. 1927. *The Attic Nights*. Translated by John C. Rolfe. Cambridge: Harvard University Press; London: William Heinemann, Ltd. http://www.perseus.tufts.edu/hopper/text?doc=Perseus:text:2007.01.0071

Glymour, Clark. 1971. "Determinism, Ignorance, and Quantum Mechanics." *The Journal of Philosophy* 68(21): 744–751. doi: 10.2307/2024947

Haack, Susan. 1974. "On a Theological Argument for Fatalism." *The Philosophical Quarterly* 24(95): 156–159. doi: 10.2307/2217720

Hacking, Ian. 1983. "Nineteenth Century Cracks in the Concept of Determinism." *Journal of the History of Ideas* 44(3): 455–475.

Hamilton, William. 1846. *The works of Thomas Reid, D.D.; Now Fully Collected, with Selections From His Unpublished Letters.* Edinburgh: Maclachlan and Stewart.

Hegel, Georg Wilhelm Friedrich. 1816. *Wissenschaft der Logik*. Nürnberg: Schrag.

Herbart, Johann Friedrich. 1842. *Kleinere philosophische Schriften und Abhandlungen: nebst dessen wissenschaftlichem Nachlasse*. Bd. 1, Ed. Gustav Hartenstein, F. A. Brockhaus.

Herbart, Johann Friedrich. 1843. "Aphorismen und kürzere Fragmente, a. Zur *Einleitung in die Philosophie*." In *Kleinere philosophische Schriften und Abhandlungen: nebst dessen wissenschaftlichem Nachlasse*. Bd. 3, Ed. Gustav Hartenstein, F. A. Brockhaus.

Herbart, Johann Friedrich. 1812. *Bemerkungen über die Ursachen, welche das Einverständnis über die ersten Gründe der prakt. Philosophie erschweren.*

Hoefer, Carl. 2010. "Causal Determinism." In *The Stanford Encyclopedia of Philosophy* (Spring 2010 Edition), edited by Edward N. Zalta. Stanford: Metaphysics Research Lab, Center for the Study of Language and Information, Stanford University. https://plato.stanford.edu/entries/determinism-causal/

Horwich, Paul. 1988. *Asymmetries in Time: Problems in the Philosophy of Science*. Cambridge: MIT Press.

James, William. 1907 [1884]. "The Dilemma of Determinism." In *The Will to Believe and Other Essays in Popular Philosophy*, 145–183. New York: Longmans Green and Co.

Jordan, Zbigniew. 1963. "Logical Determinism." *Notre Dame J. Formal Logic* 4(1): 1–38. doi: 10.1305/ndjfl/1093957391

Kant, Immanuel. 1793. *Die Religion innerhalb der Grenzen der bloßen Vernunft*. Königsberg: Nicolovius.

Kant, Immanuel. 1804. *Über die ästhetische Darstellung der Welt als das Hauptgeschäft der Erziehung.*

Kellert, Stephen H. 1993. *In the Wake of Chaos: Unpredictable Order in Dynamical Systems*. Chicago: University of Chicago Press.

Krug, Wilhelm Traugott. 1827–1829. *Allgemeines Handworterbuch der philosophischen Wissenschaften nebst ihrer Literatur und Geschichte*. Vol. i-v. Leipzig: F. A. Brockhaus.

Langer, Susanne K. 1936. "On a Fallacy in 'Scientific Fatalism.'" *Ethics* 46(4): 473–483. doi: 10.1086/208326

Laplace, Pierre-Simon. 1902. *A Philosophical Essay on Probabilities.* Translated by Frederick Wilson Truscott and Frederick Lincoln Emory. New York: John Wiley & Sons; London: Chapman & Hall Limited.

Livy, T. *Ab urbe condita libri*. http://www.thelatinlibrary.com/liv.html

Łukasiewicz, Jan. 1990. "On Determinism." In *Selected Works*, edited by Ludwik Borkowski, 110–128. Amsterdam and London: Nort-Holland Publishing Company.

Marko, Vladimír. 2011a. "Looking for the Lazy Argument Candidates 1." *Organon F* 18(3): 363–383

Marko, Vladimír. 2011b. "Looking for the Lazy Argument Candidates 2." *Organon F* 18(4): 447–474.

Maudlin, Tim. 2007. *The Metaphysics Within Physics*. Oxford: Oxford University Press.

McTaggart, John Ellis. 1908. "The Unreality of Time." *Mind* 17(68): 457–474.

Mill, John Stuart. 1843. *A System of Logic*. Vol. 1. London: John W. Parker, West Strand.

Nagel, Ernest. 1953. "The Causal Character of Modern Physical Theory." In *Readings in the Philosophy of Science*, edited by Herbert Feigl and May Brodbeck, 419–437. New York: Appleton-Century-Crofts.

Nagel, Ernest. 1999. "§V: Alternative Descriptions of Physical State." In *The Structure of Science: Problems in the Logic of Scientific Explanation* (2nd ed.), 285–292. Indianapolis: Hackett.

Popper, Karl. 1982. *The Open Universe: An Argument for Indeterminism*. Edited by W.W. Bartley III. London: Hutchinson.

Priestley, Joseph. 1779. *A Letter to the Rev. Mr. John Palmer, In Defense of the Illustrations of Philosophical Necessity.* 8vo. Bath.

Priestley, Joseph. 1780. *A Second Letter to the Rev. Mr. John Palmer: In Defense of the Doctrine of Philosophical Necessity*. 8vo. Lond.

Russell, Bertrand, and Alfred North Whitehead. 1910–13. *Principia Mathematica*, Cambridge: Cambridge University Press.

Russell, Bertrand. 1917. "On the Notion of Cause." In *Mysticism and Logic and Other Essays*, 180–208. London: George Allen & Unwin Ltd.

Ryle, Gilbert. 1953. *Dilemmas: The Tarner Lectures*. Cambridge: Cambridge University Press.

Salmon, Wesley C. 1998. *Causality and Explanation*. Oxford: Oxford University Press.

Schlick, Moritz. 1931. "Das Kausalität in den gegenwärtigen Physik." *Naturwisseschaften J. H.* 7(19): 145–162.

Schlick, Moritz. 1932. “Causality in Everyday Life and in Recent Science.” *University of California Publications in Philosophy* 15: 99–125.

Schlick, Moritz. 1939. “When Is a Man Responsible? ” In *Problems of Ethics*, 143–156. New York: Prentice Hall.

Schulz, Hans, Otto Basler, Gerhard Strauss, eds. 1999. *Deutsches Fremdwörterbuch*. Berlin, New York: de Gruyter.

Snell, Christian Wilhelm. 1789. *Über Determinismus und moralische Freiheit*. Offenbach: bey Ulrich Weiss und Carl Ludwig Brede.

Stone, Mark A. 1989. “Chaos, Prediction and Laplacean Determinism.” *American Philosophical Quarterly* 26(2): 123–131.

Suppes, Patrick. 1993. “The Transcendental Character of Determinism.” *Midwest Studies in Philosophy* 18(1): 242–257. doi: 10.1111/j.1475–4975.1993.tb00266.x

Taylor, Richard. 1962. “Fatalism.” *Philosophical Review* 71(1): 56–66. doi: 10.2307/2183681

Taylor, Richard. 1974. *Metaphysics*. Englewood Cliffs, N.J.: Prentice-Hall.

Thomson, William. 1855. *Oxford Essays*.

Tooley, Michael. 1987. *Causation: A Realist Approach*. Oxford: Clarendon Press.

van Fraassen, Bas. 1989. *Laws and Symmetry*. Oxford: Clarendon Press.

van Inwagen, Peter. 1983. *An Essay on Free Will*. Oxford: Oxford University Press.

van Inwagen, Peter. 2004. “Freedom to Break the Laws.” *Midwest Studies in Philosophy* 28(1): 334–350. doi: 10.1111/j.1475–4975.2004.00099.x

Wallace, David Foster. 2011. *Fate, Time and Language – An Essay on Free Will*. New York: Columbia University Press.

Weismann, Friedrich. 1959. “How I See Philosophy?” In *Logical Positivism*, edited by Alfred J. Ayer, 354–380. New York: Free Press.

Wołenski, Jan. 1996. “An Analysis of Logical Determinism.” Paper presented at the conference *Łukasiewicz in Dublin*, at University College Dublin, July 1996.

Vlasta Sikimić, Andrea Berber
University of Belgrade

Original Scientific Paper
UDK 179.3
340.134:[502.211:592/599

CONTEMPORARY CHALLENGES IN MORAL AND LEGAL TREATMENT OF ANIMALS

Abstract*: The purpose of the present paper is to demonstrate the inconsistencies between ethical theory and legal practice of animal treatment. Specifically, we discuss contemporary legal solutions, based on three case studies – Serbian, German and UK positive law, and point out the inconsistencies in them. Moreover, we show that the main cause of these inconsistencies is anthropocentric view of moral relevance. Finally, when it comes to the different treatment of animals living in the wild and domestic animals, we show that the current theoretical explanations are unsatisfactory.*

Keywords*: Animal rights, animal welfare, anthropocentrism, duties, positive law.*

1. Introduction

Compared to the frequency of dilemmas concerning the treatment of animals in everyday life, the question of the moral status of animals has considerably been neglected in philosophical literature. Korsgaard (2011) points out that ethical decisions on how to treat animals are made on a daily basis, whereas the moral dilemmas discussed in depth in literature, such as sacrificing an innocent person for the well-being of others, are situations which occur rarely (or never) in an average human life-span. Also, positive laws in a growing number of countries require protection of an animal's well-being. The scope of these laws can be very wide and can provide an array of duties that humans have towards animals. For instance, positive laws that guarantee the well-being of animals, prohibit torture, neglect, and abandonment of animals, etc.[1] In section 3, we will discuss in more detail particular legal solutions and present what we believe to be their correct ethical justification. The third reason why this topic deserves philosophical analysis and debate is an obvious one: we have strong intuitions and feelings concerning animal protection and welfare provision. In particular, because of these intuitions and feelings we are faced with everyday dilemmas with respect to the treatment of animals. It is precisely because of these intuitions and feelings that the number of organisations protecting and

[1] E.g. the Animal Welfare Act 2006 of the United Kingdom, the German Animal Welfare Act 1998, the Animal Welfare Act of the Republic of Serbia 2009.

advocating animal welfare and rights has been growing, and in the final instance positive laws have been brought.

The importance of philosophical analysis and ethical grounding lies in making a coherent and justifiable framework for human actions and should serve as a basis for the further development of animal-human interaction. Having in mind both the frequency of ethical dilemmas concerning human treatment of animals and the legal relevance of this subject, we conclude that a serious philosophical analysis is timely. The aim of the present paper is to (1) draw attention to the debate surrounding the moral status of animals, (2) point out and underline certain aspects of the debate, i.e. discrepancies among our intuitions, positive law, and main ethical approaches, and (3) provide some directions that might be helpful in addressing this issue.

In the rest of this section, we will briefly reflect upon the treatment of animals in the Western philosophical tradition. In the next section, we will explain the methodology employed in our paper and its organization. However, we would firstly like to advocate for a small, yet important shift in terminology. We do not refer to an animal as an "it". We believe that once animals are granted moral status, they should not be referred to as objects." In the history of philosophy, there have been many examples of denying the moral status to animals. Some philosophers, like René Descartes (1989) and Immanuel Kant (1997) were reluctant to grant the moral status to animals, because they lack the typically human characteristics such as rationality, consciousness, or autonomy. Others denied the moral standing of animals on the ground of their religious or philosophical views according to which animals are inferior to human beings and exist to serve them. One of the earliest expressions of this view can be found in Aristotle, and afterwards in the Christian philosophical tradition (see Balme 1991). Some of these approaches consider harming animals as morally detrimental in an indirect fashion, only if it has a negative influence on the treatment of humans. Animals are treated as mere objects or instruments for satisfying human goals. This kind of an approach is an unquestionable example of anthropocentrism.

One of the famous positive examples in the history of philosophy before the twentieth century is Jeremy Bentham (2007) who thought that moral considerability of animals is grounded in their sentience, i.e. their capacity to experience pleasure and pain. In the second half of the twentieth century, things started to change for the better, philosophers began to question such a simplified picture of morality in which there is no place for anyone but humans, and which does not incorporate and explain the complexity of human interaction with other species. We can distinguish three major theoretical approaches that recognise the moral significance of non-human animals: the animal welfare approach, the animal rights approach, and the environmental ethics approach. The animal welfare approach, whose most prominent defender is Peter Singer, is essentially utilitarian. According to this approach, equal interests of all sentient beings must be taken into account equally in the utilitarian calculus. This means that human interests do not have primacy over the interests of members of any other species just because they are *human* interests. However, in accordance with utilitarian logic, interests of animals can be outweighed by the aggregate interests of others (see Singer 1976). The animal rights approach, unlike the animal welfare approach with which it is often confused, has a significant deontological component. This means that there is a privileged group of beings (something like Kant's Kingdom of Ends) who possess rights that cannot be overpowered by the aggregate interests of other beings. The criterion for membership in this group is subjecthood i.e. the capacity to have propositional attitudes, emotions, self-consciousness, and awareness of the future. Every member of this group is entitled to unconditional protection of her basic interests, such as the right to life and integrity of consciousness and activity, which includes freedom to exercise the specific capabilities of a species (see e.g. Regan 1983). It is very important to understand the difference between the animal welfare and the animal rights position. The first position allows sacrificing important interests of animals, even their lives, to advance the overall well-being of other sentient beings, while the second never allows sacrificing animal rights, nor using animals as a mere means to an end, no matter how important the end itself is. Lastly, the environmental ethics approach is holistic, it places value on a live natural aggregate – a species, or on a live natural system – an ecosystem, or on the whole biosphere. From this position an individual animal is not important except for her role in the larger unit. It is considered to be justified to kill an individual or some larger number of individuals if that will contribute to the preservation of diversity within the biosphere (see Brennan and Lo 2011).

2. The methodological approach and the division of labour

The problem with the theoretical treatment of the moral status of animals is that we lack adequate moral concepts for analysing it. Our moral concepts are designed primarily for human animals. Non-human animals in some sense exist on the margins of our current theoretical moral framework. Everyday treatment of animals is not grounded on a coherent theoretical framework but on feelings, intuitions, and pragmatic considerations. We do not want to deny the importance of these feelings, intuitions, and pragmatic considerations; on the contrary, we want to emphasise the importance of developing a theoretical approach that takes them seriously and gives consistent theoretical solutions based on them.

We provide an overview of the debate surrounding the question of the contemporary moral status of animals focusing on the discrepancy among ethical theory, actual legal practice, and our intuitions. First, in section 3 we turn to the existing positive laws and show that they are mainly grounded on the animal welfare approach, i.e. utilitarianism. Therefore, in section 4, we raise questions that we believe to be problematic for utilitarianism as the potential ground for the moral relevance of animals. Namely, we show that any anthropocentric criterion including the feeling of (human-type) pain is arbitrary as a ground for the moral relevance of animals. Further on, we notice that not only is there a significant difference in the treatment of pets in opposition to other domestic animals, but there is also a big distinction in the legal and theoretical treatment of animals in human society and animals living in the wild. In section 5, we present a solution proposed by Elisabeth Anderson (2004) that emphasises precisely this social dimension. However, it is questionable whether this discrepancy in treatment is coherently grounded. In the end, we believe that it is safe to conclude that the current grounding of the moral treatment of animals is unsatisfactory and that an adequate one would need to answer all the questions raised in this paper, as explained in the concluding section.

3. The analysis of existing laws

Clearly, the existing laws rely mainly on the animal welfare approach, i.e. utilitarianism. This can be seen because of the emphasis given to the fact that animals are capable of experiencing pain or suffering.[2] Furthermore, animals are protected from unnecessary harm, i.e. "no one may cause an animal pain, suffering or harm without good reason" as stated in the German Animal Welfare Act from 1998, article 1. This leaves room for exposing animals to suffering for the "greater good", for instance, particular types of medical testing, which is precisely the moderate position of the animal welfare approach. On the other hand, the animal rights approach would not allow for such exceptions. According to the animal rights approach, a right granted to an animal or species needs to be universal in the same way as human rights are. Such a right can only be overpowered by the right of another individual, for instance, harming someone in self-defence. Since people seem to be reluctant to give up animal testing and the like that result from human primacy over animals, they are equally reluctant to grant full rights to animals.

According to utilitarianism, duties towards animals are based on the fact that animals are sentient i.e. capable of experiencing pleasure and pain. In this sense animals are bare moral patients. However, we wonder whether there are reasons to grant animals a stronger moral status. For example, the fact that animals in some situations protect and help us and each other might be one of the reasons. Mark Rowlands (2012) argued that "animals can act morally in the sense that they can act on the basis of moral emotions – emotions that possess identifiable moral content" (p. 15). According to Rowlands, animals are not bare moral patients, nor moral agents, but they are moral subjects, since they sometimes act upon moral emotions, such as tolerance, compassion, empathy, etc. On the other hand, since they are not moral agents they cannot be held responsible for their actions.

Another important note about the contemporary positive law is the fact that humans are obliged not only not to harm animals, but also to provide for their welfare. For instance, according to the Animal Welfare Act 2006 of the United Kingdom, article 9.2, humans responsible for an animal have legal duty to provide a suitable environment, diet for the animal, protect her from

[2] E.g. in the Animal Welfare Act 2006 of the United Kingdom, article 4.

pain, injuries, diseases, etc. The theoretical grounding of such *positive* rights for animals is challenging. In section 4, we present Elisabeth Anderson's (2004) position that provides significant support for positive rights towards domestic animals and captives from the wild. Also, it is questionable whether the very distinction between positive and negative rights and respective duties is actually meaningful and informative. We turn to this in sections 6.

Though mainly based on utilitarianism, the positive legal solutions are not consistently utilitarian. When it comes to the regulations of hunting, the logic of the ecological approach is used in the sense that the hunt is allowed as long as it does not endanger the survival of the species, e.g. the Law on Venison and Hunting of the Republic of Serbia, 2010, article 2. Clearly, this legal solution gives primacy to the species over an individual animal and is therefore in line with the ecological approach. Utilitarianism, on the other hand, does not recognise populations, species and ecosystem as morally relevant entities. What utilitarianism and ecological approach have in common is the fact that an individual can be sacrificed for the "greater good", i.e. for the good of a group. However, justifications for these sacrifices are different. Ecological ethics approach advocates for controlling the population of wild animals by hunting them down or even by biological war, e.g. by deliberately infecting a population with a virus with the goal of radically reducing it (Berber & Sikimić 2016). For a utilitarian it is inacceptable to scarifies an animal only for the reasons of population control. A consistent utilitarian needs to perform a complex utilitarian calculus before making a decision to sacrifice any individual.

Moreover, there is a difference in the legal treatment of pets and animals kept in zoological gardens and other domestic animals, or captives from the wild used in experiments. According to the Animal Welfare Act of the Republic of Serbia 2009, article 15, killing pets and animals kept in zoological gardens is forbidden, unless in the case of serious illness. Other domestic animals, as well as captives from the wild used in experiments, can be killed. It is not clear which ethical criterion can explain why a pig bought in the pet shop should be treated differently than the one living on a farm.

All things considered, utilitarianism as the dominant and rather coherent doctrine cannot be dismissed easily when it comes to grounding of the animal welfare acts. Utilitarian grounding of the moral treatment of animals is based on the fact that animals, express pain and pleasure in a similar fashion as humans do. In such a justification we see that animals have moral status guaranteed because they are recognised as similar to humans. Such an anthropocentric picture might lead to the conclusion that animals who do not have either humanlike demonstrations of pain and pleasure, or the physiology that is similar to the human one, should enjoy weaker moral status. This problem is the most evident when it comes to psychological and emotional pain and pleasure. We might not always be able to identify the pain and the pleasure of a being with the psychology and emotional structure radially different from ours. Thus one of our main concerns when it comes to the utilitarian criterion is that its implementation leaves room for arbitrariness and anthropocentrism. In the next section, we turn to anthropocentrism as a possible ground for the moral treatment of animals and criticise it.

4. Is anthropocentrism arbitrary?

Questions such as whether animals are like us, in which respect they are like us, and to which extent they are like us, are often in focus of the debates on the moral status of animals. Taking similarity to human beings as the only or main criterion of the moral value of a being is what is usually called "anthropocentrism". There are several problems with anthropocentrism that we want to underline. Firstly, individual capacities of a being are not the only source of moral duties towards her. This point is strongly supported by the so-called argument of marginal cases. According to the argument of marginal case, whenever one tries to justify asymmetry in treatment between human and non-human animals based on some trait or collection of traits that humans possess, it will be possible to point out to a human lucking this trait or the collection of traits. For instance, if one wants to claim that humans should have special moral status based on their cognitive abilities, we can point out to human beings who for some reason do not have the common human cognitive abilities, for example, the mentally challenged or people suffering from dementia, and towards whom we nevertheless have duties (Anderson 2004). Thus, the argument of marginal cases is based on the difficulty of finding a relevant characteristic or collection of them that captures all and only humans. Secondly, we should ask ourselves what would justify taking similarity to human beings as the criterion of moral value. The inclination to be more compassionate towards beings who are similar to us may be natural, but we cannot take this psychological tendency as the only source of moral value without a justification. Even if we presuppose that humans are superior to all other species, this does not entail that other species do not have a moral value of their own. More importantly, we hold that it is mischievous to think in categories of superiority and inferiority in this context. On the other hand, if we dismiss any anthropocentrism, it is not clear what we should take as the criterion for evaluation of non-human life. The ecological approach grants certain rights to all the living species, however, it still uses a (terrestrial) biological definition of life. An approach based on pure intelligence of an individual could grant moral status to non-biological individuals in the terrestrial sense, however, the question is whether humans are in principle able to comprehend and acknowledge intelligent forms that are potentially radically different than ours.

Note that like utilitarianism, the animal rights approach is also based on anthropocentric criteria, but these criteria focus on the cognitive capacity of an animal. The animal rights approach favours animals that have similar cognitive capacities to human ones.

The fundamental question concerning anthropocentrism, granted the basic assumption that similarity to humans is crucial for determining the moral status of a being, is what kind of similarity is more important: an emotional or a cognitive one. We can imagine that there are beings (extra-terrestrials or ones with artificial intelligence) who are hyper rational, but are not able to experience pleasure and pain. The question is whether such hypothetical beings would have any duties towards humans and whether humans would have any duties towards them.

Utilitarianism gives a moral status to all beings who are capable of experiencing human-like pain and pleasure. According to utilitarianism, humans need to guarantee welfare to an animal, or any other being with the above characteristics, even if some of them are incapable of acting reciprocally (i.e. they are mere moral patients). Now a serious question arises: would a hypothetical hyper rational being with no emotions need to have any duties (positive or negative) towards other (less rational) beings, and why?

If the utilitarian maxim aspires to be the unique and universal moral ground, it seems that according to utilitarianism purely rational beings that do not experience pleasure or pain in a human-like fashion, would need to treat all sentient beings morally, but would not need to treat each other in such a manner. Moreover, none of the purely rational beings would be a moral patient. Thus, rationality would need to give primacy to sentience, and that is highly counterintuitive. It is unclear whether purely rational beings could have their own morality based on rationality and even grant humans some moral rights, but at same time humans would not treat them as moral patients in return. They would represent a specific type of moral agents who are not moral patients. It is also unclear why such rational beings would accept that sentient beings count more in the utilitarian calculus than they do themselves, especially if they are incapable of feeling any empathy.

MacDonald Glenn (2002) offered a system that can incorporate both moral treatment of animals and of artificial intelligence. She claims that the notion of personhood cannot be defined in an absolute manner. According to this approach, each individual enjoys legal rights based on the set of traits she possesses, while she does not enjoy rights related to the set of traits she lacks. For instance, minor children do not have the right to vote, but enjoy a whole array of other rights. In this fashion, some legal rights can be granted also to artificially intelligent beings, e.g. the right not to be destroyed. Her approach offers a basis for changing the legal norms. It is important to develop an approach in ethics that could treat both animals and artificial intelligence; the position of MacDonald Glenn is a promising starting point for a comprehensive theory. We would like to point out that a candidate for such a criterion with a stronger philosophical component could be based on the family resemblance in the sense of late Wittgenstein (1986). In particular, animals share a set of traits with humans that make them morally relevant and the potential artificial intelligence might share a set of traits with humans that make them morally relevant. However, these two sets do not need to coincide. Lastly, it is conceptually possible that there are some properties that make an individual morally relevant, but that humans do not posses them, though humans certainly posses other properties which make them morally relevant.

Still, all this does not mean that the utilitarian hypothesis which claims that beings who experience pain need to be treated as moral patients, and therefore be granted certain. moral status, should be dismissed. It rather means that such a justification is insufficient to capture the phenomenon and cannot be accepted as the unique and universal moral principle, though it is very useful in capturing certain aspects of our ethical considerations. Apart from the problem of the different treatment between pets and domestic animals used for farming, the discrepancy in the treatment of animals is most prominent when it comes to the differences between the treatment of animals living in human society and the ones living in the wild. Clearly, the utilitarian criterion is not sufficient to justify these differences. In the next section, we turn to the problem of justification of the difference in the treatment between domestic and wild animals.

5. Domestic and wild animals

Elizabeth Anderson (2004) connects right to provision with membership in society, because, as she argues, only social membership can vindicate individual moral right to provision by specifying who has the obligation to provide the necessities of life to which individual. Human beings are automatically included into human society. However, two other classes of animals have also been incorporated into human society – domesticated animals and captives from the wild, which means that animals from these classes also have right to provision. According to Anderson, individual animals living in the wild, in general, do not have the right to our protection or provision even in case they need it for survival. We think that Anderson's position has a certain intuitive plausibility. It is true that we have a larger list of obligations towards domestic animals than towards those living in the wild. Nevertheless, we want to draw attention to some principal difficulties concerning this position. First, Anderson's position is related to the suspicious distinction between positive and negative rights. The usual explanation is that positive rights are the ones that oblige an action, while negative rights are the ones that oblige an inaction. In the next section we refer to the problems of this distinction. Second, we are not confident that we never have a duty to help a wild animal. What if some wild species or individual is endangered due to human activity, for instance, pollution or destruction of the animals' natural habitat? In that case, it is not justified to say that our only duty towards wild animals is to let them be. Humans at least have a positive duty to compensate for the damage they have caused.

We agree with Anderson that animals are not in a position to protect their well-being inside the human society in an adequate manner and therefore humans are obliged to help them do so. For instance, some positive laws, such as the Animal Welfare Act of the Republic of Serbia 2009, article 3.4, demand reporting animal abuse and neglect, precisely because the animal is not capable of filing such a report herself. However, we question the position that wild animals are never entitled to our protection and help. Besides the intuition that we have impulse to help an animal in need, even if she is living in the wild, and the fact that some wild animals behave protectively towards humans, e.g. dolphins, if the distinction between positive and negative duties is dismissed, it becomes arbitrary which class of duties one has towards different types of animals. The same objection can be raised against the approach developed by Donaldson and Kymlicka (2011). On the positive side, in addition to explaining the difference in the treatment between the animals that are part of human society and the animals living in the wild, they introduce a specific category of the animals who live at the margins of human communities and are not completely independent of them, such as foxes and racoons. According to this approach, the animals living in the wild enjoy a certain type of sovereignty, which means that humans are not allowed to endanger their habitats. Still, it remains unclear how much humans are allowed to interfere in the wild life in order to help or protect the animals living in the wild. Moreover, humans already indirectly affect the whole biosphere, e.g. by pollution and global warming.[3]

6. Wild life and positive duties

We will only tackle the big discussion on the distinction between positive and negative duties in order to demonstrate that it is not as unproblematic as it may *prima facie* seem. If this distinction is dismissed, then obviously there is no ground for the different treatment of animals living in human society (domesticated animals and captives from the wild) and animals living in the wild based on it. In the case of domestic animals, as already mentioned, both positive and negative duties are covered by positive laws granting their well-being. For instance, the domestic animal is entitled to proper treatment in the sense that she should not be neglected, nor abandoned. However, only the “negative” rights are assigned towards the great apes living in the wild, such as the right not to be killed.[4] Positive law also forbids for a wild animal that has grown up or spent significant time in captivity to be returned into the wilderness without special preparation that will guarantee successful integration of the animal in the new environment, e.g. the German Animal Welfare Act from 1998, article 3.

The intuitive ethical grounding that Anderson (2004) provides states that domestic animals and captives from the wild, enjoy positive rights, while animals living in the wild only enjoy negative rights. The social dimension brought up by Anderson additionally emphasises our responsibility towards animals incorporated into human society. However, we are still responsible for our environment. There is also no clear-cut differentiation between parts of the environment that are affected by humans in direct or indirect way, e.g. global warming and parts that are not. Therefore, it is unclear how much humans

[3] For more criticism of Donaldson’s and Kymlicka’s (2011) approach, see Cochrane (2013).

[4] For more details about granting basic rights to the great apes, see Singer (2006).

affect all of the wild animal habitats and what their corresponding responsibility towards the animals living in the wild is. On the other hand, our duties towards wild animals are also connected to the question of whether there is actually a big difference between positive and negative duties.

As intuitive as it may sound, it is questionable whether this distinction is real in the sense that it is meaningful and informative. Trivially, any law can be stated both in a negative and in a positive form. This is a fact granted by logic. For instance, the paradigmatic negative duty "do not kill" can be transformed into "let everyone live". Also, one famous positive duty "help others in need" can be transformed into "do not abandon others in need". Obviously, this is not a sufficient reason to reject the distinction as meaningless. One needs to consider the understanding of the action/inaction distinction. What we can conclude from the argument above is that, in formal terms, every "inaction" can be seen as a type of "action" and vice versa. Another line of argumentation claims that the distinction is not just unjustified but even harmful. For instance, Shue (1980, ch. 1–2) argued that the distinction between negative and positive rights can be harmful, because none of the necessary duties should be neglected. Moreover, this is in line with Kant's theory. Even though Kant's (2003) distinction between perfect and imperfect duties is considered to be the ground for differentiation between positive and negative rights, it is of crucial importance to keep in mind that all the duties are binding for him. This means that both type of duties are "the first class citizens" of Kant's moral architecture and it is unclear why one should have only one type of duty towards specific individuals.[5]

In a nutshell, the question of the moral status of non-human animals living in the wild is a difficult one. The main question within this debate is whether we have both positive and negative duties towards animals living in the wild. The advocates of the ecological approach desire that humans help all species in the ecosystem and assign them moral value based on the diversity of the biosphere. However, it is questionable whether the diversity of the biosphere can serve as an ethical ground. Moreover, if one wants to consistently apply the environmental approach, she would need to apply it also to human species. Clearly, humans are the ones that very much endanger the ecosystem. Yet, the principle of exterminating humans when they endanger other species or the ecosystem does not seem appealing. One the other hand, if we accept that humans only have moral responsibility for animals whose lives they have affected, it is hard to establish whether there is any large part of nature left intact and uninfluenced by humans directly or indirectly. If one accepts the idea that there are animals living in the wild that are not affected by humans, we would still feel the urge to help them if they are thirsty or badly injured,

[5] The prefect duties are the one for which it is logically impossible to be universalized, while for the imperfect duties this is practically impossible. For more details, see Babić 1991.

and it is still questionable whether such behaviour on our part is only based on our good will, i.e. whether it is morally neutral or actually morally required. Such a requirement would mean allowing positive duties towards wild animals. Lastly, if there are reasons for questioning the very distinction between positive and negative duties, it becomes unclear how to ground and justify restricted rights for wild animals.

7. Concluding remarks

In this paper, we have drawn attention to the question of the moral status of animals, provided an overview of the debate, and underlined certain aspects of it, which we believe to have been inadequately treated so far. Our main focus were inconsistencies of contemporary legal solutions w.r.t. ethical grounding. Finally, we have provided some guidelines that might be helpful in addressing these inadequacies.

The question of treating animals is a prominent question in our daily lives and as such deserves an appropriate theoretical analysis. We have analysed the question of the moral status of animals focusing on the two main alternatives, the utilitarian, i.e. the animal welfare approach, and the animal rights approach. We have pointed out that both of these approaches base the moral treatment of animals on anthropocentric criteria. However, when it comes to determining the set of morally relevant individuals, it is not clear what an alternative to anthropocentrism would be. Also, we demonstrated that certain positive laws (three case studies) appeal to utilitarian principles. Then we turned to the particular theoretical solutions that base the moral treatment of animals on duties and membership in human society. The main problem with these solutions is that they advocate different duties towards animals living in the wild and animals living in human society. We argued that this difference in treatment is not well founded. In the end, we believe that it is safe to conclude that the current grounding of the moral treatment of animals is unsatisfactory and that it is timely to provide answers to the questions raised in this paper.

Besides the distinction between animals living in the wild and the ones living in human society, the following categories of animals that are incorporated in human society can be listed: pets, captives from the wild, domestic animals in the strict sense, animals used for experiments, military purposes, etc. The question is how to justify the difference in treatment of these distinct categories inside human society. It would be beneficial to provide non-arbitrary criteria for treatment of all animals. However, the question of the moral status of animals cannot be answered in isolation, therefore it is important to first provide answers to fundamental ethical questions such as the grounding of moral principles on rationality, emotions or their combination, the question of the meaningfulness of the distinction between positive and negative duties, etc.

References

Anderson, Elizabeth. 2004. "Animal Rights and the Values of Nonhuman Life." In *Animal Rights: Current Debates and New Directions*, edited by Cass R. Sunstein and Martha C. Nussbaum, 277–298. Oxford: Oxford University Press).

Balme, D. M. 1991. Aristotle: History of Animals, Books VII-X. Cambridge, MA: Harvard University Press.

Bentham, Jeremy. 2007. *An Introduction to the Principles of Morals and Legislation*. New York: Dover Publications.

Berber, Andrea, and Vlasta Sikimić. 2015. "The Moral Status of Animals." *Collected Papers of the International Philosophical School Felix Romuliana 2008–2015*, edited by Slobodan Divjak and Jovan Babić, 479–485. Zaječar: Center for Culture and Tourism [in Serbian].

Brennan, Andrew, and Yeuk-Sze Lo. 2011. "Environmental Ethics." In *The Stanford Encyclopedia of Philosophy* (Fall 2011 Edition), edited by Edward N. Zalta. Stanford: Metaphysics Research Lab, Center for the Study of Language and Information, Stanford University. http://plato.stanford.edu/archives/fall2011/entries/ethics-environmental

Cochrane, Alasdair. 2013. "Cosmozoopolis: The Case Against Group-Differentiated Animal Rights." *Law, Ethics and Philosophy* 1: 127–141.

Descartes, René. 1989. "Animals are Machines." In *Animal Rights and Human Obligations*, edited by Tom Regan and Peter Singer, 60–86. Englewood Cliffs, NJ: Prentice Hall.

Donaldson, Sue, and Will Kymlicka. 2011. *Zoopolis: A Political Theory of Animal Rights*. Oxford: Oxford University Press.

Kant, Immanuel. 1997. *Lectures on Ethics*. Cambridge: Cambridge University Press.

Kant, Immanuel. 2003. *Groundwork for the Metaphysics of Morals*. Translated and edited by Thomas E. Hill Jr. and Arnulf Zweig. Oxford: Oxford University Press.

Korsgaard, Christine M. 2011. "Facing the Animal You See in the Mirror." *The Harvard Review of Philosophy* 16(1): 4–9. doi: 10.5840/harvardreview20091611

MacDonald Glenn, Linda. 2003. "Biotechnology at the Margins of Personhood: An Evolving Legal Paradigm." *Journal of Evolution and Technology* 13. http://jetpress.org/volume13/glenn.html

Regan, Tom. 1983. *The Case for Animal Rights*. Berkley: University of California Press.

Rowlands, Mark. 2012. *Can Animals Be Moral?* Oxford: Oxford University Press.

Wittgenstein, Ludwig. 1986. *Philosophical Investigations*. Oxford: Basil Blackwell Ltd.

Shue, Henry. 1980. *Basic Rights: Subsistence, Affluence, and U.S. Foreign Policy*. Princeton: Princeton University Press.

Singer, Peter. 1989. "All Animals are Equal." In *Animal Rights and Human Obligations*, edited by Tom Regan and Peter Singer, 148–162. Englewood Cliffs, NJ: Prentice-Hall.

Singer, Peter. 2006. "The Great Ape Debate." *Project Syndicate*, May 16. URL = <http://www.project-syndicate.org/commentary/the-great-ape-debate>

Laws

Animal Welfare Act. *Federal Law Gazette* I (p. 1094), Germany, 1998.

Animal Welfare Act. *Official Gazette of the Republic of Serbia*, no. 41/2009.

Animal Welfare Act 2006 (c. 45). *UK Public General Acts*, 2006.

Law on Venison and Hunting. *Official Gazette of the Republic of Serbia*, No. 18/2010.

Maria Clara Jaramillo
University of Cali, Colombia,

Rousiley C. M. Maia
The Federal University of Minas Gerais,

Simona Mameli
University of Bern,

Jürg Steiner
University of North Carolina, Chapell Hill

Original Scientific Paper
UDK 316.454.5:141.7
316.77:172

DELIBERATION ACROSS DEEP DIVISIONS. TRANSFORMATIVE MOMENTS[*]

Abstract: *In group discussions of any kind there tends to be an up and down in the level of deliberation. To capture this dynamic we coined the concept of Deliberative Transformative Moments (DTM). In deeply divided societies deliberation is particularly important in order to arrive at peace and stability, but deliberation is also very difficult to be attained. Therefore, we wanted to learn about the conditions that in group discussions across the deep divisions of such societies help deliberation. We organized such group discussions between ex-guerrillas and ex-paramilitaries in Colombia, Serbs and Bosnjaks in Srebrenica, and poor residents and local police officers in the favelas (slums) of Brazil. We could identify factors that help to transform discussions from low to high deliberation and risk transformations in the opposite direction. We could also identify factors that help to keep a discussion at a high level of deliberation, and, in a next step, we could determine to what extent long sequences of deliberation had a positive impact on the outcomes of the discussions. Finally, we show how our research results can have a long term effect if it is used in schools of such deeply divided societies.*

Keywords: *deliberation; speech-acts; war; Political Philosophy*

Our research aims to make deliberation relevant for the political practice.[1] The basic assumption is that from the local level to international politics, we need generally more deliberation to increase mutual understanding and trust and to arrive at political decisions of high epistemic value and legitimacy. This does not mean, however, that in our view a political system should consist only

[*] An earlier version of this paper was presented at the World Congress of the International Political Science Association, Poznań, July 23–28, 2016.

[1] For a fuller description of our research see Steiner et al. 2016.

of deliberation; we also need competitive elections, bargaining, administrative rulings, street demonstrations, and so on. If we want to learn how we can develop more deliberative behavior, we should investigate the group dynamic that helps to raise the level of deliberation and helps to prevent that its level drops again. To study these ups and downs of deliberation, we have coined the concept of *Deliberative Transformative Moments (DTM).* To have more deliberation is particularly important for countries with deep societal divisions; but these are precisely the countries, where deliberation is most difficult to be established. In our view, it is worthwhile to make an effort in this direction, since more deliberation may be the best hope to have more peace in these countries. They are critical cases for the deliberative enterprise. We will present data of group discussions of ex-guerillas and ex-paramilitaries in Colombia, of Serbs and Bosnjaks in Bosnia-Herzegovina, and of poor community residents and the police in Brazilian favelas (slums). These are critical cases in the sense that if some upward Deliberative Transformative Moments can be found, it seems reasonable to assume that they can also be found under more favorable conditions. So the crucial question for our research is whether some significant level of deliberation is possible under the unfavorable conditions of deeply divided societies.

The link of our research with the practice will be that tapes and transcripts of our group discussions will be included in the schoolbooks of these deeply divided countries and perhaps also in other deeply divided countries. Such schoolbooks will allow children at an early age to learn how deliberation works. Thus, we make a link between the deliberative literature and pedagogy. In this way, our research should have a long-term effect. Like anything else, deliberation can be learned in schools. One needs, however, the appropriate pedagogical instruments, and we hope that our research results will offer such help for schools. We can offer, for example, group discussions of Colombian ex-combatants, where common ground could be found of how to make progress in the peace process. We can also offer examples where deliberation broke down. School children can learn both from successful and failed deliberation.

In a nutshell, deliberation means that all participants can freely express their views, that arguments are well justified, which can also be done with well-chosen personal stories or humor, that the meaning of the common good is debated, that arguments of others are respected, and that the force of the better argument prevails, although deliberation does not necessarily have to lead to consensus (cf. Steiner 2012; Štajner 2015). In the course of a particular discussion, the various deliberative elements may not always be present to the same extent, and they may even be totally absent. In some sequences, arguments may be justified better than in others. Respect for the arguments of others may vary over the course of a discussion. Debates about the common good may be more frequent in some parts of the discussion than in others. Openness for all actors to speak up freely may also vary as the discussion progresses. For some decisions, the force of the better argument prevails but not for others. Thus, we are confronted with high complexity of how deliberation evolves over the course of a discussion. To get a handle at this complexity we have developed the concept of Deliberative Transformative Moment (DTM).

Before we go deeper into the deliberative model, we should be aware that deliberative research, like all social science research, has a philosophical background that needs to be reflected. Such reflection is done in a lucid way in a book-length publication by Antonio Floridia (Floridia 2016). The philosophical background of deliberation theory contrasts most clearly with the one of rational choice theory. The two theories are based on different assumptions about human nature. For rational choice theory, both politicians and ordinary citizens are always individual utility maximizers. This assumption was formulated early on in a classical way by Paul Edward Johnson, when he writes that "people are rational – they have preferences and act purposively to bring about outcomes that are desirable to them" (Johnson 1990, 610). Deliberative theory, by contrast, assumes that occasionally ordinary citizens and politicians do not exclusively pursue their individual preferences but truly care for the wellbeing of others. To be sure, to do good for others can be an individual preference, whose fulfilment gives internal psychic rewards, which would fit the assumption of rational choice that people are individual utility maximizers. In this case, to feel good about oneself would be the desired utility. Deliberative theory does not deny that many such cases exist, but it assumes that in politics there is also some genuine altruism, where voters or politicians do something for others without thinking what is in for themselves. In a referendum, for example, voters may support an expensive program for refugees without calculating that this may possibly increase their taxes. They may simply follow their conscience (Steiner 1996). Philosophically, rational choice theory goes back to Thomas Hobbes, deliberative theory to Immanuel Kant. It is, of course, not possible to say whether rational choice theory or deliberative theory are closer to truth, they simply start from different assumptions about human nature. How these assumptions are seen depends on one's psychological framing, and this framing depends on how one is socialized (Barber 1984, 67ff.). Thus, rational choice theory is particularly prominent in the United States, where it fits the strongly individualistic and competitive culture.

Deliberative theory got a push, when rational choice theory became prominent in the 1980's in the academic world. One of us set a counterpoint to the basic assumption of rational choice theory, using as example the playwright and dissident Václav Havel (Steiner 1991, 46–50). The Communist regime would have allowed him to go into exile, but Havel wanted to stay in the country to continue to raise his voice as dissident. For many months, he was sent to labor camp, where he almost died. After the fall of Communism, he became the first president of Czechoslovakia. From a rational choice perspective, one could argue that this heroic behavior was a long term investment for a political career, if one day Communism would crumble. Even rational choice theorists would not go so far, but they would argue that Havel was a rare exception that one could neglect in their theories. We argue, however, that Havel may be a rare exception but that he was exemplary. We do not deny that self-interest of both politicians and ordinary citizens is always very important. But as Nelson Mandela wrote, "man's goodness is a flame that can be hidden but never extinguished" (Mandela 1995, 622). This he said after having been treated cruelly by the white guards in prison. One can also think of the famous death toll poem of John Donne (1572–1631): "Any man's death diminishes me because I am involved in mankind, and therefore never send to know for whom the death toll; it tolls for thee" (in Dickson 2006). This poem speaks against the strong individualism of rational choice and goes to the roots of deliberation that one should care for others. Of course, politics cannot always follow the advice of John Donne, but deliberative theory assumes that some of us some of the time act in the spirit of John Donne. One should not deny that rational choice colleagues have an interesting research agenda, too. Deliberate theory and rational choice theory are just two ways to look at the world. This is fine, as long as neither side claims that their assumptions are objectively true. It is interesting and fruitful to look at politics through the lenses of deliberate theory, but one should accept that it is equally valid to use the lenses of rational choice theory. It is necessary, however, to reflect on the philosophical assumptions of one' work.

Having presented the philosophical background of our research, we now show how we proceed to identify Deliberative Transformative Moments (DTM). At an abstract level, we define them as changes from a low level of deliberation to a high level or vice-versa. To identify such situations, we use an interpretative-qualitative approach that has much to do with linguistics, social psychology, and rhetoric. We chose as our units of analysis the individual speech acts. Whenever an actor made any kind of utterance, this counted as a speech act, however brief or long the utterance was. So a speech act has a clear beginning and a clear end. When an actor makes another intervention later in the discussion, this counts as another speech act. To get an empirical handle at the concept of Deliberative Transformative Moment (DTM), we see deliberation as a continuum from no deliberation to full deliberation. On this continuum, we establish a cut-off point between high and low levels of deliberation, with the latter including no deliberation at all. The basic criterion is that at a high level of deliberation the discussion *flows* in the sense that the actors listen to each other in a respectful way, while at a low level of deliberation the discussion *does not flow* in the sense that actors do not listen to each other or do so only without respect. To determine whether a discussion is transformed from a low to a high level of deliberation or vice-versa, we use the following four coding categories for each speech act. In the 2014 issue of the Belgrade Philosophical Annual, we have presented an earlier version of the four coding categories (cf. Steiner, Jaramillo, and Mameli 2014, 39–48). Comparing this earlier version with what we present below allows the reader to see how these categories are a moving target that always still need improvement.

1. The speech act stays at a high level of deliberation

This first category is used if the preceding speech act was at a high level of deliberation and the current speech act continues at this level. The coding of the current speech act is least problematic if it fulfils all the criteria of good deliberation (cf. Steiner 2012), which means that the speaker has not unduly interrupted other speakers, justifies arguments in a rational way or with relevant stories or humor, refers to the common good, respects the arguments of others and is willing to yield to the force of the better argument. Deliberation can still remain at a high level, if speakers do not fulfil all these criteria, as long as they stay in an interactive way on topic. If a speaker, for example, supports the argument of a previous speaker without adding anything new, the discussion continues to *flow* at a high level of deliberation. Deliberation should be seen as a cooperative effort, which means, for example, that deliberative burden can be shared with some actors procuring new information, while other actors formulate new proposals, etc. The crucial aspect is that a group takes a common perspective on a topic, by which we mean a subject matter that has a certain internal consistency.

2. The speech act transforms the level of deliberation from high to low

This second category is used if the preceding speech act was at a high level of deliberation, and the current speech act transforms the discussion to a low level of deliberation. The *flow* of the discussion is *disrupted*. The topic debated so far is no longer pursued, and in the case of the Colombian ex-combatants no new topic related to the peace process is put on the agenda. Topics are mentioned that have nothing to do with the peace process and are therefore off topic. It is also possible that the speech act is so incoherent and confusing that it does not make sense. Under these circumstances, it is not easy for the other participants to continue the discussion in a meaningful way.

3. The speech act stays at a low level of deliberation

This third category is used if the preceding speech act was at a low level of deliberation and the current speech act stays at this level. Participants do not manage to give to the discussion again a direction. In the case of the Colombian ex-combatants, for example, this would mean that the speaker is unable or unwilling to put on the agenda a topic relevant for the peace process. Instead, the speaker brings up topics or stories that are off topic, or the speech act is incoherent and confusing. The key criterion for this third category is that the speech does not open new windows for the group to talk about the peace process.

4. The speech act transforms the level of deliberation from low to high

This fourth category is used if the preceding speech act was at a low level of deliberation and the current speech act transforms the discussion to a high level. Participants are successful in adding new aspects to a topic already discussed or to formulate a new topic, in the case of the Colombian ex-combatants relevant for the peace process. Success means that good arguments are presented why an old topic should be further discussed or why a new topic should be put on the agenda. In this way, the speech act opens new space for the discussion to continue in a meaningful way.

How do we proceed to apply these four coding categories to the data that we have collected? The group discussions of the Colombian ex-combatants as well as the poor community residents and police officers in Brazil were audio-recorded; in both countries for security reasons participants refused to be video-recorded. For the group discussions in Srebrenica, it was possible to use both audio– and video-recordings. As a first step in the analysis, the recordings were transcribed into Spanish for Colombia, into Bosnian for Srebrenica, into Portuguese for Brazil; then the transcripts were translated into English.[2] This was done by Maria Clara Jaramillo for Colombia, by Simona Mameli for Srebrenica, and by Rousiley Maia and her collaborators – Danila Cal, Rafael Sampaio, and Renato Francisquini – for Brazil. The translators had already acted as moderators of their respective groups, so that they were familiar with the atmosphere, in which the group discussions took place. The coding was a collective effort of the four authors, whereby Jürg Steiner had to rely on the English translations. We have looked in common at each speech act to arrive at a judgment about which of these four categories best applies to the respective speech act.

Maria Clara Jaramillo and Jürg Steiner did a reliability test choosing group 1 of the Colombian ex-combatants with altogether 107 speech acts; they agreed in 98 of these cases, which is a high rate of agreement. This does not mean, however, that we claim an *objective* nature of our coding. But the high rate of agreement is still comforting, especially because we come from very different backgrounds, Jaramillo from Colombia, Steiner from Switzerland. More important, our coding is fully transparent and therefore open for replications. The website of the Institute of Political Science (University of Bern) contains the recordings, the transcripts in the original language, and the English translations with the coding of the individual speech acts and the justification of the codes (Institute of Political Science 2016). The readers are invited to follow on this website how we interpret the dynamic that goes on in a particular discussion, and it may very well be that some readers take a different view, which would be in the deliberative spirit of how we look at our research.

We still have to justify why we focus on discussions among ordinary citizens and not among political leaders. In deeply divided countries, leaders tend to base their power on their respective group identities. Therefore, they have a vested interest that the deep divisions are kept up. In Srebrenica, for example, ordinary Serbs and Bosnjaks that we assembled for our group discussions complained that their political leaders want to keep them divided, so that their power remains intact. Emina from the Bosnjak side, for example, made the following statement: “The government just separates people; it frightens one side against the other, it says that we do not need to live together, so that they can rule us.”[3] Ordinary citizens, by contrast, are generally less constrained by such power considerations

[2] The English translations kept as close as possible to the original text to give a feeling of how participants actually expressed themselves.

[3] Cf. website for group 2 in Srebrenica (Institute of Political Science 2016).

and should therefore be more open to deliberate across deep divisions. So we had good reasons to focus on ordinary citizens. Results of their discussions, however, must reach the political leaders, and it must be made sure that these leader take seriously the results of citizens' groups. Political authorities who have the legal power to make decisions do not necessarily have to follow recommendations of citizens' groups, but they have to give good reasons if they take a different path.

We also still have to justify the choice of our three countries. Our initial choice was Colombia as an almost ideal case for a war torn country making efforts to end the war. Colombia has a long history of political violence, as presented in graphic form by Nobel Prize winner Gabriel Garcia Marquez with his novel *One Hundred Years of Solitude*. When we began our research, a program of decommissioning was under way, which gave us the chance to assemble guerrillas and paramilitaries, who a short while ago were still shooting at each other. Thus, we could submit deliberation to a particularly hard test. Would these ex-combatants be willing to meet at all, and if they did, would they ever be able to raise the discussion to a high level of deliberation? Having chosen Colombia, we looked for a similar case and found it with Bosnia-Herzegovina, another war torn country on its way out of civil war. Here, too, we had a case with a long history of political violence, as presented by another Nobel Prize winner, Ivo Andrić with his colourful novel *The Bridge over the Drina*. Within Bosnia-Herzegovina, we chose the town of Srebrenica for our research, a particularly hard case for deliberation, since it was here that in 1995 Serbs massacred a large number of Bosnjaks, men and boys. With these dreadful memories, would there be any amount of deliberation between Bosnjaks and Serbs? In Brazil, there are favelas (slums) with often war like situations, mostly linked to drug trafficking. The Brazilian police tends to exercise brutal violence with many mortal fatalities not only among the slum residents but also among the police officers. Here was another hard test for deliberation. Would the police sit together with poor slum dwellers and engage in some deliberative dialogue? We attempted to answer the question with discussion groups of local police officers and poor slum dwellers in Belo Horizonte and Belém. The three countries are similar in having deep divisions involving heavy violence. But there are also differences: In Colombia the division was based on poverty, ideology and drugs, in Srebrenica on poverty and ethnicity, in Brazil on poverty and drugs. In Srebrenica the civil war had ended, while in Colombia and Brazil violence continued at a high level. So we take a most similar approach with, however, some important differences.

Let us now show how we did the research in the three countries. We begin with *Colombia*, which is a particularly deeply divided society, in particular between leftist guerrillas and rightist paramilitaries. When we began our research, the Colombian government had a program of decommissioning under way. This program applied to combatants of both left guerrillas (in particular FARC, Fuerzas Armadas Revolucionarias de Colombia and some smaller guerrilla groups) and the paramilitary forces at the extreme right. Would ex-combatants be willing to sit around the same table? This was the challenge of our research, and it took patience to organize 28 discussion group with altogether 342 participants. The research took place in 2008. The work in the field was done by Maria Clara Jaramillo and Juan Ugarriza. In order to get a financial stipend, the ex-combatants were required to participate in a program of the Office of the High Commissioner for Reintegration. Social workers acted as tutors, and ex-combatants had to attend twice a month small-group sessions with these tutors, who helped us with a solution that gave to the ex-combatants the necessary incentives to attend our discussion groups. They could replace the bi-monthly tutorial sessions with participation in a single event and still get the full stipend. Of importance for the interpretation of what was said in the discussion groups is that politically there were strong differences between the two groups. The ex-guerrillas come much more often from a leftist family background, the ex-paramilitary from a rightist background. The clearest indicator for the deep divisions between the two groups comes to light in response to the question about their attitudes towards the combatants still fighting in the jungles. Although the participants in the discussion groups had left their former comrades, they expressed a more positive attitude towards their own side than to the other side. At the beginning of the discussions, the moderators stated the following topic: *"What are your recommendations so that Colombia can have a future of peace, where people from the political left and the political right, guerrillas and paramilitaries, can live peacefully together."* Moderators did not intervene to encourage deliberative behaviour. It was precisely our research interest to see to what extent ex-combatants were willing and able to behave in a deliberative way without any outside help. Thus, moderators let the discussion go wherever it went.

Bosnia–Herzegovina, with its recent internal armed conflict, was also a difficult place to do our research. We did it in Srebrenica, where the civil war was particularly ferocious. The research design was basically the same as for Colombia. In 2010, Simona Mameli organized six discussion groups with altogether 40 participants. For three groups, she selected the participants with a method called *random walk*. This means that she walked the streets of Srebrenica and approached people in a random way asking them to participate in our discussion groups. With random walk to select participants, we encountered two difficulties. One was related to the living pattern of the Bosnjak population. It forms the numerical majority in Srebrenica, but many Bosnjaks are only formally registered in the town and prefer to spend most of their time somewhere else. It seems that many of them come back only for elections or commemorative events for the genocide. It appears that more moderate Bosnjaks tend to live permanently in Srebrenica. This means that we likely got more moderate Bosnjaks in our sample. A second difficulty in searching for participants through a random walk was that some, both Serbs and Bosnjaks, were not willing to participate or, when they did promise to attend, did not show up. For the other three discussion groups in Srebrenica, we wanted participants, who had been exposed to a program of reconciliation and peace building, so that we could examine whether participation in such a program made a difference in the behavior in the discussion groups. The Nansen Dialogue Center, a Norwegian NGO, has such a program; its main objective "is to contribute to reconciliation and peace building through interethnic dialogue" (Nansen Dialogue Network 2011–2015).

The staff of the center helped us to recruit people, who had participated in its activities, making the selection as randomly as possible. The organization of the discussion was basically the same as in Colombia. Here the task for the group was to *"formulate recommendations for a better future in Bosnia–Herzegovina."*

In *Brazil*, poor residents in the favelas (slums) have a contentious relationship with the police. The police actions are characterized by human rights violation and abusive force, particularly against minority populations. Although the country has been re-democratized, the legacy of the military dictatorship (1964–1984) created an authoritarian culture in the police force. The growing power of criminal organizations and drug trafficking led to an escalating violence in the slums. The rhetoric of the "war on crime" leads to the view of the police as an army in face of an enemy to be destroyed. This rhetoric of war has the result of criminalizing residents of poor communities, who are seen as marginal and dangerous, as a threat to society. Although the majority of police officers come from lower classes in Brazil, they adhere to discriminatory cultural schemas and develop aggressive conduct to poor and non-white persons. This is the context in which we organized in 2014 six discussion groups, three in Belo Horizonte and three in Belém. Participants were poor community residents and local police officers, altogether 76 persons. To identify the participants, we had the help of the staff of two social projects, *Rede Escola Cidadã*, in Belém, and *Fica Vivo*, in Belo Horizonte; we also got the help of the police from the two cities. The research in the field was directed by Rousiley Maia; moderators were Márcia Cruz in Belo Horizonte and Danila Cal in Belém. The organization of the discussions followed the same guidelines as in Colombia and Srebrenica. The question to be discussed was: *"How is it possible to create a culture of peace between poor community residents and the local police?"*

We now give for each country two examples of Deliberative Transformative Moments (DTM), one upward, the other downward. We begin with the discussions of the ex-combatants in Colombia. With the following personal story ex-paramilitary Ernesto helped to transform the discussion back to a high level of deliberation.

> That is one of the things I used to say when I was young, I said, well, if I am Colombian, I am able to go everywhere I want to. Later, when I started to live with the conflict, I realized that there were places where people would tell you "go away from here, we don't know you". You knew that you were in danger. When I came to Bogotá, I was with a cousin and a friend of mine in one of the northern and wealthy neighborhoods, we were kind of lost. Then the police came, at first they asked us what we were doing; as my friend couldn't respond, at the end the police said they didn't want to see us around anymore, because neighbors had called to let them know that there were some strange and suspicious people, and they didn't want you here. What I feel is what you said about stratification, it is more than levels one, two or three of a scale; it is discrimination, that is the hard thing.

This story is relevant for a discussion among ex-combatants about the peace process in Colombia. Ernesto begins the story with his optimistic expectation that when he was young he could go anywhere in the country. He felt that as a Colombian he was not discriminated. Ernesto then continues that later in life in the context of the civil war he had to learn that unfortunately discrimination existed in Colombia and that he encountered this at a very personal level. He illustrates this claim with a story about a bad experience that he had in a wealthy neighborhood in Bogotá. Because he, his cousin and his friend looked suspicious, wealthy neighbors called the police to chase them away. Ernesto characterizes this episode as putting them in danger, because they were anxious not knowing what the police would do with them. This story is relevant for the peace process, because Ernesto can show to the other participants that there are huge social and economic inequalities in Colombian society. More specifically, he can show how ex-combatants in particular suffer under these inequalities. Through his story, Ernesto tells the other participants that these inequalities are not just a legal concept with abstract levels of one, two and three, but something that is revealed in everyday life as real discrimination. Ernesto does not explicitly link such discrimination to the ongoing civil war, but he tells his story in such vivid terms that it is implicitly clear that such inequalities are a major obstacle on the way to peace. Discrimination of ex-combatants is particularly damaging for the peace process, because their successful reintegration into society is a key pillar of the governmental peace plan of decommisioning and reintegration. If ex-combatants are dissatisfied with their situation, they may go back to fight in the jungle, as many have already done so. All this shows that the story of Ernesto touched an important nerve in the peace process. His story helps to make the argument that discrimination of the ex-combatants and more generally of the large masses of poor people has to be overcome if there is any chance for peace. The story helped the group to take a perspective on their common discrimination as ex-combatants, irrespective whether they come from the side of the guerrillas or the side of the paramilitaries. In this way, the story helped the group to develop a common life world in the sense of Habermas (Habermas 1981, 159). Laura W. Black also sees great potential in storytelling to enhance deliberation; for her "stories encourage listeners to understand the perspective of the storyteller. In this way, storytelling can provide group members with an opportunity to experience presence, openness, and a relational tension between self and other" (Black 2008, 109).

As we see in the following example, stories can also have a negative influence on deliberation. In this case, ex-guerrilla Hernando complains about the demobilization program and then tells his story.

> I've been demobilized for almost three years. The military card. What happened? From there I even was in jail in Picaleña for some crimes I had committed over there.

To this jail story, Beatrix, another ex-guerrilla, reacts with the following question:

> You mean you have not yet been cleared?

To this question Hernando answers as follows:

> Well, right now, it took me around life imprisonment, and I don't know what. I have to go to. Until you are not. They are not going to find a solution for us.

Hernando begins his story in a way that could have been of interest to the other participants. He informs them that he is demobilized for almost three years, which is longer than for most ex-combatants. So the group would have been interested to learn from Hernando how things stand after such a long time of demobilization. He mentions that he got a military card, which means that he was enrolled in the regular Colombian military. This was not an exceptional situation for ex-combatants; thereby, one must know that many of them were forced to enroll with illegal means. Hernando does not say, how he joined the military and what his experience was in the regular armed forces. He continues his story in telling the group that he committed some crimes and was put to jail. Again, he withholds from the group what exactly happened, which crimes he committed and what was his experience in jail. Beatrix, also an ex-guerrilla, asks him in a respectful way whether he has not yet been cleared. The context of the question is that the Colombian government makes a distinction for ex-combatants between military actions and ordinary crimes. For ordinary crimes they are persecuted like everyone else. Thus, Beatrix wanted to know whether Hernando was cleared from ordinary crimes. He is taken aback by this question not knowing how to answer and rambling along. The group only learns that he has not to go for life in prison, but otherwise Hernando does not give any further information of what happened to him in the almost three years since his demobilization. When Hernando spoke up, the conversation did flow at a high level of deliberation. Why did his story not help to keep the conversation at this high level but transformed it down to a low level? Since Hernando had a long experience of being decommissioned, his story had the potential to tell the group much about the process of reintegration. The group could have learned from him how the government differentiates for ex-combatants between military actions and ordinary crimes. The group also could have learned whether joining the regular armed forces was a good option for ex-combatants to be reintegrated into society. Hernando did not give any useful information about these questions, neither on the process of reintegration in general. His story lacked specifics and was not related in any intelligible way to the peace process. The case of Hernando shows that Sharon R. Krause is correct when she warns that personal stories may also have a detrimental effect on the quality of deliberation and that one should "distinguish between deliberative and nondeliberative forms of expression" (Krause 2008, 61). The story of Hernando was clearly a nondeliberative form of expression, not adding anything substantial to the discussion on the peace process.

We now turn to the discussions of Serbs and Bosnjaks in Srebrenica and present here, too, two cases of Delibereative Transformative Moments (DTM), one upward, the other downward. Milena from the Serb side offers a good example of how a rational argument can help to transform the discussion back to a high level of deliberation. Before she spoke up, Svetlana, also from the Serb side, had expressed utter despair claming that political parties hand out jobs only among their supporters, and as protest she will not give her vote to any party. With such despair, she keeps the discussion at a low level of deliberation. Milena picks up the election issue with the following rational argument:

> If you don't vote for anyone, those votes will help the current authorities.

Milena is interactive and offers Svetlana an argument why abstention in elections is counterproductive because it helps the current authorities. This argument is based on good knowledge of how elections work, and Milena links in a rational way a cause with a conclusion, transforming the discussion back to a high level of deliberation in opening space to discuss of how to use elections in an effective way.

An example of how the discussion in Srebrenica was transformed from a high to a low level of deliberation stems from Sladjana, whom we have already met above. With the following statement, she expressed utter despair and hopelessness:

> Here I am, a single mother I'm not protected by any law. I thought of that. No law. I had a problem, I faced the first three to four years (as single mother), and with whomever I spoke they told me that there is no law.

Sladjana seems comfortable enough to talk about her problems as a single mother. She does not say what her problem is but expresses despair that single mothers are not protected by any law. Bosnjak Tarik enlarges the despair of Sladjana to a more general level:

> What do you think, madam, how I am protected? I am a male. But neither women nor men are protected by laws. Neither you nor me. So, there is no law. For those who survived, there is no law.

Tarik refers to the massacre in Srebrenica, when he talks about those who survived. He enlarges the point of Sladjana that not only women, but men, too, are not protected by any laws. So he reinforces the despair of Sladjana keeping the discussion at a low level of deliberation.

Finally, we turn to two examples from the discussions in the Brazilian favelas between police officers and local inhabitants. A good case of an upward Deliberative Transformative Moment (DTM) was launched by Carolina, who at the time was only a 14 year old high school student:

> The people in the community only have bad things to say about policing, which is rude, but they do not see the sacrifice the police makes every night, right? Oh, I think what is missing is for the community to communicate with the police. When they have their break, community members should come up and tell the police what they think, to communicate with them. Because I think that it is a lack of communication between them. Because if you have perfect communication, the people will become more relaxed about security.

As a teenager, Carolina shows great wisdom in making a proposal very much in a deliberative spirit. At first, she shows good will towards the police acknowledging their sacrifices. Then she identifies the reason for the lack of a culture of peace that the community does not make any effort to communicate with the police. Furthermore, Carolina makes a concrete proposal how the situation can be remedied in asking the members of the community to come up to the police officers when the latter have their regular work breaks and to tell them what they have in mind. She concludes that such communication would relax the relations between the police and the community. This is all very well argued; the problem is clearly stated, and a specific solution is proposed how the problem can be solved. To emphasize the importance of communication is a key element in the deliberative model, and it is amazing how well Carolina is able to express it in simple terms. As the next speaker, police officer Roberto agrees with Carolina that communication is key and applauds the "interaction as we do it now (in the discussion group)." So the discussion continues with Roberto at a high level of deliberation.

A case of a downward Deliberative Transformative Moment (DTM) in the Brazilian favelas was triggered by a sarcastic remark of 19-year-old inhabitant Larissa. She told two police officers that when she was almost robbed in a park she did not get any help of a by-standing police officer. The two police officers then attempted to explain to Larissa what could have happened. Perhaps the criticized police officer did not realize the situation or he was not on duty. They also told Larissa that she should have called on the police officer. In this exchange, Larissa got angrier and angrier and finally made the following sarcastic remark, transforming the discussion to a low level of deliberation:

> I think he was there for a walk rather than to do his job.

With this sarcastic remark, Larissa mocks the two police officers for what they try to explain to her and conveys contempt for the work of the police in general. The police officers now also got angry telling Larissa that according to the rules there must always be two police officers together, so that the criticized officer must live in the neighborhood and was off-duty. This whole exchange shows how sarcasm can derail a discussion. Larissa, instead of listening to the police officers according to the deliberative principle of reciprocity, lashed out with her sarcasm at the professional honor of the police.

Based on the anaysis of all our cases from the three countries, we arrived at our conclusions. Our baseline null hypotheses was that, given the deep divisions, the group discussions would mostly be at a low level of deliberation with minor fluctuations up and down. This null hypothesis is rejected. There were many cases where the group dynamics led the discussion from a low to a high level of deliberation and vice-versa. What mechanisms helped to transform a discussion from a low to a high level of deliberation? Our initial interest focused on the comparison between the effects of rational arguments and personal stories. We tried to throw light on the controversies in the deliberative literature on the role of these two mechanisms (cf. Steiner et al. 2016, chs. 1 and 2). We found that rational arguments and personal stories were about equally successful to transform discussions from a low to a high level of deliberation. When it came to transformations in the opposite direction, from a high to a low level of deliberation, the responsibility was much more often with personal stories than with rational arguments. There was indeed only a single case where a rational argument was presented with so much arrogance that the other participants were intimidated. We conclude from these findings that rational arguments keep the upper hand for their deliberative functions; they often help to transform a discussion to a higher level of deliberation and are hardly ever responsible, when a discussion drops to a lower level. Personal stories, by contrast, have about equally often a positive and a negative influence on the level of deliberation. Deliberation is most helped when an actor makes a rational argument and supports it with a relevant personal story.

Besides rational arguments and personal stories, we found other mechanisms that helped to transform discussions from a low to a high level of deliberation or vice-versa. Good chosen humor can have a positive effect on deliberation, but when it turns to sarcasm, the effect can be negative. A mute reaction to an offensive remark can help that the discussion quickly returns to a high level of deliberation. At the individual level, we found that there were actors who played the role of deliberative leaders or deliberative spoilers. For upward DTM's it is particularly noteworthy that self-criticism and respectful criticism can have positive effects on deliberation. For downward DTM's, it is not surprising that in these war-torn countries the expression of despair and hopelessness often functioned as a deliberation killer.

We are aware that linking such factors to Deliberative Transformative Moments, we cannot speak in a proper way of causality. To establish proper causality, an experimental research design would have been needed with different treatments for individual groups. In some groups, for example, moderators could have insisted that no personal stories are told, whereas in other groups, on the contrary, moderators could have encouraged the telling of personal stories. With such an experimental design, we could have established in a true causal way the effect of personal stories on Deliberative Transformative Moments. We decided against an experimental research design, because we wanted to have our groups as close as possible to real life with the moderators not intervening and letting the discussions go wherever they went. Ours is not an experimental but a qualitative-interpretative approach with all its advantages and disadvantages.

We were not only interested in the factors that lead to Deliberative Transformative Moments (DTM) but also in what happens after such moments. Here, our focus was on what happened when the discussion continued for a long stretch at a high level of deliberation. Under these ideal deliberative conditions, actors of both sides of the deep divide were indeed often able to reach some agreements by the force of the better argument, which gave more legitimacy to the outcome. According to the research design, the moderators did not put issues to a vote but let the discussion go freely wherever it went. There were also no cases where participants organized a vote on their own. Therefore, we define an agreement between the two sides, if there is open accord from participants of both sides and no open objection of either side. A good example of such an agreement stems from a discussion between police officers and local residents in the Brazilian favelas, where there was a sequence with 24 speech acts at a high level of deliberation. There were 12 participants, three police officers, Michel, Gustavo, Cynthia; six teenagers, Thiego, Cibele, Yago, Thaiane, Nathália, Eric; three adult residents, Isadora, Margarida, Milena. At the beginning of this sequence, 15-year-old Eric had transformed the discussion back to a high level of deliberation with the following statement:

> Teenager Eric: I think more communication, more conversation between the police and the community is needed, because sometimes it is just a lack of communication. For example, the police gets here and says something that is not true, it needs to have this conversation first: "How was it? How did it happen?" I think it is a lack of communication.

Police officer Gustavo agrees with "the young man here" that there must be "more dialogue." As a policy measure, Gustavo proposes that "the community has to invest more in cultural and educational projects." He justifies this investment that "there will be less crime, if the population is more educated." The discussion then turns to the question of discrimination by the police. Police officer Cynthia argues that for police work to be successful, profiling is necessary in the sense that not all community residents are approached in the same way. She justifies her argument in the following way:

> Police officer Cynthia: If you are walking down the street, and a boy is coming with a book, a backpack on his back, no matter his skin color, you do not check him; another boy is coming with dyed hair, tattoos, boxer shorts, walking in a strange way. Of whom of the two boys will you be afraid, who do you think will rob you?

Teenager Eric agrees with police officer Cynthia that the police should check only the second boy and adds: "I have no tattoos." This is a remarkable agreement across the divide between the community and the police. Community resident Margarida does not object to profiling but warns that "appearances can be deceiving because today bad guys dress better than good persons, so that the police does not think they are criminals."

Police officers Gustavo then gives advice to the community residents, especially to the teenagers.

> Can I give a tip for you guys? Walk with your documents. Let us say the truth, there are still a lot of aggressive police officers. Then you say, look if you want my identity it is right here in my pocket. The guy who is malicious does not want to be identified.

Gustavo then thanks our university research group, "because the community residents can already see some of the difficulties that we have in the police, and we know that everything can be improved." Police officer Cynthia adds that "it is a matter of both sides understanding that there are mutual difficulties but that we want to help each other. The community has difficulties with social issues, and the police has difficulties, too. When both sides understand this, things will change."

As a concrete measure, both sides agree that reporting of crimes should be improved. Community resident Margarida offers that "the community can help the police in cases of theft; sometimes our mobile phone gets stolen, and we think it is a silly thing and do not register it with the police." Police officer

Gustavo thanks Margarida for this offer and repeats "no more impunity, we need to report."

In this sequence of the discussion police officers and community residents address in a straight forward way their troubled relations and find ways of agreements how these relations can be improved. We found similar agreements between ex-guerrillas and ex-paramilitaries in Colombia and between Serbs and Bosnjaks in Srebrenica. In this paper, we do not have space to present examples from these two countries, for which we refer to our book-length publication (Steiner at al. 2016, ch. 8). The general important conclusion is that long stretches of high deliberation often leads to approaches and even agreements across deep divisions. In this way, deliberation can contribute to increased democratic legitimacy of political outcomes.

Our research has focused on the very micro-level of deliberation in studying the internal dynamic of group discussions. We are aware, however, that an important topic in the current deliberative literature is the analysis of deliberation at the system level (Parkinson and Mansbridge 2012). In this literature, the focus was up to now to analyze in a synchronic manner the various discourses in the system and how they are connected. The concept of Deliberative Transformative Moments can also be applied at the system level and will help to give to the analyses a longitudinal dimension. Let us take the United States, where it seems that since the late 1980's the level of deliberation has strongly dropped (cf. Muirhead 2014). The times have past, when President Ronald Reagan and Speaker Tip O'Neill had drinks together after work. There is no longer such a common life world between Republicans and Democrats in Congress, no longer any real deliberation. It is remarkable that this change cannot be explained institutionally because the institutions remained unchanged over this period of time. One could try to establish at the systemic level to what extent the level of deliberation actually dropped and what possible causes and consequences could be. Generally speaking, it will be fascinating to use the concept of Deliberative Transformative Moment (DTM) to connect deliberative theories at the micro– and macro-level. Possibly, there is a grand theory in waiting about the dynamic development of deliberation at all levels of society.

Our research should be relevant for the practice of deliberation. To be sure, some participants in our group discussions may have become more deliberative as a result of participating in these events. But this is not enough. There must be wider implications of our research. Of prime importance is that school children learn to deliberate. John Dewey is the classic author who has inspired much scholarship in this respect (Dewey 1902). More recently, in a general book about political education in schools, Eamonn Callan has stressed that moral dialogue in schools would seem necessary if we are to cultivate the respect for reasonable differences. ... Moral education requires ongoing dialogue with children as they grow up, and the requirement holds in schools and not just in families (Callan 1997, sec. 56).

From a philosophical perspective, Tomas Englund argues in the very title of his paper that schools can be "sites of deliberation" (Englund 2011). He begins in a creative way telling the story of pianist and conductor Daniel Barenboim who for many years brought together in the West-Eastern Divan Orchestra young talented musicians from both sides of the conflict between Israel and Palestine for musical events and political dialogue (Ibid., 236). According to Englund, such dialogue across deep divisions should also be possible in schools, "namely as spaces for encounters between students from different environments exercising common interests, political dialogue and fraternization" (Ibid., 237). He postulates that the universal human right for education should mean "every child's right not just to learn basics, but also to come into contact with different and conflicting world views" (Ibid.). Englund wants an interactive universalism in which schools constitute an arena for encounters between different social, cultural, ethnic and religious groups that attaches importance to developing an ability and willingness to reason on the basis of the views of others and to change perspectives (Ibid., 244–245).

This focus of Tomas Englund for schools to overcome deep divisions fits exactly what we have in mind as practical conclusion of our research. We want students to be exposed to authentic material of our research about deliberation across deep divisions. The website to our research contains the recordings of the discussions and the transcripts both in the original language and in English translation (cf. Institute of Political Science 2016). The prime task will be to make future and current teachers familiar with the deliberative model. The research material will help in this task. In listening to the recordings and reading the transcripts of our group discussions teachers get an understanding what it means to deliberate. They learn what factors help and what factors hurt deliberation. Teachers will then have to be taught of how our research material can be used as a teaching tool. We suggest that schoolchildren should learn to deliberate in critically evaluating what went on in our discussion groups. Children in Bosnia-Herzegovina, for example, should listen to the recordings of the discussions of Serbs and Bosnjaks in Srebrenica and evaluate what reduced the division between the two ethnic groups and what increased the division. Was is helpful, for example, that both sides agreed on the construction of a shelter for stray dogs? Were the arguments well presented? Were participants listening to each other with respect? Was the wellbeing of the entire city considered or was each ethnic group only looking for their own interest?

To be successful, teachers have to use the right pedagogy to bring our research material into the class room. It would be in a deliberative spirit if teachers would somewhat stand back and let the students analyze for themselves the research material. This should preferably be done in small groups, where all the students can get actively involved. In this way, students learn not only about deliberation in our discussion groups but get themselves a hand-on experience in deliberation. A good pedagogical devise would be if the small groups would then report their results to the entire class, where a discussion in a larger circle can take place. Here, students learn to speak up to a larger audience, a necessary skill for their later role as citizens. In all such activities, teachers have a delicate and important role. Without intervening too much in the discussions of the students, they still should give some deliberative guidance. When a student gives an argument without justification, the teacher should ask the student why he or she makes such an argument. Such teaching is very challenging and needs a lot of training and preparation before. So it is key that teachers become very competent in the field of deliberation.

To overcome deep divisions in countries like Colombia, Brazil, and Bosnia-Herzegovina is a long term project. Short term measures are not likely to be successful. This was shown in an investigation of Juan E. Ugarriza and Enzo Nussio in Colombia. They brought together discussion groups of ex-combatants and residents of communities that were particularly struck by the civil war. With a randomized controlled experimental design, they investigated whether encouragement to follow deliberative standards had an effect on the discourse quality in the ensuing discussions. While some groups got such encouragements, others did not. Comparing the two sets of groups did not reveal any significant differences in the discourse quality. Ugarriza and Nussio conclude: "A core finding from our experimental design is how short-term efforts aimed at providing people with a basic understanding of deliberative standards, while also encouraging them to act accordingly, cannot overcome the structural limitations deriving from low levels of education within marginalized communities" (Ugarrizza and Nussio 2016, 160). Using a medical metaphor, Ugarriza and Nussio caution in the very title of their paper: "There is No Pill for Deliberation."

This conclusion fits well with what we propose ourselves. Teaching the skills of deliberation must be a long term process beginning already at an early age in schools. Having understood deliberative lessons, children may also influence their parents leading to a snowball effect up the generations. It would also be helpful if the media, in particular social media, report about such new teaching experiences. When schoolchildren become later citizens, they should have learned to respect people with whom they differ with regard to ideology, ethnicity, race, religion, social class and other such aspects. We hope that in this way a culture of peace and tolerance will develop. Our practical argument is that deliberation is a skill that can be learned like any other skill. It would be gratifying for our research team if our research material could help in this learning process of deliberation. We are aware, however, that even when students have learned to deliberate in schools, these deeply divided countries may have so much power inequalities that effective deliberation in political practice may be difficult. But perhaps efforts to engage in deliberation by young people may help to reduce existing power inequalities.

To investigate whether our hope that schoolchildren can learn to deliberate based on our research material is our next research project. The prime focus are the three countries where we have done our current research, Brazil, Bosnia-Herzegovina, and Colombia. Other countries can be included, but there teachers and students have to rely on translations of the transcripts to their respective languages. The important element in this further research will be that there are control groups of students that do not get the "treatment" of deliberation. Only working with such control groups can we establish, whether teaching deliberation has a positive effect. So see whether such an effect exists, students in both the control groups and the experimental groups have to fill out questionnaires before and after the latter groups get the deliberative treatment. To measure the attitudes towards deliberation, the following items are used (the response categories are for all four items: strongly agree, agree, disagree, strongly disagree, don't know).

> Item 1: When we have disagreements with other people, we should fight as much as possible for our own position.
>
> Item 2: When we have disagreements with other people, we should try to find a solution acceptable to everyone.
>
> Item 3: In politics, all are fighting for their personal interests.
>
> Item 4: We should not give up hope that we find political candidates who care for the common good of all of us.

If there are no changes in these items for the control groups, but there are changes in a deliberative direction for the experimental groups, the hypothesis is confirmed that the deliberative treatment had the expected effect. To check whether this effect is enduring, the items have to be answered again one month afterwards in both sets of groups. A further step in the research is when we check whether the impact of the deliberative treatment is not only on attitudes but also on behaviour. To do this further check, still a month later the two sets of groups participate in discussions across deep divisions, which may be based on ethnicity, race, religion, social class or any other identity creating attribute. The hypothesis would be that the students in the control groups would discuss in a less deliberative way than the students who got the deliberative treatment. Not only children but also adults can learn the skills of deliberation, and this also in deeply divided societies, as our research has shown. We hope that governmental agencies, nongovernmental organizations, universities, and other national and international agencies will join in our effort to set up discussion groups in deeply divided societies, bringing together people from across the deep divisions. Thereby, such discussion groups can also involve political leaders, either discussing among themselves or with ordinary citizens. Special attention must be paid to the role of the moderators. In the research reported here, our moderators only put the question to be discussed and then let the discussion freely go wherever it went. In our view, this is in a deliberative spirit in the sense that the moderation is taken over by the groups themselves with deliberative leaders emerging. There will also be deliberative spoilers, but the group has to decide itself how to handle them. With this approach, citizens are taken as "mündig" in the sense of Immanuel Kant (Kant 1784), a concept that is not quite captured but comes close to the term "mature". We acknowledge, however, that the level of deliberation may possibly be increased, if the moderator acts as facilitator, urging, for example, participants to give better justifications for their arguments or to be more respectful. The disadvantage of an active moderator is, however, that participants may feel like in school, discouraging them to speak up in an independent way, fearing to say the wrong thing. For each project, one has to weigh carefully the advantages and disadvantages of active moderators.

Finally, there is the problem of *scaling up* from discussion groups to the policy process. Such discussion groups will always be small in number, involving only a minimal part of the entire population. Although participating in such groups has a value in itself, the policy impact can only be attained if the results of their discussions reach the political decision makers. We have made an effort in this direction. In Srebrenica, for example, the groups of Serbs and Bosnjaks put together letters with policy recommendations that were hand delivered to the office of the High Representative. These letters may or may not have had any influence. To make sure that an influence exists, the relationship between the discussion groups and the political authorities must be more institutionalized. In our view, this is very successful done by the Italian region of the Toscana, where this consultation process has as basis a law decided by the regional parliament (Regione Toscana. Consiglio Regionale 1998–2016). The law determines the issues that are important enough to be submitted to the consultation process of discussion groups. Quite a large amount of money is allocated on a yearly basis for this purpose. The ultimate legal responsibility remains with the political authorities, but according to the law they must seriously consider the recommendations that come from the various discussion groups. Thereby, it is important that the law stipulates that the recommendations of the groups must be published, so that a public debate can ensue. How the Toscana Region has institutionalized the relationship between discussion groups of ordinary citizens and political authorities can serve as a good model elsewhere, including in deeply divided countries.

Deeply divided societies are most in need of deliberation but encounter also the greatest obstacles to deliberation. Our research has shown that these obstacles make deliberation difficult but not impossible. The challenge is to put this finding into political practice in the many countries and regions of the world with political violence resulting from deep societal divisions of many kinds.

References

Barber, Benjamin. 1984. *Strong Democracy. Participatory Politics for a New Age*. Berkeley: University of California Press.

Black, Laura W. 2008. "Deliberation, Storytelling, and Dialogic Moments." *Communication Theory* 18(1): 93–116. doi: 10.1111/j.1468–2885.2007.00315.x

Callan, Eamonn. 1997. *Creating Citizens. Political Education and Liberal Democracy*. Oxford: Clarendon.

Dewey, John. 1902. *The Child and the Curriculum*. Chicago: Chicago University Press.

Dickson, Donald, ed. 2006. *John Donne's Poetry*. New York: W.W. Norton.

Englund, Tomas. 2011. “Potential of Education for Creating Mutual Trust. Schools as Sites for Deliberation.” *Educational Philosophy and Theory* 43(3): 236–248. doi: 10.1111/j.1469–5812.2009.00594.x

Floridia, Antonio. 2016. *From Participation to Deliberation. A Critical Genealogy of Deliberative Democracy*. Colchester: ECPR Press.

Habermas, Jürgen. 1981. *Theorie des Kommunikativen Handelns*. Frankfurt a.M.: Suhrkamp.

Johnson, Paul Edward. 1990. “We Do Too Have Morals: On Rational Choice in the Classroom.” *PS: Political Science and Politics* 23(4): 610–613. doi: 10.2307/419906

Kant, Immanuel. 1784. “Was ist Aufklärung?” *Berlinische Monatsschrift* Dezember: 481–494.

Krause, Sharon R. 2008. *Civil Passions: Moral Sentiment and Democratic Deliberation*. Princeton: Princeton University Press.

Mandela, Nelson. 1995. *Long Walk to Freedom*. New York: Little, Brown and Company.

Muirhead, Russell. 2014. *The Promise of Party in a Polarized Age*. Cambridge: Harvard University Press.

Parkinson, John, and Jane Mansbridge, eds. 2012. *Deliberative Systems – Deliberative Democracy at the Large Scale*. Cambridge: Cambridge University Press.

Steiner, Jürg. 1991. “We Need More Politicians Like Havel and We Should Tell Our Students So.” *PS: Political Science and Politics* 24: 46–50.

Steiner, Jürg. 1996. *Conscience in Politics. An Empirical Investigation of Swiss Decision Cases*. New York and London: Garland Publishing.

Steiner, Jürg. 2012. *The Foundations of Deliberative Democracy. Empirical Research and Normative Implications.* Cambridge: Cambridge University Press.

Steiner, Jürg, Maria Clara Jaramillo, Simona Mameli. 2014. “The Dynamics of Deliberation.” *Belgrade Philosophical Annual* 27: 39–48.

Steiner, Jürg, Maria Clara Jaramillo, Rousiley Maia, and Simona Mameli. 2016. *Deliberation across Deeply Divided Societies. Transformative Moments*. Cambridge: Cambridge University Press.

Štajner, Jirg. 2015. *Osnovi deliberativne demokratije: empirijsko istraživanje i normativne implikacije*. Beograd: Službeni glasnik; Sarajevo: Fakultet političkih nauka.

Ugarriza, Juan E., and Enzo Nussio. 2016. “There is No Pill for Deliberation: Explaining Discourse Quality in Post-conflict Communities.” *Swiss Political Science Review* 22(1): 145–166. doi: 10.1111/spsr.12195

Online sources

Regione Toscana. Consiglio Regionale. 1998–2016. “Autorità per la partecipazione, Consiglio regionalle della Toscana.” *Regione Toscana. Consiglio Regionale.*

http://www.consiglio.regione.toscana.it/oi/default?idc=47&nome=PARTECIPAZIONE

Institute of Political Science. 2016. “Deliberation.” *Institute of Political Science*. www.ipw.unibe.ch/content/research/deliberation

Nansen Dialogue Network. 2011–2015. “Vision and Mission.” *Nansen Dialogue Network*.

http://www.nansen-dialogue.net/index.php/en/who-are-we/vision-and-mission

Ivan Matić
University of Belgrade

Original Scientific Paper
UDK 321.728:321.01
141.7

THE CONCEPT OF MIXED GOVERNMENT IN CLASSICAL AND EARLY MODERN REPUBLICANISM

Abstract: *This paper will present an analysis of the concept of mixed government in political philosophy, accentuating its role as the central connecting thread both between theories within classical and early modern republicanism and of the two eras within the republican tradition. The first part of the paper will offer a definition of mixed government, contrasting it with separation of powers and explaining its potential significance in contemporary political though. The second part will offer a comprehensive, broad analysis of the concept, based on political theories of four thinkers of paramount influence: Aristotle, Cicero, Machiavelli and Guicciardini.*[1] *In the final part, the theories and eras of republican tradition will be compared based on the previous analysis, establishing their essential similarities and differences.*

Key words: *Mixed Government, Classical Republicanism, Florentine Realism, Early Modern Republicanism, Aristotle, Cicero, Machiavelli, Guicciardini.*

Introduction

The roots of republican political thought can be traced to the antiquity and the theories of ancient Greek and Roman philosophers. As might be expected in the two and a half millennia that followed, our understanding of republicanism has been subject to frequent and, sometimes, substantial change. Yet, one defining element has always remained: the concept embedded in its very name, derived from the Latin phrase *res publica* (public good, or, more broadly, *commonwealth*), its implicit meaning being that the government of a state is meant to be accessible and accountable to all citizens, its goals being the goals not merely of certain classes and factions, but of society as a whole.

Today, republicanism is generally associated with the political systems of parliamentary democracies in the majority of European states, the United States of America and the members of the British Commonwealth. The philosophical

[1] I use the terms classical and early modern republicanism to distinguish between the two eras within the tradition; it is worth noting that the concept of republicanism doesn't appear in theoretical discourse until the Enlightenment. Thus, referring to the theories of these authors as republican is essentially the same as talking about Plato's aesthetics, the term having been coined by 18-th century philosopher Alexander Baumgarten.

foundation of this type of government can be traced back to the Enlightenment ideas that inspired the French Revolution, primarily those of Montesquieu (Montesquieu 1989, 156). In an attempt to combat absolutism, the system of monarchic government that was in place across the whole of Europe, with the exception of Great Britain, the Netherlands and the Polish-Lithuanian Commonwealth, and in order to counteract the possibility of its re-emergence, the concept of separation of powers was developed, comprehending that the legislative, the executive and the judicial are three specific, mutually independent branches of government.

The main purpose of this division was the prevention of executive overreach – its proper application would insure that the wielders of executive authority can neither craft laws to suit their interests, nor can they arbitrarily involve themselves in judicial procedures. States that have successfully applied the concept of separation of powers have distinguished themselves by political and economic stability, rule of law, individual freedom and minority rights. However, many of these states also face an increasingly obvious problem of unequal political representation stemming from economic inequality.

What is the cause of this inequality? Political theorist and prominent Machiavelli scholar John McCormick notes that, despite the formidability of the concept of separation of powers in combating absolutism and, more recently, totalitarianism, it is also characterized by a premise, that while rooted in the spirit of the Enlightenment and the ideas of human rights and natural equality, has no foundation in political reality: namely, the idea that a majority-elected government, owing its authority to the will of the people, will strive to represent the interests of all its citizens equally (McCormick 2011, 1).

McCormick claims that the majority of contemporary democratic theorists and policy analysts cannot answer a question that was central to pre-eighteenth-century republicanism: "What institutions will prevent wealthy citizens from dominating a government that is supposed to serve the entire citizenry?" Seeing as this was a presupposed consequence in the ancient and early modern period unless suppressed institutionally, he bases his prescriptions for increasing the level of political equality in contemporary systems on the theories of early modern political thinkers from Renaissance Florence, the so-called Florentine Realists: Niccolo Machiavelli and Francesco Guicciardini. Despite the two millennia long historical "gap", these theories share many similarities with those of political philosophers from the epoch of classical republicanism, mainly Aristotle and Cicero, and greatly correspond to the political systems of both ancient city-states and Northern Italian late fifteenth and early sixteenth century republics.

The Concept of Mixed Government

The central moment that distinguishes classical and early modern republicanism from that in the Enlightenment and in our own era is the concept of mixed government. What does this concept comprehend and what makes it different from separation of powers? Quite similarly to our modern systems, it also entails a (frequently tripartite) separation of powers, the difference being that in this case, the separation is not based on governmental branching, but on economic strata and property assessments. The division that can, in one form, be argued to originate from Plato (Plato 2005, 239), is between monarchy, aristocracy and democracy as the "good" forms of government, with the "bad" ones, tyranny, oligarchy and anarchy being essentially their distortions.[2]

Monarchic government, despite the obvious implication, can, in practice, be simultaneously held by several people, usually from the wealthiest and most influential political families, with their positions often being hereditary. Furthermore, contrary to what the term may imply, these need not be kings or princes, but can also be the ancient and early modern equivalents of presidents, prime ministers and chancellors. Aristocratic government is commonly held by the nobility, usually within the so-called council of elders, or senate, the positions within which can be either hereditary or elective. The fulcrum of democratic government is the popular assembly or council of representatives, the members of which are elected from among the common people. The concept of mixed government comprehends that for a society to achieve the long-reaching goals of justice, freedom and stability, all three types of government must coexist in a system of checks and balances, the alternative being that a single mode, allowed to grow into its extreme, will inevitably deviate into its opposite.

Classical and early modern republican theorists founded this concept on the presupposition that, for lack of any of these forms of government, the social stratum that corresponds to it will necessarily fail to achieve political representation, which, in turn, will make the entire system unjust and unstable. For example, Sparta, which is historically remembered as a monarchy ruled by two kings, had a mixed government in Aristotle's view, with the council of elders and the council of representatives corresponding to the aristocratic and democratic element, respectively (Aristotle 1998, 52). Machiavelli held that Rome only became a republic after the establishment of the popular tribunes, following the ousting of Tarquinius Superbus, with all three forms of government having achieved representation: the consuls corresponded to the monarchic element, the senate – the aristocratic, and the tribunes – the democratic element (Machiavelli 1996, 14).

In addition to the need for all social strata to achieve representation through their respective forms of government in order for a society to be stable, the concept of mixed government simultaneously supervenes on the implicit virtues of the different forms of government while protecting from their respective vices. The virtue of monarchy is held to be in the paramount political knowledge, skill and virtue of the one or the few in authority; that of aristocracy, in the competence and experience of the ruling minority, required to handle the

[2] The terms used to describe these forms vary greatly from one theorist to another: in Aristotle's terminology, for example, kingship corresponds to monarchy, and polity to well-ordered democracy; Machiavelli names monarchy principality, aristocracy government of the great and democracy government of the people. However, in spite of the terminological distinctions, these names retain the same essential meaning throughout all theories.

complexities of challenging political issues, and the virtue of democracy is held to be in the freedom and equality for of the majority and in the legitimacy that stems from its consent.

As all of these forms of government are susceptible to corruption and deviation, their mixture is also meant to prevent their regression into their distortions, the primary distinction being that the "good" forms of government strive for the well-being of the entire society, while the corresponding "bad" forms serve the narrow interests of the individuals or groups in power (Aristotle 1998, 78). Thus, the only purpose of tyranny is the boundless accumulation of power and wealth by the tyrant; that of oligarchy is the exclusion of the poor from the government for the unimpeded realization of the interests of the rich, while that of anarchy is the full redistribution of wealth for the benefit of the poor and to the detriment of all others.

Good constitutions and forms of government are, therefore, also distinguished from bad ones by their laws, which, besides being clear and comprehensive, also need to be consistently followed, establishing the groundwork for a free, stable and just society. Actually creating such a society is an age-old problem to which the concept of mixed government should be regarded as a republican solution founded on realist principles: by building upon the lawful basis of a good constitution, it seeks to appropriate and combine virtues inherent in the three good forms of government, while counteracting the vices of their deviations, constantly ensuring that all social strata are represented, with partaking in the government being seen as the only meaningful way of realizing their political interests.

Aristotle, Cicero, Machiavelli, Guicciardini

As we present a basic overview and analysis of the theories of these four great thinkers, we'll primarily focus on each one's particular understanding of mixed government. The first step in this analysis will be to describe the social and political circumstances they wrote in, as well as the particular intention with which they approached their works. The second will be to explore their worldview in the descriptive sense, as the foundations upon which their political prescriptions are based. The third will be to focus on the substance of their theories, analyzing their views on how government should be constituted, what its role and limits are and how these differ across various circumstances, if there is an ideal form of government and, for our inability to create it, what its best alternatives are.

Aristotle was born a free man to a well-off family in Stagira, northern Greece. At a young age, he was sent to Plato's academy where he went on to become a teacher and researcher. He would eventually become the tutor and mentor of the famous conqueror, Alexander the Great (Reeve 1998, xvii). As a man who enjoyed considerable privilege, Aristotle challenged few social and cultural dogmas of his day, his position towards slaves and women perhaps being the most notorious one he held (Aristotle 1998, 8). He was nevertheless a man of immense practical wisdom who approached political theory from an entirely different angle from his teacher, Plato, whose system he criticizes before presenting his own. Rather than going into great detail about the ideal society and state, his main work on the subject, *Politics*, is filled with practical prescriptions for reforming existing constitutions, based on comparative historical analysis and greatly benefitting from his remarkable knowledge of history.

In the ancient world, Greece was not a unified state that we know today, but rather a large collection of independent city-states that stretched from the western coast of Asia Minor to Southern Italy and Sicily, with vast social differences and political rivalries that quite often resulted in large scale wars. These city-states were, nonetheless, connected by a long thread of common culture, religion and customs, and while there were as many forms of government as there were city-states, ranging from monarchies, across oligarchies, to democracies and vast numbers of mixed regimes in between, all of them were populated by a minority of free citizens with political rights generally dependent upon property assessments, while the women, slaves and foreigners had no institutionally recognized power whatsoever.

Rather uniquely, and in great contrast to the social contract theories of the Enlightenment, Aristotle saw the state as a natural phenomenon (Reeve 1998, xlviii), and humans as animals with a natural tendency toward forming society, with anyone that didn't need it by virtue of being self-sufficient being seen as "either a beast or a god" (Aristotle 1998, 5). Aristotle's brand of realism is centered around the acceptance of certain political facts to the exclusion of attempting their change: for example, that enslavement is no less legitimate than hunting or any other form of property acquisition (Ibid., 14), that "vulgar craftsmen and hired laborers" can never be truly equal to free men (Ibid., 74), and that virtue naturally escapes the common man, which is why the majority can attain it only through military service (Ibid., 77).

His division between good and bad forms of government is based on two primary criteria: their lawfulness, or lack thereof and the state's apparent (as opposed to stated) goals. The good constitutions, whether in the hands of kings (*kingships*), the rich (*aristocracies*), or the poor (*polities*), are characterized by just laws and a clear governmental focus on benefitting the entire society rather than the narrow interests of the ruling social strata (Ibid.). By contrast, their deviations are typically either lawless, or their laws are routinely ignored and the ruler(s), whether an individual (*tyranny*), a rich minority (*oligarchy*) or a poor majority (*democracy*) holds power, seek nothing but their own benefit to the constant detriment of society as a whole (Ibid., 78).

As the foundational elements in his political theory, Aristotle defines the city-state as a community of households and families whose end is a complete and self-sufficient life (Ibid., 81) and civic virtue as the capacity to rule and be ruled in turn (Ibid., 72). Briefly identifying a true, natural aristocracy as the ideal form of government (Reeve 1998, lxxii), with all of the extant aristocracies and other regimes being its deviations, Aristotle immediately recognizes the fact that most city-states are generally either oligarchies, democracies or a mixture between the two.

Elaborating upon his view of the three basic good and bad forms of government, Aristotle acknowledges that they have several different varieties each, based on the different methods of acquisition and transfer of power, property assessments, wages and taxes for assemblies, as well as other factors. In this context, and in regards to his briefly mentioned ideal constitution of natural aristocracy, he goes on to point out that "one should not study only what is best, but also what is possible" (Aristotle 1998, 102) and identifies tyranny as the worst form of government, being the furthest removed from a well-ordered constitution, oligarchy as the second worst, and democracy as the most moderate.

Aristotle opposes the general definition of oligarchy as the rule of the few and democracy as that of the many, as there have been examples, albeit a few, of city-states where the ruling majority was rich and the subject minority poor; he proposes a more narrow definition, which asserts that democracy is the rule of the free and oligarchy that of the rich (Ibid., 106). The need for mixed constitutions, therefore, naturally arises from the fact that most existing constitutions are of the bad forms, by and large being either oligarchic or democratic. Their potential remedies also represent something inherently more virtuous, the two opposing good forms being defined as such not only on the basis of their lawfulness and commitment to the common good, but also on mixed government, with polity being a mixture of oligarchy and democracy that leans toward the latter, while aristocracy tends to lean toward the former (Ibid., 115).

Seeing as the institutional enactment of this mixture presents a peculiar problem, Aristotle proposes a solution that is based on combining the different policies of oligarchies and democracies, all of which are purposed toward facilitating the level of the ruling strata's participation in government. Thus, oligarchies fine the rich for failing to appear in court cases, while democracies pay the poor to appear; oligarchies require a large property assessment for membership in the assembly, while democracies require none; the former conduct the election of public officials by vote, the latter by lot (Ibid.). Mixed constitutions, whether referring to aristocracies, polities, or something in between, should, therefore combine all of these elements, thereby creating a middle ground between existing legislative practices, that results in a greater balance of political power, preventing any social strata's representation from attaining too much power and transforming the constitution into its opposite, deviant form.

Furthermore, Aristotle places great emphasis on moderation and equality: a city-state in which the middle class is the most numerous will tend to have the greatest longevity, seeing as they neither desire the property of others, nor is their property desired (Ibid., 120). He insists that the political community dominated by those in the middle will be the best and the most well governed; that the middle class should, therefore be stronger than both the rich and the poor or, failing that, at least one of them. Both extreme democracy and unmixed oligarchy tend to give rise to factionalism, which often results in tyranny, whether in case of the rich minority choosing a guardian to suppress the masses in their name, or in case of the poor majority electing a champion to combat the will of the oligarchs (Ibid.). In either case, the ensuing factional tension merely represents the stepping stone for a would-be tyrant to rise to power by presenting himself as the defender of the disadvantaged.

Further illustrating the danger of unmixed constitutions, Aristotle notes that the reason why most city-states tend to give rise to either oligarchies or democracies is because those in the middle tend to be fewer and weaker than the rich and the poor, with either of the latter two taking superiority resulting from their dissentions as reward for their victory, and proceeding to establish neither a common constitution, nor an equal one (Ibid., 121). Therefore, he sees a state with just laws and a mixed government founded upon the middle class as the most stable and lasting system that preserves both freedom and wealth, while guaranteeing political rights through representation to all social strata.

***

Cicero was a Roman statesman and philosopher who lived during the turbulent era of the republic's impending downfall. Born to an equestrian family, he was a proud member of Rome's ancient aristocracy who had attained his political influence through both virtue and noble blood. Having exposed a conspiracy to overthrow the Roman republic as a consul, he was regarded as an outstanding statesman whose support was sought by many politicians, while having experienced first-hand the many machinations of various factions, including both the first and the second triumvirate. He was also personally hurt by them: his opposition to Caesar (Featherstonhaugh 1829, 8) meant that he was forced to spend some of his life in exile. A notable patriot and staunch defender of republicanism, he saw the rapid shift toward imperialism as a paramount threat to Rome's political institutions and society and sought to halt the influence of ambitious individuals and divisive factions by returning to the roots of what had once made his country into a free and stable republic (Ibid., 2).

A knowledgeable scholar of Greek philosophy, Cicero introduced the concepts of most major Greek schools of philosophy into Roman thought. His particular admiration for Plato, whom he praises on several occasions (Cicero 1829, 42), is evident in the way he writes, his main political treatise, *On the Commonwealth*, having been written as a platonic dialogue, with the famous Roman statesman and general Scipio Africanus taking the role that would normally belong to Socrates. Despite the considerable similarity of style, the substance of Cicero's work, a historically based analysis of republicanism founded upon realist assumptions about man and society, bears far closer resemblance to the works of Aristotle; indeed, their political theories hold enough similarities to rightly be considered a part of the same school of political though that would act as the foundation of the republican tradition.

Primarily basing his analysis on a detailed knowledge of Roman history, Cicero's goal in his treatise is twofold: on the one hand, much like his Greek predecessors, he offers a theoretical framework of government, which, in his case, much like in Aristotle's, is meant to be highly applicable to republics; on the other hand, his primary goal is to find the salvation from Rome's current decline, which he sees in returning to the republic's early constitution, it being a well-balanced system of institutions, ensuring the realization of each social stratum's political interests.

Cicero opens his discourse by describing a strange natural phenomenon of two visible suns, which could have been meant as a metaphor for the disunity between the senate and the people, resulting from factional strife, and signaling the republic's impending collapse (Cicero 1829, 53). He goes on to offer what could be the earliest historical definition of a republic, outlining that "A republic or commonwealth then, is the wealth or common interest of the people" (Ibid., 56). To insure its permanence, it must be governed by that authority which has a strong relation to the causes by which the republic came into being, namely, "the three modes of government" (Ibid., 57).

These modes aren't the specialized Enlightenment era branches of government, but rather public offices corresponding to the three different social strata. Acknowledging the traditional division between the three basic forms of government and their respective deviations, Cicero immediately points out their many shortcomings, while accepting that democracy is still the best among them for its common, equal enjoyment of freedom (Ibid., 60). In his view, subjects within kingdoms are too much deprived of common rights; under an aristocratic government, the people are not admitted into public office and therefore, cannot realize their interests, and under the majority, equality itself becomes unjust, as it admits no allowance for degrees of rank based on nobility and virtue (Ibid., 58).

Cicero claims that all extremes of a seemingly agreeable nature are inevitably converted to their opposites: the destruction of an aristocratic government proceeds from its unlimited power; the people's slavery arises from their boundless freedom just as tyranny springs from the most licentious liberty (Ibid., 75). These observations can be linked to Aristotle's belief that both unmixed oligarchy and extreme democracy tend to give rise to tyranny. Their stances differ in this regard though, as Aristotle believes a natural aristocracy to be the ideal constitution, while Cicero prefers the monarchy for its stability, as well as the danger of oligarchy and democracy lapsing into tyranny (Ibid., 77). He further promises an outline of the ideal constitution, much like the one in Plato's republic (Ibid., 80), but, as is the case with some others, the portion of the book in which he discusses it was, sadly, lost to history (Featherstonhaugh 1829).

Recognizing the monarchy as the most stable mode of government, democracy as the freest, and aristocracy as having the highest degree of political skill and virtue among its leadership, Cicero is nonetheless dissatisfied with their shortcomings, even in their good, uncorrupted forms: "Therefore I think a fourth kind of government, moderated and mixed from those three of which I first spoke, is most to be approved" (Cicero 1829, 60). He further says: "separately I do not approve of any of them; but should prefer to every one of them, a government constituted out of all three" (Ibid., 65). In pointing out that the state was never sound when the senate governed alone and that under kingdoms, the disadvantages were worse yet (Ibid., 61), he is likely referring to the early history of the Roman republic – the ousting of the last king, Tarquinius Superbus and the ensuing domination of the nobility in which the senate, alongside the consuls, enjoyed a monopoly on political power, though this could also be ascribed to kingdoms and aristocracies in general.

The implicit crucial moment of change here, which we'll see explored in greater detail in Machiavelli's theory, is the introduction of the institution of popular tribunes: upon its creation, all social strata had gained political representation through their respective modes of government, with the consuls representing to the monarchical mode of government, the senate corresponding to the aristocratic, and the tribunes to the democratic form (Cicero 1829, 107). However, as the enactment of the tribunes did little to curb the power of the senate in practice, a further push for democratic reform was made, by relinquishing political power to a new body, the decemvirate, tasked with reforming Roman laws after those of Athens. Their leader, Apius Claudius, who had previously shown tyrannical tendencies, reverted to his old ways, causing his government to fall to a revolution supported by both the people and the nobility (Ibid., 109), with the republic's mixed constitution being restored and functioning in relative harmony for several following centuries.

Cicero holds that the strength of laws rests in punishment and not in our natural justice and, furthermore, that natural right does not exist (Ibid., 121). He places special emphasis on the dangers of extreme democracy, likening Claudius' tyranny to those of Dionysius in Syracuse and Pisistratus in Athens (Ibid., 128). Finally, Cicero laments about the posterity and suggests that the republic's immortality might have been possible had the institutions of the Roman forefathers been preserved (Ibid., 126).

***

Renaissance Florence – home to both Niccolo Machiavelli and Francesco Guicciardini – was the vanguard of cultural and social transformation, while being embroiled in a centuries old internal struggle between the various factions holding power in Italy. Like ancient Greece, late fifteenth and early sixteenth century Italy was not the unified country that we know today, but rather a tumultuous mixture of small domestic maritime republics, interspersed with lands controlled by great foreign powers and the Vatican. The discord arising from their frequent wars meant that they could never contend with the large countries of Western Europe in terms of political and military power; nonetheless, they were at the forefront of technological and cultural development, never more so than during the Renaissance.

While the earliest roots of the cultural and social rebirth can be traced to the rediscovery of the works of Aristotle in the twelfth century, the fifteenth century can undisputedly be seen as its true beginning, with a drastic decrease of the church's influence (Mansfield and Tarcov 1996, xvii). The Renaissance can best be described as a revival of classicism and humanism, which had until then been repressed by the church: in addition to the reinvigoration of antiquity-inspired arts and sciences, philosophy, political though in particular, gained much traction, especially among the Italian states that much resembled the ancient city-states of Greece.

On the eve of the sixteenth century, the period of substantial cultural change also took a radical social turn in Florence: having been a relatively stable oligarchic republic for several centuries, its political structure was deeply shaken by the government of Gonfalonier (chief executive) Piero Soderini, whose reforms, centered on the creation of the Great Council, led to him being labelled a class traitor by most of the nobility (McCormick 2011, 40). This is the primary source of difference in the perspectives of Machiavelli and Guicciardini, with the former having been a member of the emergent class of professional bureaucracy within the new democratic government, while the latter served as an ambassador based on his lineage, a noble family of paramount influence that had produced no less than sixteen gonfaloniers in the past (Moulakis 1998, 27).

Having achieved his position in the Soderini government based on merit, Machiavelli still felt the brute end of the nobility's nepotism when they stifled his appointment as ambassador to the Holy Roman Empire. As the republic's wealthiest families held great influence in spite of the democratic reforms, the Gonfalonier was forced to submit to their demands (McCormick 2011, 40). This turn of events in addition to Machiavelli's ousting from political life after the fall of the Soderini government left the famous political thinker embittered toward the great, both for their misdeeds against the republic and against him.

The *magnum opus* of Machiavelli's political theory, *The Discourses on Livy*, even to this day remains in the shadow of the much shorter and more poignant work, *The Prince*. Harshly judged for immorality and rejection of religion by contemporaries (Kahn 2010, 240) and later interpreters (Strauss 1957, 36) alike, this brisk guide for a would-be ruler has no theoretical aspirations beyond basing an impromptu government on borderline extreme assumptions regarding man and society, the grim measures being undertaken to ensure the success of a goal directly linked to the turbulent period in which Machiavelli wrote: the unification of Italy and its liberation from "the barbarians", the many foreign powers that ruled a number of its regions (Machiavelli 1998, 101). And while the republican beliefs that he would greatly elaborate upon in his main theoretical work do surface in *The Prince* occasionally, the most significant being that "the end of the people is more decent than that of the great, since the great want to oppress and the people want not to be oppressed" (Ibid., 39), they are frequently overshadowed by the permeating advocacy of ruthlessness.

With quotes such as "it is much safer to be feared than loved" (Ibid., 66), "it is necessary to a prince, if he wants to maintain himself, to learn to be able not to be good, and to use this and not use it according to necessity" (Ibid., 61) and "men should either be caressed or eliminated, because they avenge themselves for slight offenses but cannot do so for grave ones" (Ibid., 10), we simply cannot be puzzled by the infamy that Machiavelli had garnered, though it should be noted that Guicciardini, while choosing his words more carefully, paints a no less grim picture of political reality.

Machiavelli has been described by some as a complex and deliberately enigmatic writer, much of which has to do with the period in which he wrote (Pocock 2010, 144). The obscurity of his true intentions in *The Prince* notwithstanding, the substance of his theory greatly transcends the pages of that book, as it barely scratches the surface of the concepts he would come to explore in much greater detail in his *magnum opus*, *The Discourses on Livy*: a vast, all-encompassing work of political philosophy, rooted in deep historical analysis, *The Discourses* forsake the particular, time-bound goals of contemporary politics in favor of establishing a system founded upon timeless laws of state and society.

With tremendous apparent influence from the works of Aristotle and Cicero, Machiavelli seeks to establish a system of institutions conforming to the unseen laws of political reality, while using republics both contemporary and ancient as a fount of practical examples. His main inspiration is the Roman republic, whose political reforms and institutional practices he takes as the paramount example of republicanism firmly based in and corresponding to realist assumptions about the nature of politics.

The main part of his discourse is centered on the institutional developments that followed the ousting of Rome's last king, Tarquinius Superbus, which would come to shape the politics of the republic for several following centuries. He begins by comparing the various ways in which republics can be created, concluding that Rome's foundation by Romulus established its independence from the beginning (Machiavelli 1996, 9), while the freedom of its citizens was enhanced by a flexible lawmaking process, which, unlike that of Sparta, meant that laws were enacted and evolved over a long period of time in order to adapt to the changing political circumstances (Ibid., 10).

Accepting the traditional tripartite division of government, Machiavelli immediately points out that principality easily becomes tyrannical, aristocracy regresses into government of the few and the popular government inevitably turns licentious, because of the likeness that virtue and vice have in their cases (Ibid., 11). With the three good modes having been shown to be short-lived, and their deviations malign, Machiavelli arrives at the conclusion quite similar to those of Aristotle and Cicero: that "those who prudently order laws having recognized this defect, avoiding each of these modes by itself, chose one that shared in all, judging it firmer and more stable: for the one guards the other, since in one and the same city there are the principality, the aristocrats, and the popular government" (Ibid., 13). Thus, he praises Lycurgus of Sparta for establishing a mixed government through a clear and just system of laws that ensured the state's longevity, while criticizing Solon of Athens for establishing a popular government that regressed into tyranny within his lifetime.

Recognizing Rome's initial kingship as one of the state's many defects that had to be fixed over time, he contrasts republics that were stable and well-ordered from the beginning with the ones that were forced to evolve through the circumstances arising from their imperfection. As mentioned previously, Rome had not yet become a true republic, as classical and early modern theorists understood the concept, immediately upon the ousting of the last king: with political power being shared between society's elites that expressed their will through the consuls and the senate, the people had no voice of their own within the government, which is what leads Machiavelli to the conclusion that the creation of the popular tribunes made the republic more perfect (Machiavelli 1996, 15). What makes his analysis of exactly how this transpired particularly original and provocative is the claim that the enactment of the office of tribunes resulted not from peaceful cooperation, but from disunion and collision between the senate and the people: "I say that to me it appears that those who damn the tumults between the nobles and the plebs blame those thing that were the first cause of keeping Rome free, and that they consider the noises and the cries that would arise in such tumults more than the good effects that they engendered" (Ibid., 16). This is fully within the spirit of his consequentialist approach, perhaps best formulated by his maxim: "when the deed accuses him, the effect excuses him" (Ibid., 29).

Finally, Machiavelli places great emphasis on well-ordered legal institutions, insisting that the Roman practice of public accusations was instrumental to both checking the power of the nobility, as well as keeping order in society. Using the example of Coriolanus, an enemy of the popular faction who attempted to starve the people into submission, he claims that, had the tribunes not levelled a public accusation against him, he would have been killed in a riot. The functioning of such legal institutions insures that justice is served even if the accuser is acquitted, inspiring fear in those who conspire against the common good, while allowing the people to vent their anger against the nobility (Ibid., 24). In addition to good laws, practices like this one are meant to ensure the republic's stability even in the most tumultuous of times.

***

The personal and political experience of Machiavelli's contemporary and fellow citizen, Francesco Guicciardini, whom he admired greatly, in spite of their disagreements (Moulakis 1998, 35), had been very different. Born to a noble family of paramount influence, he was groomed for high office from his youth; having received a doctorate in law, he joined the opposition against Gonfalonier Piero Soderini (Ibid., 28). In order to pacify the nobles, who retained much of their traditional privileges despite his democratic reforms, Soderini appointed Guicciardini as ambassador to King Ferdinand of Aragon. Whilst abroad, Guicciardini bore witness to the steady deterioration of Florence, which eventually prompted him to write a discourse on bringing order to popular government, naming it *Discorso di Logrogno*, after the town in Northern Spain where it was written.

In one of the great ironies of history, he finished the *Discourse* merely several days after the Soderini government fell to the Medici restoration (Ibid., 30). While being strongly critical of the ideas of Friar Savonarola, whose teachings inspired Piero Soderini, he admired the essence of his democratic reforms, particularly the creation of the Great Council, for the constitutional balance it provided (Ibid., 83). In accordance with their different social station, Machiavelli and Guicciardini took opposite views of political power in society: while the former took a chancery view of the emerging modern state, focusing on professional civil servants, the latter saw social continuity and coherence as dependent upon the political class whose members were distinguished by their knowledge and experience in the affairs of state (Ibid., 80).

Taking a generally elitist approach to government, Guicciardini insists on the importance of qualified majorities for the election of most officials and in the Senate's decision-making process (Ibid., 102). He prohibits debates within the Great Council, making them exclusive to the Senate from which legislation originates, while perceiving the office of the Gonfalonier as a guarantee of stability, unity and effectiveness of executive action (Ibid., 109). This system does not, however, prevent the people from taking an active role in politics, but rather relegates that role to legitimating and ratifying decisions made by the Senate and the Gonfalonier, based on Guicciardini's belief that the many lack the capacity and expertise required to deal with the complexities of government, but can be trusted to defend liberty, check tyrannical aspirations and advance the general welfare (Ibid., 105).

Establishing that "Political rule and command are nothing but violence over subjects, occasionally mitigated by a pretense of decency" (Guicciardini 1998, 121) and further, that "Liberty is nothing but the supremacy of law and public decrees prevailing over the desires of individuals" (Ibid., 122), Guicciardini founds his reforms upon firmly realist principles, while centering them on the aristocracy and its corresponding office, the Senate. Rather originally, he actually rejects the tripartite division of government, considering it to be open to abuse by demagogues, while claiming that Florence is too accustomed to liberty to have any need for its generality (Moulakis 1998, 99); nonetheless, his division of government is apparently based on the traditional system, as clearly evidenced by both the presence of all three modes in the offices of the Gonfalonier, the Senate and the Great Council, and in their respective roles, embedded in their connection with the social strata they represent.

Being the highest office in the republic, the primary virtue of the Gonfalonier, elected for life from among the most influential noble families, is political skill and experience (Guicciardini 1998, 124); however, with the power that enables the embodiment of that virtue also comes the risk of the people falling under the chief executive's influence, which may give rise to tyranny: "Experience shows and reason confirms that as a result of its weakness the multitude is never ruled by itself, but always seeks an allegiance and a prop" (Ibid., 125). Therefore, "To maintain true and complete liberty one of the most important things is surely this: that there be a mean by which the ignorance of the multitude can be controlled and the ambition of the Gonfalonier kept in check" (Ibid.).

The delicate position of the Gonfalonier must be carefully managed, as excesses could make him either useless, or a tyrant. In the process of his selection, three candidates should be proposed by the Senate, with the final election being left to the Great Council (Guicciardini 1998, 135). In its primary capacity as the barrier against the nobility's political monopoly, the Great Council has the final say in most essential laws and decrees, being barred only from decisions that require secrecy, as is the case with espionage and warfare. In this regard, it's important to note that while Guicciardini shares Machiavelli's grim view on human nature, he perceives the many, and not the few, as the embodiment of the worst in humanity, seeing the Great Council as rampant with ambition, evil and ignorance (Ibid., 126).

Nevertheless, being the body that represents the largest majority of the republic's citizens, its cooperation is the condicio sine qua non of any political undertaking: its contestatory role is based around the power to bar the decisions of any other office, including those of the Senate and the Gonfalonier, as well as those of smaller, elite bodies like the Council of Eighty, the Ten of War and the Signoria. In this regard, the originality of Guicciardini's system lies in withholding participatory powers from the Great Council, limiting its role to legitimation and ratification (Ibid., 128). While no decision, regardless of its perceived wisdom among society's elites, can justifiably be considered legitimate without the consent of the majority that it affects, Guicciardini argues that their deeper involvement in the process of decision-making would cause confusion, when taking into account the Great Council's vast membership, as well as the members' limited political competence (Ibid., 139).

The main body of deliberation and lawmaking is, therefore, the Senate: it represents balance and brings together the extremes of the one (Gonfalonier) and the many (Great Council), safeguarding liberty through laws (Ibid., 136). Its membership should consist of all the republic's wise men, chosen both for their noble lineage and their virtue, as "in fact, the entire weight of government ultimately rests on the shoulders of very few" (Ibid., 137). In defining the functioning of the Senate and its relation to the Gonfalonier and the aforementioned elite bodies of government, emphasis must be placed on well-informed deliberation, ensuring that members of the Senate are both presented with detailed proposals and given enough time to debate about them.

To further facilitate discussion, participants should be allowed to present multiple opinions, even if these run contrary to the organs they are part of, with even the Gonfalonier himself being allowed no more power than the senators during the process (Guicciardini 1998, 141). Here, Guicciardini contrasts his well-ordered system with the somewhat arbitrary rule of Piero Soderini, who, despite having made considerable progress toward a republican government by establishing the Great Council, still demonstrated dangerous ambition by pitting different factions against each other and manipulating votes (Ibid., 140).

Mixed Government: Tracing the History of an Idea

The republican tradition of western political thought is notoriously difficult to frame: while the term "republicanism" only entered broad theoretical discourse during the Enlightenment, its central concepts could be argued to have reached their zenith a few centuries before that, while separated from their original intellectual foundations by no less than two millennia. Though the idea of separation of powers became the defining characteristic of republicanism under the sweeping influence of Montesquieu's theory, its substantial difference from mixed government cannot simply be dismissed as the product of a lack of theoretical sophistication in the earlier period, as the political reality of our era has exposed the defects of the branching of government based upon the modes' role, and not the strata they stand to represent.

By contrasting the concepts of separation of powers and mixed government, we identified this distinction as the primary source of their difference. Furthermore, in following the history of the latter idea, we accentuated its central role in the theories of the four great thinkers, which displayed substantial similarity, in spite of their different personal and political circumstances and of the long gap between the two eras. On the one hand, the similarities can be somewhat comprehended when considering that the Renaissance was marked by the reinvigoration of interest in the classical period and its cultural and intellectual products. On the other, the subtle, but notable differences between the two eras point to the idea's survival through tenuous transformation, raising an intriguing philosophical question: did the early modern republican theorists simply excavate the long-forgotten concepts of the classical thinkers, or might have they, perhaps, arrived at similar conclusions through a different approach?

To answer this delicate question, we might have to look beyond their conclusions, and into the foundations of their discourses. The overview of each of the four theories has demonstrated the undisputed, central role of mixed government, stemming from the brevity of the good modes of government and the malice of the bad ones, as each author strongly criticized all three forms, and emphasized their defects in either theory or practice. Their insistence on the importance of laws as the guarantee of a lasting system is also directly linked to the concept of mixed government, as unchecked power of any social stratum is seen as one of the most pernicious threats to legality, and the greatest source of arbitrary rule. While making a just system based on any of the three modes of government is acknowledged to be theoretically possible, the difficulty of realizing it in practice mainly stems from the fact that any stratum, having been given too much power, will inevitably seek to use the state to its own benefit, even if it means abridging laws and causing detriment to others.

Closely linked to the issue of clear, just laws is the matter of institutions that embody them, their functioning being among the few substantial guarantees of preventing arbitrary rule; this applies both to legal institutions and to bodies of government that, while taking into account the small differences between each author's division and regulation of political power, in all cases serve as a system of checks and balances against each other, their active engagement with the affairs of state ensuring that the stratum they correspond to is well represented, while preventing others from acquiring boundless power. In this regard, the theories of both classical and early modern republicanism expose a significant weakness in comparison to that of the our era, as the particular roles of the different modes of government are often glanced over and vaguely defined, resulting in a lack of clarity when it comes to division of authority.

While each of the four aforementioned theorists has a number of unique and original ideas about the particulars, these are mostly relegated to non-essential issues: preferring one mode of government over another, while illuminating in terms of the theorists general perception of politics, is of little consequence when a mixed government is claimed by all to be superior. Likewise, the slight variations in formulating particular laws, managing policy prescriptions and organizing institutional arrangements can be viewed as minor disagreements that merely point to a general broad consensus on all crucial issues, with seemingly no major issue being the cause of separation between either different theories within the same era, or the two eras themselves, returning us to the original question of whether the strong apparent similarity between them may hide a foundational distinction of approach.

Therefore, we must turn to the subtle question of ideal constitutions, the kind that was deeply explored in Plato's *Republic*. While the matter may seem insignificant, it poignantly reveals the vastest sphere of division between the theories of classical and early modern republican thinkers. Without going into it in nearly as much detail as his teacher, Aristotle still defined a natural aristocracy as the ideal mode of government, with all the other modes, including other aristocracies, being its deviations (Reeve 1998, lxxii); Cicero, having promised to offer his view of the ideal constitution (Cicero 1829, 80), did so in a part of his discourse which was lost to history (Featherstonhaugh 1829), though, judging from his observations about the existing modes of government, we likely wouldn't be remiss in assuming that he saw it as a kind of perfectly ordered monarchy.

Quite strikingly, neither Machiavelli, nor Guicciardini offer any conceptions about the ideal constitution, and the best description of what was probably their shared reasoning on the issue can be found in one of former's most illuminating quotes: "Since my intent is to write something useful to whoever understands it, it has appeared to me more fitting to go directly to the effectual truth of the thing than to the imagination of it. And many have imagined republics and principalities that have never been seen or known to exist in truth; for it is so far from how one lives to how one should live that he who lets go of what is done for what should be done learns his ruin rather than his preservation" (Machiavelli 1998, 61).

The implications of this approach point to a different perception of the very origin of society and state; thus, they also imply a seemingly subtle, but nonetheless crucial distinction in understanding its foundations. While the classical thinkers viewed political community as a naturally emergent phenomenon (Aristotle 1998, 5), the early modern theorists preferred to avoid the issue altogether, seeing no relevance in it for lack of a possible concrete application. However, in a quote strikingly similar to what would later become the basis of the social contract theory, Machiavelli offers his understanding of the formation of the state, observing that people were few and far between, living like beasts in a sparsely populated world; pressed for resources and defense as their numbers grew, they were essentially forced to create a political body as a means of insuring survival (Machiavelli 1996, 11).

Though he offers no elaboration of this thesis, the Florentine Realists' abandonment of the classical concept of the natural human tendency toward statehood may have paved the way for the social contract theory, signifying one the crucial moments of the modern political thought's departure from its ancient roots. On a more general note in regards to the history of the idea of the state's causes, the Florentine Realists' historically inclined apathy toward the issue can be argued to represent the transitional stage between the ancient, organic, and the modern, individualistic conception of society.

Despite the undeniable influence of classical thinkers on early modern republican theorists, firmly embedded within the spirit of the Renaissance as a period of cultural and intellectual transformation, the differences in perceiving the political phenomenon of the state's initial causes as natural in the former case, and historical in the latter, certainly played a significant part in the somewhat different structuring of theories, offering an explanatory perspective on the generally tenuous, subtle distinctions between the two eras.

Conclusion

In spite of their small differences, most of which can be reduced to cultural and social circumstances, the theories offered by the early developers of the republican tradition achieve a broad consensus on a number of crucial issues, and thus may hold the key to solving the problems inherit in very foundations of the later currents within the tradition. While clearly surpassed in the area of specifically defined roles of the modes of government, classical and early modern republican thought, founded on the concept of mixed government, could offer an additional layer of branching that would enable us to overcome the problem of a lack of political representation stemming from economic inequality.

John McCormick lays the foundation for exploring this concept, arguing that the democratic element is lacking in modern politics, particularly in the United States: he explores a potential solution through a thought experiment, proposing the enactment of the institution of popular tribunes via a constitutional amendment (McCormick 2011, 179). Whether or not such a development ever really occurs, the very recognition that solutions can be found in the past is hopeful in itself.

Specifically, in the context of mixed government being contrasted with separation of powers, the two different tripartite divisions of government offer divergent, but potentially compatible solutions to the problems presented by political reality. And while the deficiencies of the present system of governmental branching could eventually lead to a clamor for its complete replacement with the alternative concept, the shortcomings of mixed government necessitate a union between the two which, much like the systems proposed by the four theorists, would seek to combine the virtues of the two, while transcending their vices.

Changes to existing systems that would improve the political balance between the social strata would, of course, be extremely complex, and would require great knowledge of individual systems in which their application would be attempted, but their general characteristics would include introducing elections to certain offices by lot, which tend to negate the political privileges of wealth and recognition, an institution in the spirit of public accusation discussed by Machiavelli, eligible to be leveled at any public official, regardless of his station, and a citizen body, much like Piero Soderini's Great Council, the contestatory role of which would greatly improve the balance in representation and fundamentally increase the legitimacy of the decision-making process.

References

Aristotle. 1998. *Politics*. Translated by C. D. C. Reeve. Indianapolis and Cambridge: Hackett Publishing Company.

Cicero, Marcus Tullius. 1829. "On the Commonwealth." In *The Republic of Cicero*, translated and edited by G. W. Featherstonhaugh, 33–148. New York: G. & C. Carwill.

Guicciardini, Francesco. 1998. "Discorso di Logrogno/On How to Order Popular Government." In *Republican Realism in Renaissance Florence* edited by Athanasios Moulakis, 117–149. Lanham, Maryland: Rowman & Littlefield Publishers, Inc.

Featherstonhaugh, G. W. Introduction to *The Republic of Cicero*, translated and edited by G. W. Featherstonhaugh, 2–31. New York: G. & C. Carwill.

Kahn, Victoria. 2010. "Machiavelli's Afterlife and Reputation to the Eighteenth Century." In *The Cambridge Companion to Machiavelli*, edited by John M. Najemy, 239–256. Cambridge: Cambridge University Press.

Machiavelli, Niccolo. 1996. *Discourses on Livy*. Translated by Harvey C. Mansfield and Nathan Tarcov. Chicago: The University of Chicago Press.

Machiavelli, Niccolo. 1998. *The Prince*. Translated by Harvey C. Mansfeld. Chicago: The University of Chicago Press.

Mansfield, Harvey C., and Nathan Tarcov. 1996. Introduction to *Discourses on Livy* by Niccolo Machiavelli, xvii–xliv. Chicago: The University of Chicago Press.

McCormick, John P. 2001. "Machiavellian Democracy: Controlling Elites with Ferocious Populism." *American Political Science Review* 95(2): 297–313.

McCormick, John P. 2011. *Machiavellian Democracy*. Cambridge: Cambridge University Press.

Montesquieu, Charles-Louis. 1989. *The Spirit of the Laws*. Edited by Anne M. Cohler, Basia Carolyn Miller and Harold Samuel Stone. Cambridge: Cambridge University Press.

Moulakis, Athanasios. 1998. *Republican Realism in Renaissance Florence*. Lanham, Maryland: Rowman & Littlefield Publishers, Inc.

Plato. 2005. *Republic*. Translated by C. D. C. Reeve. Indianapolis and Cambridge: Hackett Publishing Company.

Pocock, J. G. A. 2010. "Machiavelli and Rome: The Republic as Ideal and as History." In *The Cambridge Companion to Machiavelli*, edited by John M. Najemy, 144–156. Cambridge: Cambridge University Press.

Reeve, C. D. C. 1998. Introduction to *Politics*, by Aristotle, xvii–lxxix. Indianapolis and Cambridge: Hackett Publishing Company.

Strauss, Leo. 1957. "Machiavelli's Intention: The Prince." *The American Political Science Review* 51(1): 13–40.

Nikola Đurković,
Department of Philosophy,
University of Belgrade

Original Scientific Paper
UDK 165.191:159.9

THE PSYCHOLOGICAL PLAUSIBILITY OF RELIGIOUS FICTIONALISM

Abstract: *In this paper I explain the psychological plausibility of religious fictionalism by finding the proper form of fiction that is analogous in relevant aspects to religious practice. First I examine the forms of fiction that are commonly listed in literature and explain why participating in these forms does not resemble taking part in religious community. After that, I establish characteristics of religious practice that the appropriate form of fiction would have to share, and find that acting shares most of these features. However, I argue that acting is still significantly different in some aspects, and propose the method acting as practiced by Daniel Day-Lewis as the only form of fiction that is analogous in all relevant facets. Then, by exploring how a method actor relates to the objects of the fictional world and to the real objects that do not belong to that world, and by finding similarities in his method and in religious fictionalism, I explain how the religious fictionalist differs from the realist in the content of his belief. I also use this analogy to explain the motivation for being a religious fictionalist: namely, to still be able to participate in religious practice because we find something worthy in it, although some doctrines of religion directly contradict the facts of science and our moral ideals.*

Key words: *religious fictionalism; psychology; religion*

Religious fictionalism is often defined in opposition to religious realism and religious positivism. Religious realism interprets statements of religion as either true or false statements about transcendental reality[1]. A believer claims that these statements are most probably true while an atheist claims that these statements are most probably false[2]. The religious positivist interprets statements of religion as either true or false, not about transcendental reality but about our moral and spiritual lives (LePoidevin 1996, 112). Religious fictionalism does not interpret statements of religion as either true or false, but as useful fictions[3] for

[1] Or even physical reality, if pantheism is an option.

[2] Here, I assume that both believer and atheist are in a certain sense epistemically modest, and that they do not know with absolute certainty that there is God. Most rational people would leave an open possibility that they could be mistaken about beliefs so complex like religious ones.

[3] Fictions can still be either true or false (it is possible that *Lord of the Rings*, for example, describes some real state of affairs in a galaxy far away), but are taken to be highly improbable, and most significantly, their probability is not important for the role they play.

providing moral guidance and bringing people together. Although the entire religion is understood as fiction, it is supposed to be the kind of fiction that is so powerful that it could produce deep emotions, function as social glue that binds people together, and guide moral lives of entire communities.

One common objection to religious fictionalism is that it is psychologically implausible (LePoidevin 1996, 112 and Eshelman 2005, 190). How could people celebrate Christ's resurrection if there was no Biblical Christ after all? How could people fall on their knees and give their entire being to the Lord if they believe there is no God? How could they pray if nobody is listening? How can the Ten Commandments motivate our moral lives when we know that they are just a part of tradition that existed a long time ago, and that was, according to our common modern judgment, racist, misogynic, and had a very strict social stratification?

It is important to correctly understand what the focus of this objection is. It does not say that it would not be useful to practice religion, once we understand its statements as fiction. It does not deny that faith understood this way can promote moral values or bind people together. What it says is that engaging in religious practice seems **psychologically** implausible once we interpret statements of religion as fiction. It would take some really strange psychological state for somebody to pray to God that is not real. Possibly an act of self-deception (Eshelman 2005, 193) or even some kind of madness.

An analogy with our enjoyment of the usual forms of fiction is often made to explain and to motivate the plausibility of religious fictionalism. There is nothing strange in people having deep emotions about something that they consider not to be real. The "red wedding" scene in *Game of Thrones* caused numerous people to weep and yet hardly any of them thought that Rob and Catelyn Stark where real people. The actions of the elder Zosima from *Brothers Karamazov* morally transformed many people, but hardly any of them believed that he was a real person. Fiction is a powerful medium that can greatly influence all kinds of persons and there is nothing psychologically strange about it.

In this paper I will inquire as to what is the form of fiction most analogous with religious practice, and how far this analogy can be stretched. In the end, the psychological plausibility of religious fictionalism will depend on the answer to this question.

Comparison is often made between religion and the forms of fiction that we enjoy as observers. Thus, LePoidevin refers to Hitchcock's *Psycho*, Dickens' novels and TV soaps (LePoidevin 1996, 115, 119). Eshelman is more careful here, but he still talks about experience of watching a play (Eshelman 2005, 193). Nonetheless, as these authors sometimes notice, this parallel falls short in one important aspect. Religion is a practice, and requires our participation, but our enjoyment of, say, reading a book and watching a film is of a different form; when we watch *The Master* or *Casablanca* we experience emotions as if that fictional world is real but **we do not act** as if it is real. We are always aware of the boundaries of that world, whether they are the edges of a TV or a projector screen. Even when there is a live recreation of the fictional world, as it is in theater plays or operas, we do not act as if that fictional world is actually there. If we rushed on to save Romeo and Juliet from committing suicide, people would say that we misunderstand what theatre is about. Our presence in that world is like the presence of an eye/mind which is able to see parts of the world that the author wants us to see, but which cannot interact causally with.

However, when we engage with religious material, our attitude is clearly different. We interact with God when we pray to him and we celebrate Easter with people that share our beliefs. We are part of a community that participates in a fictional world, not just observers of it. As Eshelmans puts it

> /.../ the fictionalist view/.../ requires not simply that one be moved by a theistic narrative but to take part in the fiction through rule-governed ritual behaviors. In other words, the analogy here is not that of viewing a play but of being in the play and following the script (Ibid., 193).

Therefore, our participation is fundamental for our engagement in religious practice.

Although Eshelman mentions the analogy with acting in plays, he quickly moves on to the analogy with sport preparation, which, because it lacks the emotional and moral aspects, is not similar enough to a religious practice. LePoidevin also mentions an analogy that includes activity – children playing a game of make-believe together (LePoidevin 1996, 116) – but then, because such an analogy includes only quasi emotions and lacks duration he moves on back to TV soaps (Ibid., 121). Both of them do not develop analogies which involve active participation enough to encompass all of the relevant elements of religious practice. I will try to do that in the rest of this paper. However, first we must establish conditions which determine something as a genuine religious practice, conditions that any fictionalism we are looking for has to meet.

Genuine religious practice has to:

1.) **involve participation** – we have to be part of that (fictional or real) world.

2.) **excite genuine emotions**[4] – It cannot resemble something like a children's make-believe game because a child wouldn't play a game of running from a monster if her fear was the same as fear of an actual monster.

3.) **be pervasive, long-lasting and continuous**, occupying a significant place in our lives.[5]

[4] Genuine emotions could be loosely defined as emotions that have the same **strength** as the typical manifestations of these emotions. A child that is pretending to run from a tiger usually does not have genuine emotions, because she knows that there is no tiger there, and that there is nothing really to fear about. Her fear is fear "light".

[5] One possible objection to this list is that it sets extremely strict criteria for something to qualify as a religious practice. However, I insist on it, because I want to explore whether religious fictionalism can replace deep and powerful religious practice of the realists, not be some kind of religion "light". The other possible objection is that the religious practice does not have to involve any emotions. That is, the relation with God and the religious practice could be entirely based on the rational reasoning and respect for the higher power. However,

The element of duration is especially hard to capture, because our involvement with fiction is often periodical and it doesn't last more than few hours. But religion, even fictional, should be constant although practice itself isn't. LePoidevin tries to capture the element of duration with TV soaps, saying that they last for a long time (in terms of episodes) and occupy a significant percentage of our thoughts (LePoidevin 1996, 121). But that analogy is still problematic because enjoying TV soaps is not participation-based and is not continuous. An analogy with acting in plays that Eshelman makes is the best candidate, but it still obviously lacks duration and the question of quasi emotions is still a legitimate one. It seems that there is no obvious form of fiction which shares all three properties that religious practice should have. This is the reason why these authors often change between different forms of fiction, depending on the property of religion that they discuss. However, in the end it seems that there is no proper enjoyment of fiction which is analogous to religious practice in all relevant aspects, and that religious fictionalism is therefore psychologically implausible.

Still, I think there is one curious form that has not yet been considered. I want to propose a strange practice of method acting popularized by Daniel Day-Lewis as a form of fiction similar in all relevant aspects. First I will briefly explain what makes ordinary practice of acting successful, and then expand it by introducing the idea of method acting.

In his "Understanding acting" Richard Horbny lists three crucial characteristics of good acting performance and two of those three concern us here. The first one is the ability to of an actor **to relate** with every object that belongs to the fictional world of the play (Horbny 1983, 28). An actor should be able to relate to props on the set as if they are actually the things they represent: to use a fake plastic gun as a genuine killing device, to turn away from a paper mask as if it were a real disfigured face of some freak. He should also be able to relate to the other actors as if they are those persons they play. The actor also ought to have the capacity to relate to imaginary beings of the world of the play. If he is just looking at the camera, with the cameraman and the director behind it, but he is supposed to see an approaching tiger, he should have a vivid representation of the beast in his imagination.

The second characteristic is **the pursuit of the objectives** (Ibid., 31). Almost every character of some fictional story has certain goals that he is chasing after and every last bit of an actor's behavior should be performed with those goals in his mind. Relating is not enough. It brings only spatial dimension of the fictional world. If an actor only holds a gun as if it is a real killing device, but without having a goal of stopping a villain and saving his beloved, his performance will be lacking. The goal directed behavior of an actor brings the temporal dimension to the fictional world of the play in question. He is not relating just to the real and imaginary parts of the fictional world around him, but also to what that

I do not think that this is a plausible view since the emotion of love plays very important role in all major religions.

world has been, what it will be, and how he can bring about important changes in that fictional world.

We can see how this correlates to first characteristic of religious practice. Both relating and goal oriented behavior are crucial for genuine participation in the fictional universe. Indeed, good acting and religious practice must include both of them. When we apply these two rules to the behavior of religious fictionalist we will see that, for example, she has to relate to the bread and wine as if they are the holy relics of Jesus Christ. She should relate to the fictional God as if he is a real deity. But the fictional universe of religion must also have temporal dimension and her behavior has to be goal oriented. In prayer, she is praying for her goals to be realised in the future, and her religious practice is related, as far as her memory serves her well, with everything that she did in the past.

However, the question of genuine emotion and the problem of continuity still remain. Actors leave the fictional world once that play is over or that camera stops shooting, and they return to normal life. They might be still emotionally attached to that world but the prop gun is now just a prop, and his goal is not the goal of the character that he is playing but of his own goal. As Hornby said: “It is important that an actor be sensitive and imaginative, but not that he be a lunatic” (Horbny 1983, 28).

But what if the actor, in his attempt to reach the highest connection with the fictional world, behaves in a certain way that would resemble lunacy to an outsider? What if he didn’t break a role for the whole duration of the shooting? That is the requirement of method acting, and one famous actor – Daniel Day-Lewis brought that method to the extreme. In *My Left Foot* (1989), Day-Lewis spent eight weeks at a cerebral palsy clinic in Dublin, learning to speak as the character he played spoke, and to write and paint with his left foot, as his character did. He famously said that: “films don’t begin only when the camera starts rolling”. During the whole process of filming he stayed in character at all times, never leaving his wheelchair. He was lifted into and out of the car that brought him to filming, and over the cables that littered the set, and was fed by members of the crew (The Telegraph 2013). While playing Hawkeye in *The Last of the Mohicans*, Day-Lewis learned how to capture and skin animals, construct canoes, battle with tomahawk axes and shoot a 12-pound flintlock while running. Day-Lewis insisted on carrying his gun wherever he went, even bringing it to lunch with his family on Christmas day. When Day-Lewis completed filming, he suffered from hallucinations and claustrophobia; “I’ve no idea how not to be Hawkeye”, he told to the director (Ibid.). He had a similar approach to many of his other films, like *There Will Be Blood, Gangs of New York,* and *Lincoln*.

This approach to fiction has the element of the continuity that religious practice requires. It is also all-encompassing, occupying the entire actor’s life during the shooting. Concerning the question of quasi-emotions, I would claim that method acting produces emotion as strong as they can be when produced by fiction and clearly as strong as emotions produced by real events of that kind. However, I am not a method actor, and perhaps the levels of the emotions in questions are so nuanced that only they could tell us the truth.

Therefore, it seems that there is a good analogy with fiction after all that can make religious fictionalism psychologically plausible. In the practice of method acting there are all three important properties of religious practice: engaging with the community that is related to the fictional world, feeling genuine emotions and participating in activity that is pervasive and continuous. The only problem for the analogy is the impact that advanced method acting has on the persons practicing it. Remember the word of Day-Lewis: "I've no idea how not to be Hawkeye". It could seem to an outsider that such deep and continuous engagement with some fictional world can temporarily erase the boundary between the fictional and the real. And since the shootings of the movies end after a few months or years, but religious practice last for a lifetime, interpreting fictionalism in a way similar to method acting could prove problematic for the very existence of that position. That is, according to this analogy, fictionalism would be psychologically plausible but it would dissolve into realism when applied to religion.

To determine whether this happens we need to fully understand how the method actor experiences the world during filmmaking. However, to be able to do that one has to have first person experience of method acting. Since I do not possess that kind of experience, as an alternative I am going to try to guess what happens based on my third person point of view.

It seems to me that any good method actor has two related but distinct abilities. The first one is to suspend his beliefs about fictional objects being unreal (I will call them contextual unreal objects). The second one is to suppress the beliefs about the existence of some real objects that do not belong to the context of fictional world (I will call them non-contextual real objects). Concerning the former, he is able to interpret, say, fictional character's motives as real motives and a prop sword as a real sword. Concerning the later, he is still aware of the camera and the microphone that follow him throughout the shooting and he stills listens to a director for the instructions between the scenes, but at the same he is not fully conscious of them. He is like a person who is driving a car, which concentrates on the road ahead, but does not pay immediate attention to it.

It is important to notice that, if my interpretation of his world is correct, it follows that his experience is not the same as that of the realist. This is because their beliefs have different content. Although it could be argued that his attitude towards contextual unreal objects is such that during his highest focus of method acting he actually perceives them as real, it would be implausible to claim that he isn't aware at all of non-contextual real objects like the camera, the script, the director, the lighting and so on. His awareness is not full, but it is still some kind of awareness. Thus, this form of fictionalism in method acting does not dissolve into realism.

Now we need to inquire if the same happens when this form of fictionalism is applied to religion. If something corresponding to the non-contextual real objects exist in the religious practice of the fictionalist, then interpreting fictionalism as analogous to method acting would not lead to the merging of the religious fictionalism and realism. At the first glance in the world of the religious fictionalist there are no objects similar to the non-contextual real objects. By the presupposition of religion the entire universe was created by God, and thus everything that exists belongs to the world of fiction. However, for many fictionalists there are actually some objects that do not belong to that world of religion and these are the facts of science and moral ideals that directly contradict the religious doctrine. Theory of the evolution, with the fossil and geological records supporting it, would be an example of these facts that contradict the idea that man is somehow special and divine, and that he was made by God to be radically different from other animals. Some parts of physics, like celestial mechanics would contradict the idea that the world was built according to the well-thought design of the divine creator. The belief that animal sacrifice is morally wrong would be the example of the other kind. That belief would be put aside when a religious fictionalist celebrates that God send an ox to Abraham to sacrifice instead of his son (*Genesis* 22).

If the religious fictionalist accepts these scientific theories and moral ideals, his relation towards them during strong religious focus would be similar to that of the method actor to non-contextual real objects. Specifically, he would still believe that those theories are true and the ideals correct, but he would suppress that during the religious practice. Therefore, it seems that there is a strong correlation between the fictionalism in method acting and the religious fictionalism, and the later does not become religious realism if we apply the analogy.

Although the final stage of this argument may seem odd, I think that it plays the perfect part for most of the religious fictionalists, since it appears to me that the motivation behind this theory is to find a form of the religion that can incorporate the changes that happened in science and in our moral judgment during last five centuries.

Conclusion

This analysis has led us to the understanding of how religious fictionalism can be psychologically plausible. It turned out that for something to qualify as a religious practice, it had to meet rather strict criteria concerning involvement and dedication, and that the only practice related to fiction which fulfills these criteria is the kind of method acting championed by Daniel Day-Lewis. The fact that this is the only proper analogy shed some vital light on the distinction between religious fictionalism and realism, and, in the end, uncovered the link between the motivation for religious fictionalism and its psychological plausibility. It seems to me that the educated and liberal person is the one who chooses religious fictionalism over realism because she sees something valuable in religion, but she cannot interpret religion as a realist because that contradicts her worldview. Despite this she can still participate in the religious practice, relate to the objects and people of that practice, and suppress her belief in facts of science and moral ideals in the same way that a method actor does in order to relate to the fictional universe during filming. This shows us that it is psychologically possible to participate in the complex and demanding world of religion and still be fictionalist about it. From the great proponents of the method acting – primarily Daniel Day-Lewis – we can learn more about how that could be done.

References:

Eshelman, Andrew S. 2005. "Can an Atheist Believe in God?" *Religious Studies* 41(2): 183–199. doi: 10.1017/S0034412505007602

Horbny, Richard. 1983. "Understanding Acting." *Journal of Aesthetic Education* 17(3): 19–37.

LePoidevin, Robin. 1996. *Arguing for Atheism*. London and New York: Routledge.

The Telegraph. 2013. "The Method Madness of Daniel Day-Lewis." *The Telegraph*, Jan 23. http://www.telegraph.co.uk/culture/culturepicturegalleries/9819469/The-Method-Madness-of-Daniel-Day-Lewis.html?frame=2459178

LIST OF REFEREES FOR THE BELGRADE PHILOSOPHICAL ANNUAL (2014, 2015, 2016):

Jonelle DePetro (Eastern Illinois University)
Anand Jayprakash Vaidya (San Jose State University)
Brian Huss (York University)
Jan Narveson (University of Waterloo)
Louis Groarke (Saint Xavier University)
Werner Sauer (University of Graz)
Martin Vacek (Australian National University; Slovak Academy of Sciences)
Jurg Steiner (University of North Carolina, Chapel Hill)
Allan Franklin (University of Colorado, Boulder)
Angelo Corlett (San Diego State University)

Jovan Babić (University of Belgrade)
Milan Ćirković (Astronomical Observatory of Belgrade; Oxford)
Leon Kojen (University of Belgrade)
Ljiljana Radenović (University of Belgrade)
Milos Adžić (University of Belgrade)
Duško Prelević (University of Belgrade)
Kosta Došen (University of Belgrade)
Dejan Vuk Stanković (University of Belgrade)
Ivan Mladenović (University of Belgrade)
Leon Kojen (University of Belgrade)
Dusko Prelevic (University of Belgrade)
Mašan Bogdanovski (University of Belgrade)
Slobodan Perović (University of Belgrade)

CIP – Каталогизација у публикацији
Народна библиотека Србије, Београд

1

FILOZOFSKI godišnjak = Belgrade philosophical annual / editor Slobodan Perović. – God. 1, br. 1 (1988)– . – Belgrade : Institute of Philosophy, Faculty of Philosophy, 1988- (Belgrade : Službeni glasnik). – 24 cm

Godišnje. – Glavni stvarni naslov od br. 28 (2015) Belgrade philosophical annual. – Tekst na engl. jeziku.

ISSN 0353-3891 = Filozofski godišnjak

COBISS.SR-ID 15073792