

Tina Radkohl & Kerem Akdogan

The Physics of Free Agency

as part of the Paradoxes VU (2022)

by Gemma De las Cuevas

University of Innsbruck, Austria

2022

Contents

I. Introduction	3
1. Motivation	3
2. Definitions	3
II. Free Agency and Determinism	5
1. Compatibility of Free Agency and Determinism	5
2. Incompatibility of Free Agency and Determinism	5
III. Free Agency and Indeterminism	7
1. Incompatibility of Free Agency and Indeterminism	7
1.a. Luck Objection	7
1.b. Mind Argument	7
2. Compatibility of Free agency and Indeterminism	8
2.a. Projective Simulation	8
2.b. Relating Projective Simulation to Free Agency	10
IV. Conclusion	11
V. References	12

I. Introduction

This essay is a brief introduction on the physics of free agency whereby we reflect on the compatibility of free agency in an (in)deterministic world. We do not claim the world to be either deterministic nor indeterministic as this is another topic for discussion and goes beyond the scope of this text. We thus consider arguments for and against the compatibility of free agency in a deterministic (Section II) as well as in an indeterministic world (Section III). All arguments in sections II and III are based on examples from literature, however, for the purpose of showing different perspectives only some main arguments are presented. For details and further reading, we refer to the sources indicated. Section IV concludes the essay with a short summary of the reflections.

1. Motivation

The goal of this overview is to illuminate the four possible combinations of free agency and (in)determinism:

Free agency is ... compatible with determinism.

... *incompatible* with determinism.

... compatible with *indeterminism*.

... *incompatible* with *indeterminism*.

To achieve that, we will first define the most important terms and take a quick look at the meaning of free agency and the required criteria.

2. Definitions

The following definitions are highly inspired by Müller and Briegel and their work on “A Stochastic Process Model for Free Agency under Indeterminism” [1], where they describe each term in more detail and illustrate them with examples.

Definition I.1 (Determinism). The laws of physics and the current state of the world fully *determine* its unique evolution. Hence, there is only one real possibility how the future can turn out.

Definition I.2 (Indeterminism). The negation of determinism.

Definition I.3 (Free Agent). Entity that exhibits free agency.

Definition I.4 (Free Agency). The capacity of free agents to act freely at least some of the time.¹

Definition I.5 (Free Action). An agent doing something (or refraining from something) and thereby *actively* influencing its environment.

It is important to note that a free action lies between *mere behaviour* and *acting out of free will*. As Müller and Briegel point out, the former refers to pure causal attribution while the latter implies moral attribution.

Definition I.6 (Free Will). The ability to do otherwise.

And: The conscious decision to choose one's actions and thus being morally responsible for the action.

A discussion on the physics of *free will*, as was originally intended, seems therefore out of reach as (to our knowledge) no physical law has any expressiveness on morality. However, we agree with Müller and Briegel that free agency forms a precondition for free will and thus believe it forms a good starting point for our discussion.²

¹Müller and Briegel offer 3 necessary and 1 sufficient criteria for free agency: Free agency (1) has causal influence, (2) is non-rigid, (3) is flexible and adaptive, and (4) is sensible. For details please see [1].

²Furthermore, the lack of a proper common definition of what is and what is not part of free will is another reason why we will follow Müller and Briegel and focus on free agency.

II. Free Agency and Determinism

1. Compatibility of Free Agency and Determinism

Living in a world under the rule of Laplace's Demon, there is not much space for free agency. No choice between different actions, no choice between several futures. But is this Demon actually realistic or just a fear from the old days? Laplace couldn't know at his time that there is a fundamental limit to the information we can collect about a system due to the Heisenberg Uncertainty Principle. Thus, his Demon is fiction. To determine the agent's future decision our knowledge about the state the agent's brain is in has to be can exceed a certain limit such that it is not predictable anymore. Hence, it depends on how far our prediction should reach the future - in this sense free agency is compatible with determinism.

2. Incompatibility of Free Agency and Determinism

Referring to Laplace's Demon seems to be outdated since the interpretation of one of physics' most successful theories is based on true randomness.

However, besides noting that quantum randomness, which cant be influenced by anything, destroys our notion of free agency, there are hints which call for a deterministic fundamental theory, which quantum mechanics could be derived from. One of these hints is the measurement problem. The dynamic of a system can be derived by the Schrödinger equation, which is linear. But the measurement process in quantum mechanics isn't, so it's incompatible with the Schrödinger equation.

Thus, you either drop the reductionism (which is unscientific) or state that the Schrödinger equation is not fundamental.

So can quantum mechanics be some sort of statistical approximation to a deterministic theory?

The violation of bells inequality tells us, that the assumption of a local hidden variable theory is violated. That's why the underlying deterministic theory is commonly refused.

What most physicists don't know, there is another assumption that could be violated which would allow you to keep the local hidden variable theory.

The assumption is called statistical independence, which states that the time evolution

of a prepared state depends on detector settings

$$p(\lambda) \neq p(\lambda|detector),$$

where λ are the hidden variables of the state.

Any local hidden variable theory which fits our observation has to violate statistical independence. A superdeterministic theory is one that violates the assumption of Statistical Independence [2].

Thus, having a superdeterministic fundamental theory would prune our free agency defined as an ability to have done otherwise [3].

III. Free Agency and Indeterminism

1. Incompatibility of Free Agency and Indeterminism

The main arguments suggesting incompatibility of free agency and indeterminism are the *Luck Objection* and the *Mind Argument*. Both arguments are available in several different versions (see e.g. [4, 5]) and have been argued against equally often (see e.g. [6, 7, 1]).

Interestingly, advocates of the luck and mind argument generally come from *libertarians*, which assume free agency exists and infer from this that determinism must be true. However, it is a fallacy to conclude that the universe is deterministic without first showing that (a) free agency is incompatible with indeterminism *and* (b) free agency does exist. Since we are only concerned with (a) in this section and do not make any statements whether or not indeterminism is true, the luck and mind argument are sufficient.

1.a. Luck Objection

The key points of any Luck Objection are that (a) undetermined actions occur due to luck, where luck refers to randomness and (b) the presence of luck prevents the agent from being in control, i.e. acting freely. Thus, instead of indeterminism freeing the agent, the agent is now dependent on luck and cannot have free agency.

1.b. Mind Argument

The Mind Argument claims that an agent cannot act freely because it has no control over what is undetermined. The difference to the Luck Objection is that here, no luck is required instead the indeterminism itself leads to loss of control.

However, as we will see shortly, both arguments are refutable for the same reason.

2. Compatibility of Free agency and Indeterminism

The view that free agency is compatible with indeterminism is approached from two perspectives.

On one hand, it is derived by advocates of the incompatibility of free agency and determinism, who assume free agency to be true. However, this argument fails to acknowledge that free agency might simply not exist and is thus not satisfactory³. One of its proponents is van Inwagen, who admits that this argumentation leads to a paradox:

If free will is incompatible with determinism, we are faced with a mystery, for free will undeniably exists, and it also seems to be *in*compatible with indeterminism. That is to say: we are faced with a mystery if free will *is* incompatible with indeterminism. Perhaps it is not.[4]

On the other hand, a new perspective has been brought forward that argues directly for the compatibility of free agency and indeterminism (see [8] and [1]) and is considered in more detail below.

2.a. Projective Simulation

Before directly arguing for the compatibility of free agency and indeterminism we first give an overview of the parts of projective simulation that will play an important role in the argumentation below. All details of this model on how the agent interacts with its environment can be found in “Projective simulation for artificial intelligence”[9].

Important definitions:

Definition III.1 (Intelligence). The capability of the agent to perceive and act on its environment in a way that maximizes its chances of success.[9]

Definition III.2 (Creativity). The distinguished capability of dealing with unprecedented situations and of relating a given situation with other *conceivable* situations.[9]

Definition III.3 (Clips). Clips represent basic (but variable) units of memory which will be accessed, manipulated, and created by the agent.[9]

Projective simulation describes a model of an intelligent agent that not only builds on experience but also uses creativity to achieve its goal in a three-step process.

³For the same reason as discussed above in section III.1.

- The environment is perceived through sensors and provides the input data (percepts).
- A computational device calculates the output using the projective simulation model.
- The agent acts upon the environment through actuators (actions).

Projective simulation relies on two core elements:

1. Projection Simulator (PS)

The PS allows the agent to project itself into a simulation of a possible future. The agent thereby tests if a certain action leads to a preferable future or if the consequences of this action are not desirable. This gives the agent the opportunity to test outcomes of several different actions before real action is taken.

2. Episodic & Compositional Memory (ECM)

ECM is a special type of memory and consists of a *stochastic* network of clips, which change dynamically as the agent interacts with its environment. The projection simulator uses these clips as building blocks for its simulation.

a) Episodic Memory:

Here episodic memory refers to clips representing previous real experiences of the agent.⁴ The number of clips changes with experiences and the probability distribution of the network is adapted accordingly.

b) Compositional Memory:

Compositional memory describes the combination of clips which allows the agent to create different and new simulations. Its simplest form is the combination of various previous real experiences, which is comparable to associative memory. A more sophisticated form additionally creates new fictitious experiences, which are variations of existing clips but have itself never been experienced before.

The combination of projection simulator and episodic & compositional memory therefore “allows an agent to project itself into fictitious situations, which are self-generated by the agent [...] and which influence its future actions. Projective simulation [...] introduces an elementary notion of creative action.”[9]

⁴This model of episodic memory does not assume encoding of time or dating experiences.

2.b. Relating Projective Simulation to Free Agency

Perhaps the strongest argument that projective simulation enables free agency is the detachment of the agent from its strict causal embedding into the environment. The agent not only modifies past experiences to create new ways of responding to inputs but also simulates their outcomes so that real action is taken only when the outcome is preferable. Briegel argues that projective simulation is possible for sufficiently complex systems and compatible with physical law, which leads him to say:

To put it provocatively, even if human freedom were to be an illusion, humans would still be able, in principle, to build *free robots*. Amusing. [8]

Moreover, as Müller and Briegel [1] point out, projective simulation is an inherently stochastic model, unlike other approaches that assume a deterministic model and add randomness. This difference allows the agent to be reliable and flexible: If a given output repeatedly leads to success, the agent is highly likely to perform the corresponding action, while the agent remains flexible if different considerations lead to different preferred actions. Indeterminism thus shines in a new light under projective simulation: Instead of subjugating the agent to randomness it is the stochastic process that frees the agent from its past and allows it to develop new "thought patterns."

Franklin wrote in "Farewell to the luck (and Mind) argument" in 2011 [7] an argument that perfectly coincides with how projective simulation makes use of indeterminism and how that invalidates the Luck and Mind argument. The key is *when* randomness takes place. While proponents of the Luck and Mind argument place randomness *after* the deliberation process of the agent, Franklin proposes to use randomness only *during* the deliberation process but not directly before an action. Therefore, the agent does not fall victim to randomness but actively uses it as was later shown possible by the model of projective simulation.

IV. Conclusion

Several arguments for and against the free agency in an (in)deterministic world have been considered whereby it became apparent that - as of now - neither physics nor philosophy allow any conclusions on their (in)compatibilities.

From a deterministic viewpoint, there are two different conclusions. Either we refer to Heisenberg's Principle and come to the conclusion that it is not possible to predict the future state of a system from its initial conditions with reasonable accuracy and therefore accept free agency, or we consider a superdeterministic theory as a fundamental theory in which the violation of statistical independence erases any hope for free agency.

If we live in an indeterministic world, the model of projective simulation, as well as the argumentation of Franklin, give reason to believe that free agency is possible. However, both arguments fail to explain how the agent decides if an outcome is preferable, and thus, while the agent might choose the path it takes to reach its goal, the goal itself might still be forced upon the agent. The consequences thereof on free agency (and later on free will) are unclear and remain to be clarified.

At this point, we would also like to remind the reader that this essay only gives a small excerpt of the arguments in this debate and that much more could be said in favor of and against free agency. However, the main point - we do not know the answer yet - would remain the same.

V. References

- [1] Thomas Müller and Hans J. Briegel. “A Stochastic Process Model for Free Agency under Indeterminism”. In: *Dialectica* 72.2 (2018), pp. 219–252. DOI: <https://doi.org/10.1111/1746-8361.12222>.
- [2] Sabine Hossenfelder and Tim Palmer. “Rethinking Superdeterminism”. In: *Frontiers in Physics* 8 (May 2020). DOI: 10.3389/fphy.2020.00139. URL: <https://arxiv.org/pdf/1912.06462.pdf>.
- [3] Gerard 't Hooft. *Free Will in the Theory of Everything*. 2017. DOI: 10.48550/ARXIV.1709.02874. URL: <https://arxiv.org/abs/1709.02874>.
- [4] Peter van Inwagen. “Free Will Remains a Mystery: The Eighth Philosophical Perspectives Lecture”. In: *Philosophical Perspectives* 14 (2000), pp. 1–19. ISSN: 15208583, 17582245. URL: <http://www.jstor.org/stable/2676119> (visited on 07/20/2022).
- [5] Richard E Hobart. “Free will as involving determination and inconceivable without it”. In: *Mind* 43.169 (1934), pp. 1–27.
- [6] Robert Kane. “Responsibility, Luck, and Chance: Reflections on Free Will and Indeterminism”. In: *The Journal of Philosophy* 96.5 (1999), pp. 217–240. ISSN: 0022362X. URL: <http://www.jstor.org/stable/2564666> (visited on 07/20/2022).
- [7] Christopher Evan Franklin. “Farewell to the luck (and Mind) argument”. In: *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition* 156.2 (2011), pp. 199–230. ISSN: 00318116, 15730883. URL: <http://www.jstor.org/stable/41487732> (visited on 07/20/2022).
- [8] Hans J. Briegel. “On creative machines and the physical origins of freedom”. In: *Scientific Reports* 2.1 (July 2012), p. 522. ISSN: 2045-2322. DOI: <https://doi.org/10.1038/srep00522>.
- [9] Hans J. Briegel and Gemma De las Cuevas. “Projective simulation for artificial intelligence”. In: *Scientific Reports* 2.1 (May 2012), p. 400. ISSN: 2045-2322. DOI: <https://doi.org/10.1038/srep00400>.