


A Note on Weakening Free Choice in Quantum Theory

Working Paper

Author(s):

Fourny, Ghislain 

Publication date:

2019-03

Permanent link:

<https://doi.org/10.3929/ethz-b-000334217>

Rights / license:

In Copyright - Non-Commercial Use Permitted

A Note on Weakening Free Choice in Quantum Theory

Ghislain Fourny

March 2019

Abstract The non-extensibility of quantum theory into a non-trivial, non-contextual deterministic theory is based on a strong assumption of free choice, in which the physicists pick a measurement axis independently of the rest of the world. This same strong assumption of free choice is at the core of Nashian game theory. However, recent game-theoretical research based on a weakened version of free choice lead to non-trivial solution concepts with desirable properties.

In this note, we share our view that a similar change of assumption to the modelling of free choice in the foundations of quantum theory opens a non-trivial avenue of research towards a deterministic and non-trivial version of quantum theory with, at least in theory, an improved predictive power.

At this point, this note is a draft that will be regularly updated and completed based on feedback and discussions.

1 Introduction

While there are numerous competing interpretation of quantum physics (Copenhagen, Many-Worlds, ...), most theoretical physicists are aligned on one fundamental assumption: free choice.

Under the assumption that an experimenter freely¹ chooses a measurement axis, several contributions in the field of quantum physics have been made that demonstrate that quantum physics is inherently random, and thus cannot be extended to a fully deterministic theory.

G. Fourny
ETH Zürich
Department of Computer Science
E-mail: ghislain.fourny@inf.ethz.ch

¹ We will make the meaning(s) of free choice precise in this note in subsequent sections

“If indeed there exist any experimenters with a modicum of free will, then elementary particles must have their own share of this valuable commodity.” (Conway and Kochen, 2006)

However, and in spite of evidence in the field of neuroscience in contradiction with a strong account of free choice, there was little research done so far on the consequences of dropping, or weakening, the assumption of free choice in quantum theory.

In this note, we advocate that, beyond intimate convictions, the strong assumption of free choice is not ontologically necessary, and advocate that a different approach in which a weaker version of free choice is assumed can potentially be equally consistent, and also useful. We also sketch what such a more complete theory of quantum physics would then look like in terms of mathematical formalism, based on existing research on the same lines in game theory.

2 Free Choice

The notion of free choice, or free will, is probably as old as philosophy, and there are many ways that it can be defined. We argue here that, for the purpose of this note, two equally reasonable definitions of free choice, strong and weak, can be given.

Free choice is often defined, or thought of, in contrast to the ability to predict was an agent endowed with free choice will decide before they do. This apparent incompatibility between being fully predictable and having free will is embodied in Newcomb’s problem.

As has been argued in Gardner (1973), Newcomb’s problem is to free will what Schrodinger’s cat (Schrödinger, 1935) is to quantum entanglement: it is a thought experiment that takes a seemingly abstract notion and ties it to something tangible².

2.1 Newcomb’s problem

Newcomb’s problem (Gardner, 1973) is typically found under the following form.

An agent is presented with two boxes. One of the boxes is transparent, and it can be seen that it contains \$1,000. The other box is opaque and its contents cannot be seen, but it is known that it is either empty, or contains \$1,000,000.

The agent has the choice between either taking the opaque box, or both boxes. Whichever amounts are inside become hers. But there is one catch.

² Early mentions of these thoughts were made in a paper by Jon Lindsay written in J.-P. Dupuy’s class at Stanford in 1994.

A while before this game took place, somebody predicted the agent's decision, and prepared the contents of the opaque box accordingly: if the predictor predicted that the agent would take one box, he put \$1,000,000 inside the opaque box. If, however, the predictor predicted that the agent would take two boxes, then he put nothing inside the opaque box.

Regarding the skills and accuracy of the prediction, past records of the game with other agents – some of which took one box, some of which took two boxes – are available, showing that the predictor has made 1000 correct predictions out of the 1000 games played so far.

What should the agent do?

2.2 One box or two boxes?

One line of reasoning, which can be qualified of Nashian, is based on a dominant-strategy reasoning: calling x the amount in the opaque box, utility is maximized by picking $1000 + x$ over x , that is, both boxes should be taken. The people reasoning this way are casually named two-boxers. With a correct prediction, they get \$1,000.

Another line of reasoning which makes as much sense is that one box should be taken, leading to \$1,000,000 if the prediction is once again correct. This is because, if two boxes had been taken instead, then the prediction would have been correct too, that is, the predictor would have predicted that two boxes would be taken, and would have put nothing in the opaque box. Had the agent taken two boxes, he would have got only \$1,000 in total, which is less.

The resolution of this apparent paradox (Dupuy, 1992)(Fourny et al, 2018) lies in the modelling of the prediction, more exactly, in the counterfactuals: two-boxers assume that the prediction is correct, but would have been the same (and incorrect) if the agent had made the other choice. One-boxers assume that the prediction is correct and would also have been correct if the agent had made a different decision.

2.3 Two definitions of free choice

The core divergence between the two reasonings is thus whether the prediction *would have been*³ correct, counterfactually. It is precisely this which allows us to distinguish between two fundamental approaches to free choice.

In the two-boxer approach, free choice means that the choice is independent of (or uncorrelated with) anything that does not lie in the future light cone of the decision. This is the strong version of free choice. With this definition of free choice, the past is fixed and (in modal logics terms) necessary, so that any prediction made in the past can be made incorrect by the predicted agent simply by making a decision different than the prediction. There is nothing the

³ The use of the conditional tense is paramount in counterfactual statements.

agent can do at t_2 so that the prediction would have been different at $t_1 < t_2$. (Dupuy, 1992)

In the one-boxer approach, free choice means that the agent could have acted otherwise, but their decision may be correlated to other events not in the future, including in particular the prediction of their decision. This is a weaker, compatibilistic version of free choice. In the actual world, the prediction is correct. In another, hypothetical world, the decision is different but the prediction as well, and it is also correct. The prediction is (in modal logics terms) necessarily correct, but the past is no longer necessary.

A solution to the paradox along the same lines, but based on a statistical framework, was also given by Baltag et al (2009).

2.4 Counterfactuals

The one-boxer approach is often discarded on the grounds that only the first approach is consistent with causal consistency, because an event cannot cause another event that is not in its future light cone. However, as Dupuy (1992) argues, causality is not the only kind of dependency: a dependency between two events may also be due to a correlation, or even a quantum entanglement. In the one-boxer approach, there is a correlation between the decision and its prediction (I pick one box and it has been predicted I would pick one box, but if I had picked two boxes instead, it would have been predicted that I was going to pick two boxes).

Counterfactual implications have been formalized by Lewis (1973) based on lining up alternate world around the actual world with the notion of a distance to the actual world. The counterfactual implication $A > B$ then means that, in the closest world where A is true, B is true as well. It is thus to be distinguished from a logical implication $A \implies B$ equivalent to $\neg A \vee B$, which holds trivially in the actual world if A does not hold, but also from the notion of a necessary implication that would hold in all accessible worlds: $\Box(A \implies B)$.

Coming back to Newcomb's problem, the assertion by a one-boxer that "if he had picked two boxes, the opaque box would be empty" means that, in the closest world in which he picks two boxes, the predictor predicted so and put nothing inside the opaque box. It can thus be seen that there is no causal implication directed to the past in this reasoning. A well known and broadly mentioned example of a non-causal, counterfactual dependency is when there exists a common cause (e.g., in the EPR experiment, the particles are prepared together to have them entangled).

Likewise, the assertion by a two-boxer that "if he had picked one box, the opaque box would be empty as well" means that, in the closest world in which he picks one box, the predictor wrongly predicted he would pick two, and put nothing inside the opaque box.

Whether a counterfactual implication is true or not is thus a matter of assumption on the way possible worlds are organized, depending on whether

they assume strong free choice (the past is the same in all possible worlds and my decision is independent of that past), or a weaker version (the prediction is correct in all worlds, and I could have acted otherwise).

3 Proofs of non-extensibility of quantum theory

Several proofs of the non-extensibility of quantum theory, or put equivalently, on the inherent randomness in quantum theory, are found in literature. These proofs all have in common a fundamental assumption: that the physicist performing a quantum measurement freely chooses the measurement axis.

The underlying notion of free choice is based on a strong assumption that a decision taken freely is independent from the past. This assumption, coupled with the Kochen-Specker theorem (Kochen and Specker, 1967), implies that it is mathematically impossible for the outcome of the quantum measurement to be predicted correctly and consistently *for any choice of the measurement axis*.

Conway and Kochen (2006) established that “if indeed there exist any experimenters with a modicum of free will, then elementary particles must have their own share of this valuable commodity”. The definition of free will, regarding the physicists freely choosing the measurement axis, is formally stated like so: “the choice of directions in which to perform spin 1 experiments is not a function of the information accessible to the experimenters.” Conway and Kochen (2006) conclude that “the free will assumption implies the stronger result, that no theory, whether it extends quantum mechanics or not, can correctly predict the results of future spin experiments.”

Renner and Colbeck (2011) define the strong version of free choice as follows: “our criterion for A to be a free choice is satisfied whenever anything correlated to A could potentially have been caused by A”, which is another way of stating that A is a free choice whenever it is uncorrelated to anything not its future light cone. Formally, in this definition, A as well as the aforementioned “anything” is modelled as a Spacetime Random Variable (SV), which is a random variable with four-dimensional spacetime coordinates. A slightly weaker, non-relativistic version, but still with the same strong idea of independence, only requires for it to be uncorrelated to anything in its past-light cone.

With this strong version of free choice, which is “common in physics, but often only made implicitly” (Renner and Colbeck, 2011), coupled with the assumption that quantum theory is correct, they conclude that “no extension of quantum theory can give more information about the outcomes of future measurements than quantum theory itself.”

In 2018, the Big Bell Test (Abellán et al, 2018) involved a large number of people in order to experimentally perform a Bell Test to further confirm the impossibility to extend quantum theory under the existence of free will. For this experiment, it is assumed that the involved experimenters are endowed with free will in the sense that it a “free variable.”

4 Epistemic omniscience and Kripke semantics

The weakening of the free choice assumption was explored in depth in the field of game theory and rational choice. Dupuy (1992) was the first to point out an analogy between Newcomb’s problem and the Prisoner’s dilemma. Dupuy (2000) then suggested a new approach to rational choice theory based on the notion that the prediction of a rational agent’s decision is correct in all possible worlds. He gave a few examples on a few simple games known as take-or-leave as well as centipede games, and conjectured that under these assumptions, the underlying solution concept, which he called Projected Equilibrium, always exists, is unique, and is always Pareto-optimal. Pareto-optimality is the desirable feature in economics that no other possible outcome of the game gives a better payoff to all players, i.e., that the equilibrium is never suboptimal.

The solution concept was formally defined for all games in extensive form in general position (no ties) in 2004 (Fourny et al, 2018) as the Perfect Prediction Equilibrium and the three conjectures were proven. The key to the reasoning is the use of a forward induction mechanism – in contrast to the Nashian backward induction – that eliminates outcomes one by one until the last one remains. An outcome is eliminated – we also say: preempted – if its own prediction causes a deviation to a different, incompatible subtree. In other words: all outcomes subject to a Grandfather paradox are eliminated. The Perfect Prediction Equilibrium is the only outcome that is immune to its prediction, in the sense that, knowing it in advance, the players play towards this very outcome.

From a logical perspective, the Perfect Prediction Equilibrium is the solution of a fixpoint equation going backward and forward in time, namely, that the outcome must be caused by its prediction, the latter being counterfactually dependent on it.

The main idea underlying perfect prediction is that, if agents can indeed predict each other in all possible worlds, then this induces consistency constraints over what the actual world can look like because of the way counterfactual dependencies interfere with causal dependencies. The actual world must indeed be consistent with both a correct prediction and a consistent timeline.

The twin solution concept for games in normal form, which can be played by spacelike-separated players in separate rooms, was formalized in 2017 (Fourny, 2017) as the Perfectly Transparent Equilibrium. It is based on the iterated elimination of non-individually rational outcomes. It follows the same logics than its extensive-form counterpart. However, even though it is unique and Pareto-optimal as well, it does not always exist.

An epistemic characterization was given (Fourny, 2017) by formalizing the concepts of necessary rationality (or rationality in all possible worlds) and necessary knowledge of strategies (or knowledge of strategies in all possible worlds) into Kripke semantics. The latter can also be referred to as a form of epistemic omniscience, in the sense that the agents know all the events that happen in their world.

There are numerous other papers published in Game Theory with similar, weakening approaches to free choice: for example, Halpern and Pass (2013) researched what happens when agents are translucent, which means that in contrast to perfect prediction, some information leaks but not all of it. Shiffrin et al (2009) has another non-Nashian approach to games in extensive forms, that one could also call translucent. These other approaches, in contrast to the PPE and PTE, consider other possible worlds not to be impossible possible worlds.

5 A fixpoint-based theory of physics with more predictive power

There are a number of intriguing similarities between the above economic framework and quantum theory. First, the notion of possible worlds is present both in Kripke structures, with a set of possible worlds equipped with an accessibility relation, and in quantum theory with the Hilbert space containing the quantum states. In the many-worlds interpretation (Everett, 1973), possible worlds are even explicitly part of the model.

Then, the notion of counterfactuals is at the core to quantum theory: the outcome of a measurement was A, but it could have been B. The cat is alive, but it could have been dead. Quantum theory deeply embeds the notion of unrealized possibles in its mathematical framework. Whether these unrealized possibles are real or not is the subject of intense debates between supporters of the Copenhagen interpretation and of the Many-worlds interpretation (Everett, 1973).

Likewise, quantum theory also has, as its core, counterfactual dependencies in the way envisioned by Lewis (1973): in the EPR experiment, two entangled particles are prepared, for example an electron and a positron, and sent far away to two physicists. The initial state of the joint system is:

$$|\phi\rangle = \frac{|\uparrow\downarrow\rangle + |\downarrow\uparrow\rangle}{\sqrt{2}}$$

The first physicist, Peter, measures the spin of his particle against an axis of his choice, say z. With 50% of probability, he obtains +1 and his half of the system collapses, from his perspective, to:

$$|\phi'_P\rangle = |\uparrow\rangle$$

The entire system has actually collapsed to:

$$|\phi'\rangle = |\uparrow\downarrow\rangle$$

So, assuming they both agreed in advance to measure along the same axis, then Peter knows, with certainty, that Mary measured -1 on her particle:

$$|\phi'_M\rangle = |\downarrow\rangle$$

More importantly, Peter also knows that, had he measured -1 instead, Mary would have measured $+1$ instead. This is a counterfactual statement that follows the laws of quantum theory. Since Peter and Mary can be spacelike separated when they perform their measurement, this is a real-world example of a counterfactual dependency in the absence of any causal dependency.

The possibility of local hidden variables was excluded with the Bell inequality (Bell, 1964). It is more accurate today to speak of Bell inequalities, as there is a large number of them (Brunner et al, 2014). A system modelled with local hidden variables must fulfil a Bell inequality, but actual experiments can break such inequalities and are thus in contradiction with local hidden variables theories. This discards Einstein's local realism, and this is known as Bell's theorem.

Note that some deterministic, non-local hidden variable interpretations of quantum physics are known such as the de Broglie-Bohm theory (Bohm, 1952). Its predictive power is identical to other interpretations, as it essentially factors out the randomness into an unknown initial configuration, separating possible worlds in the same way as Everett's many worlds interpretation.

Later on, the Kochen-Specker theorem (Kochen and Specker, 1967) completed Bell's theorem by weakening its assumptions. This theorem states that it is impossible to assign, in advance of even the choice of measurement axis by the experimenting physicist, a value to each observable for each choice of measurement axis, in a way that is consistent – and this, even if we only require consistency for observables that commute mutually (which is a stronger result).

What we point out here is that it is implicitly assumed, when the Kochen-Specker theorem is used to discard non-contextual hidden variable theories, that the choice of measurement axis by the physicist is unpredictable, i.e., that the physicist is endowed with free will in the strong sense.

But if we weaken this assumption into its one-boxer equivalent, namely, that the physicist *could have picked* a different measurement axis, then we can imagine a setup in which:

- The choice of measurement axis is known in advance (correct prediction)
- Values are only assigned to observables for that known choice of measurement axis, and are an element of reality
- The physicist could have picked a different axis (weak free will)
- The choice of measurement axis would also have been known if the physicist had picked a different axis (substituting fixity of the past for perfect prediction)

With this setup, we can exclude possible worlds in which a different choice of (and correct prediction of) measurement axis would counterfactually lead to an inconsistent world (Grandfather's paradox...). In Kripke semantics, this is known as an impossible possible world (Kripke, 1963)(Rantala, 1982). Eliminating inconsistent worlds in a way similar to game theory (PPE, PTE) can thus narrow down which actual worlds are allowed by the laws of physics,

possibly only one. The theoretical ability to be able to narrow down possible worlds to just one is precisely closing the feedback loop, as envisioned by Dupuy (2000): if we can do so, then we can compute in theory the choice the physicist will make, and even further, we would also have computed the choice the physicist would have made, had it been different.

The notion of epistemic omniscience modelled in (Fourny et al, 2018) can thus directly be put in perspective with an interpretation of quantum theory based on nonlocal hidden variables, in a way that these variables can be calculated as solutions of a fixpoint equation, as was shown to be both feasible, non-trivial, and leading to interesting results, in game theory. This interpretation could span a new class of quantum theories that are essentially augmented Everettian or Bohmian theories, in which the actual world (or part of it) is not contingent, but necessary because of the additional constraints entailed by the postulated predictability of the choice of measurement axis.

Theories in this wider class would have a stronger predictive power than quantum theory in its current state, making themselves falsifiable, so that it may be within our technological reach to design an experimental setup that can confirm or deny them, independently of matters of taste or of personal opinions on the debated topic of free choice. Such experiments would be accessible to us if the hypothetical, global fixpoint equation can be solved partially in certain closed setups.

Either nature fundamentally works that way (weak free will), or it does not (strong free will). It is thus something worth exploring to find out.

Whether this stronger predictive power would be something that we could harness with our current state of technological advancement or if it would remain, in the short to middle term, useless in practice, is, of course, another matter.

6 Acknowledgements

I am first and foremost indebted to Jean-Pierre Dupuy, who laid down the philosophical foundations of projected time, underlying the Perfect Prediction Equilibrium and the Perfectly Transparent Equilibrium. Jean-Pierre pointed out at multiple occasions that there is a strong link between Newcomb's problem and Schrödinger's cat, and that this should be investigated. I am also thankful to Stéphane Reiche, with whom I collaborated on building the algorithmic framework behind the Perfect Prediction Equilibrium.

The general idea presented in this note was mentioned for the first time in a talk on the Perfect Prediction Equilibrium that I gave back in 2009 at a lunch seminar, kindly hosted by Renato Renner from the ETH Zurich institute for theoretical physics, after I visited his Quantum Information course. I am thankful to Renato for various exchanges of emails as well as a few offline discussions on the matter, and to Roger Colbeck and other members of the group for pointing me to related papers in physics and getting me started. I am also indebted to Marcello Ienca for numerous conversations on the neuroscience

aspects of free will. Marcello gave me pointers to many relevant papers on this topic.

I would also like to mention, even though I cannot cite them all, the participants to the Solstice of Foundations, hosted at ETH Zurich in 2017 for the 50th anniversary of the Kochen-Specker theorem with whom I had exciting and motivating discussions. I am more generally thankful to all my colleagues and friends who encourage me to pursue this kind of long-term investigations, sometimes even although they are definite two-boxers.

References

- Abellán C, Aín A, Alarcón A, Alibart O, Andersen CK, Andreoli F, Beckert A, Beduini FA, Bendersky A, Bentivegna M, Bierhorst P, Burchardt D, Cabello A, Cariñe J, Carrasco S, Carvacho G, Cavalcanti D, Chaves R, Cortés-Vega J, Cuevas A, Delgado A, de Riedmatten H, Eichler C, Farrera P, Fuenzalida J, García-Matos M, Garthoff R, Gasparinetti S, Gerrits T, Ghafari-Jouneghani F, Glancy S, Gómez ES, González P, Guan JY, Handsteiner J, Heinsoo J, Heinze G, Hirschmann A, Jiménez O, Kaiser F, Knill E, Knoll LT, Krinner S, Kurpiers P, Larotonda MA, Larsson JÅ, Lenhard A, Li H, Li MH, Lima G, Liu B, Liu Y, López-Grande IH, Lunghi T, Ma X, Magaña-Loaiza OS, Magnard P, Magnoni A, Martínez-Prieto M, Martínez D, Mataloni P, Mattar A, Mazzeo M, Mirin RP, Mitchell MW, Nam S, Oppliger M, Pan JW, Patel RB, Pryde GJ, Rauch D, Redeker K, Rieländer D, Ringbauer M, Roberson T, Rosenfeld W, Salathé Y, Santodonato L, Sauder G, Scheidl T, Schmiegelow CT, Sciarrino F, Seri A, Shalm LK, Shi SC, Slussarenko S, Stevens MJ, Tanzilli S, Toledo F, Tura J, Ursin R, Vargyris P, Verma VB, Walter T, Wallraff A, Wang Z, Weinfurter H, Weston MM, White AG, Wu C, Xavier GB, You L, Yuan X, Zeilinger A, Zhang Q, Zhang W, Zhong J, Collaboration TBBT (2018) Challenging local realism with human choices. *Nature* 557(7704):212–216, DOI 10.1038/s41586-018-0085-3, URL <https://doi.org/10.1038/s41586-018-0085-3>
- Baltag A, Smets S, Zvesper J (2009) Keep “hoping” for rationality: a solution to the backward induction paradox. *Synthese* 169:301–333, URL <http://dx.doi.org/10.1007/s11229-009-9559-z>, 10.1007/s11229-009-9559-z
- Bell J (1964) On the Einstein Podolsky Rosen Paradox. *Physics* 1(3):195–200
- Bohm D (1952) A Suggested Interpretation of the Quantum Theory in Terms of ‘Hidden Variables’ I. *Physical Review*
- Brunner N, Cavalcanti D, Pironio S, Scarini V, Wehner S (2014) Bell nonlocality. *arXiv*
- Conway J, Kochen S (2006) The Free Will Theorem. *Foundations of Physics* 36(10)
- Dupuy JP (1992) Two Temporalities, Two Rationalities: A New Look At Newcomb’s Paradox. *Economics and Cognitive Science*, Elsevier pp 191–220

- Dupuy JP (2000) Philosophical Foundations of a New Concept of Equilibrium in the Social Sciences: Projected Equilibrium. *Philosophical Studies* 100:323–356
- Everett H (1973) The theory of the universal wavefunction. In: DeWitt B, Graham N (eds) *The Many-Worlds Interpretation of Quantum Mechanics*, Princeton UP
- Fourny G (2017) Perfect Prediction in Normal Form: Superrational Thinking Extended to Non-Symmetric Games. arXiv
- Fourny G, Reiche S, Dupuy JP (2018) Perfect Prediction Equilibrium. *The Individual and the Other in Economic Thought: An Introduction*, Routledge pp 209–257
- Gardner M (1973) Free Will Revisited, With a Mind-Bending Prediction Paradox by William Newcomb. *Scientific American* 229
- Halpern JY, Pass R (2013) Game theory with translucent players. arXiv URL <https://arxiv.org/abs/1308.3778>, 1308.3778
- Kochen S, Specker E (1967) The problem of hidden variables in quantum mechanics. *Journal of Mathematics and Mechanics* 17:59–87
- Kripke SA (1963) Semantical considerations on modal logic. *Acta Philosophica Fennica* 16(1963):83–94
- Lewis D (1973) *Counterfactuals*. Harvard University Press
- Rantala V (1982) Impossible world semantics and logical omniscience 35
- Renner R, Colbeck R (2011) No extension of quantum theory can have improved predictive power. *Nat Commun* 2(411)
- Schrödinger E (1935) Die gegenwärtige Situation in der Quantenmechanik. *Naturwissenschaften* 23:807–812, DOI 10.1007/BF01491891
- Shiffrin R, Lee M, Zhang S (2009) Rational Games - ‘Rational’ stable and unique solutions for multiplayer sequential games. Tech. rep., Indiana University