Four Verses from the Dàodéjīng

Jochen Szangolies*

1 Introduction: A Phase Transition

At some point during human evolution, a phase transition must have occurred: from a succession of instinct-driven organisms, a being emerged that was capable conceiving of itself as an individual agent within a larger world, acting within it according to its needs and desires.

Take the sphex wasp. This digger wasp follows a strict ritual: it drags its prey to a point a certain distance from its nest, inspects it, and then drags the prey inside. If, during the inspection, the prey is moved further away, the wasp will proceed to drag it back, and once more inspect the burrow. This cycle may be repeatedly iterated.

Thus, the wasp's behavior seems largely automatic: given a certain set of stimuli ('input'), a certain behavior ('output') results. It apparently lacks the capacity to reflect upon its actions and their consequences within the world, and thus, is in this sense nothing but a line of physical causality briefly made flesh.

Eventually, along the evolutionary line that connects automaton-like beings with modern humans, some predecessor of ours must have started to conceive of itself as separated from the world; as an actor within external circumstances, not as a mere extension of them. At this point, choices open up to our early ancestor: paths to be taken or ignored, aims to be realized, goals to be attained. Deliberation is added to instinct in the determination of behavior.

We see something like this moment repeated within small children: up to about the age of 20-24 months, they do not seem to conceive of themselves as separate entities within the world, for instance failing to recognize themselves in the mirror. Afterwards, however, an understanding develops that allows them to identify this reflection with that unique object in the world that they identify as themselves. It is at this point (at the latest) that the child must have developed the idea of being a self-contained entity within the world that is picked out by the special indexical "I".

In order for this transition to occur, a unique capacity of the human mind is necessary: that of constructing internal models of the world, and of ourselves as distinguished entities within it. In the following, I will argue that the only way we ever interact with the world is through such models—that our perception of the world is ultimately of the models we construct. I will give an account of how this faculty operates, and what its limits are. Finally, I propose that the question "What is fundamental?" only makes sense within any given model, and that we must take care not to confuse our models with the world they (imperfectly, by necessity) reflect.

2 Modeling

A *model* is any system that stands for or reflects certain key attributes of another system (the *object*). More formally, a system $\mathcal{M}$ is a model for an object $\mathcal{O}$ if they obey the relation depicted in Fig. 1 (Rosen, 1991).

Both the model $\mathcal{M}$ and the object $\mathcal{O}$ can take on a variety of forms. First, they may both be physical systems: think of an orrery modeling our solar system, for example. On the other hand, either or both may be abstract structures: a formal mathematical system may be employed to model a natural system,

*jochen.szangolies@gmx.de

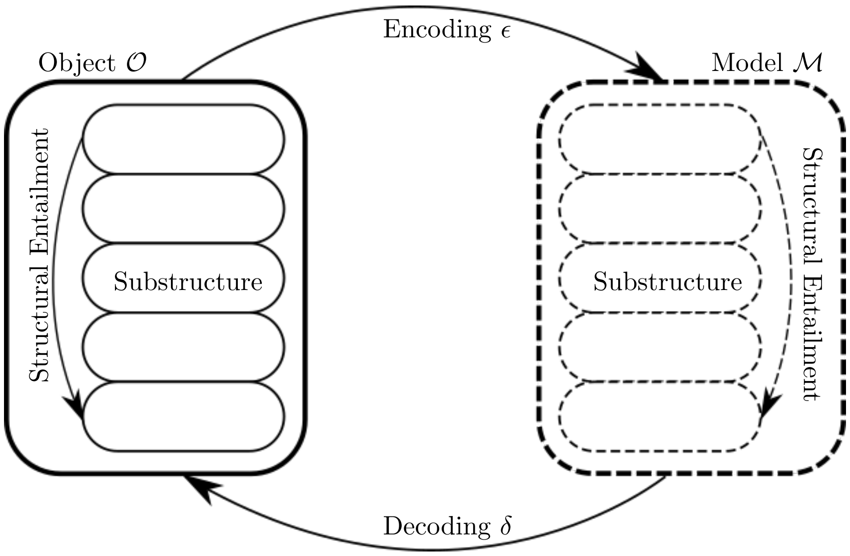

Figure 1: The modeling relation

or, indeed, vice versa—a physical system may be used as a model of a mathematical structure, as happens, e.g., with computation (cf. next section). Finally, and most important to our present considerations, there are mental models, the pictures we have in our minds of the outside world.

The role of models in the mind is to mediate our access to the world. Whenever we perceive, or interact with, parts of the world, what we really perceive is a model constructed from the impressions relayed to the brain via the senses. When we predict the consequences of our actions, we make use of a model in order to rehearse them. Finally, we use models whenever we explain something: an explanation is nothing but a recipe for constructing a mental model that captures relevant features of whatever phenomenon we wish to explain.

Modeling may be iterated: a physical system may be modeled by an abstract formal one, which in turn is then implemented by means of a computer, leading to a simulation of the system. Thanks to this, our ability to use external models is grounded in our ability to use mental models.

This works the same way that a sentence in an unknown language may be understood after having been translated into one we know: we start with the original sentence in, say, Chinese (the model), and then connect it to its meaning (the object) by means of an auxiliary model, a sentence in English connected both to the Chinese sentence and to the same meaning. Analogously, we start with a state of the orrery, and connect this to a state of the solar system via an intermediate state of our mental model of the solar system.

But what, exactly, is it that makes a system into a model of another? As shown in Fig. 1, there are two essential components: first, there must be an encoding/decoding relationship between the two systems, such that the state of the model can be used to support conclusions about the state of the object. Furthermore, there must be a kind of continuity of structure between both systems: what I have called the 'structural entailment' of one system must be mirrored within the other.

The notion of 'structure' to be used here depends on the nature of the system. In a physical system, one may have, for instance, causal structure: the particular way one state implies the following state in temporal succession. Additionally, one may have spatial structure: say, the relative placement of the planets of the solar system, which is mirrored within the orrery in such a way that for each configuration of the orrery, a matching configuration of the solar system exists.

A formal system, on the other hand, possesses a notion of logical entailment. If such a system is thus used to model a physical system, logical derivations represent causal relations, as when one calculates the precise arc a baseball traces out through the sky: just as the initial conditions of the throw cause the ball to land in a certain spot, so do the initial values fed into the right differential equation determine its solutions.

Models are at the heart of our engagement with the world. When we think about a tree, there is no tree present in our thoughts; rather, we use a mental model in order to draw valid conclusions about the actual physical system. This carries an immediate caveat as a corollary: the tree that we perceive, think about, or speak about is not that tree out there in the world, but rather, an element of our model of the world. It may have been this that Lǎozǐ cautioned about (Laozi, 1972):

> The Tao that can be told is not the eternal Tao
> The name that can be named is not the eternal name

We thus must be careful not to confuse our models for the parts of the world they model. In the following, I will argue that the question "What is fundamental?" is subject to just such a confusion—and moreover, that once we clear up the confusion as far as is possible, no sensible question remains. But first, to set the stage, I will introduce an account of modeling that views it as essentially computation, and derive some interesting implications on our ability to model the world.

3 Modeling as Computation

Computers are a very special class of model system defined by their unique versatility. Indeed, one is hard pressed to think of anything that could not be modeled by a computer of sufficient capacity. In analogy to the famous Church-Turing thesis (Church, 1936; Turing, 2004), we have thus grounds to accept the following:

Proposition 1 (Universal model)**.** *Anything that admits of a model at all admits of a computer model.*

More formally, this essentially amounts to a claim that every model can itself be modeled by a system whose internal structural entailment constitutes a computable function. Assuming the Church-Turing thesis, every model then possesses an equivalent model realized in the form of a Turing machine.

Conversely, that at least sometimes, a computation models an object system needs no argument: we model the solar system, the weather, the stock market, and even the entire universe by means of computers running the right sort of program. However, in other situations, this is less obvious: what is a computer running a word processor, or merely the operating system, modeling?

Inasmuch as any computer implements some computable function, we can take that as an answer: it is a model for the abstract structure defined by that function. Thus, we also stipulate:

Proposition 2 (Computation as modeling)**.** *Every computation models some object system.*

Taken together, then, propositions 1 and 2 imply that computation and modeling are in fact co-extensive—modeling is computation, and vice versa.

By their nature, and like the Church-Turing thesis, propositions 1 and 2 are unprovable, and can only be accepted provisionally. Nevertheless, I will argue that accepting them holds great explanatory power, and helps clear up some otherwise mysterious phenomena.

Consequently, I do not propose here to undertake a further defense of the above propositions, and will instead concentrate on their effect on our investigation of the world, and in particular, the attempt to uncover its fundamental nature.

But let us first take stock of some immediate implications. In essence, we have seen that understanding something is equivalent to being able to construct some computational structure that acts as a model of that system; consequently, understanding the motion of the planets ultimately amounts to creating a mental simulation of an orrery. This seems plausible on its face: if we understand some process well enough, we are able to formulate it in terms of a step-by-step description—in other words, an algorithm.

However, we can also deduce some intriguing limitations to modeling itself: take again a look at Fig. 1. So far, we have concerned ourselves exclusively with the systems $\mathcal{M}$ and $\mathcal{O}$. But what about the encoding/decoding mapping that connects them? What provides this translation?

The paradigm example of a system that takes one type of data, performs some manipulation on it, and returns another type is, of course, the computer. Thus, we may suppose that the mapping could be done computationally: taking states of the object $\mathcal{O}$ as input, and returning as output states of the model $\mathcal{M}$. This is an alluring thought, but recall that we have stipulated that computation, itself, is nothing but modeling; thus, a computer implementing this sort of encoding would act as a model for the abstract mapping associating model- and object-states.

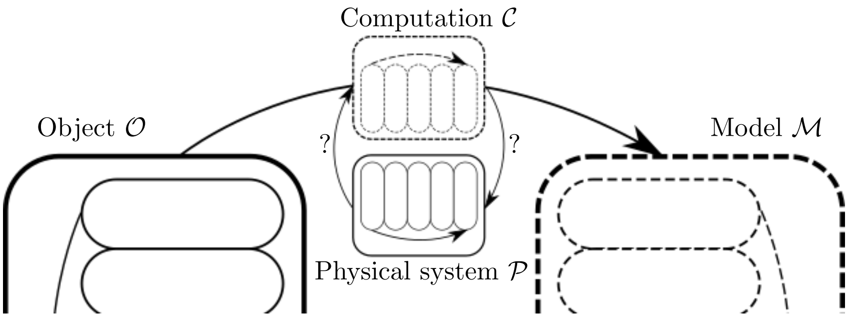

Figure 2: The infinite regress: trying to realize the encoding ϵ via a computation $\mathcal{C}$ raises the question how that computation in turn is implemented (arrows marked by '?').

This, however, will not do. The reason is that we are faced with an infinite regress (see Fig. 2): in order to implement the encoding computationally, we need a physical system $\mathcal{P}$ whose states are mapped to those of some computation $\mathcal{C}$. But this system itself then would be in need of such a mapping, associating states of the computer with the abstract structure of the code map. If we now suppose that this mapping were again to be realized computationally, we are faced with the original problem again.

Interestingly, this sort of self-referentiality also appears in the original text of the *Dàodéjīng*, whose first line, in Pinyin Romanization, is:

> Dào kê dào, fei cháng dào.

Some translations make an attempt to preserve this structure (Laozi, 1913):

> The Reason that can be reasoned is not the eternal Reason.

In our context, we might bastardize this as "modeling that can be modeled is not true modeling".

Due to this vicious regress, it cannot be computation that furnishes the mappings in Fig. 1. This simple insight has startling consequences. For one, since we have found something that cannot be realized computationally, we also cannot model it—thus, to our explanatory abilities, relying as proposed on models, it must be entirely opaque.

Furthermore, if the above is right, no computational theory of the mind can exist—it could grasp, at best, the models themselves, but must necessarily fail when it comes to how they are connected to the objects they model.

We may come to the same conclusion by a different path. We saw above that external models can ultimately be reduced to mental models; that it is our interpretation that, for instance, makes a particular arrangement of beads into a model of the solar system. But then, it must be the same for computation: only by interpreting a given system as modeling a certain abstract structure is that system turned into a computer. Computation is thus a mind-dependent notion, and consequently, any attempt to explain the mind in terms of computation must lead to circularity.

The situation is the same as that of interpreting some string of symbols as meaningful text: to somebody not fluent in its language, the text is meaningless; to somebody using a different mapping between symbols and meanings, it might mean something entirely different. The meaning is not in the text, but in the mind interpreting it. So, too, for computation, with the states of a computational system furnishing the symbols, and the meaning being the computation the system is interpreted as performing.

In the last consequence, we thus see that there are aspects of the world—namely, the connection between (mental) models and their objects—that cannot be modeled; consequently, no model of the world *in toto* can exist.

Few will grant the above conclusion without a good reason to grant first the premises—propositions 1 and 2—on which it rests. Indeed: one may be tempted to see it as a reductio of either.

Accepting these premises, however, confers some striking benefits to our understanding. We have already established that they mandate that the connection between model and object is not amenable to being modeled itself, and thus, seems opaque to reason. Moreover, this entails an inability to communicate these connections: communication occurs if there is a model that can be shared; that is, if we can produce text (such as this) capable of generating in others an understanding of the text's object. Clearly, this entails the necessity to possess a model in the first place.

Consequently, our mental models must be realized by means of a faculty that seems wholly opaque to our understanding, and which we cannot share with others; yet which, in some way, acquaints us with the object at hand, with its properties and characteristics.

These points are shared by the notion of *qualia*, the subjective 'what it's like'-ness of experience (Nagel, 1974). Qualia are generally supposed to be ineffable (Dennett, 1988), that is, incapable of being communicated. Furthermore, they form a direct connection to the qualitative nature of the objects of our experience. Thus, I propose that it is our qualitative (phenomenal) experience that furnishes the necessary connection between our mental models and the objects in the world we experience: by virtue of having the phenomenal experience of an apple, said apple is modeled within our minds.

In the following section, I will take a brief look at how this identification helps solve a number of ordinarily vexing problems in the philosophy of mind. Bolstered by this, I will then derive some consequences for the question "What is fundamental?" from this view of the world.

4 Phenomenology: Binding Object to Model

In the previous section, we have seen that the connections between object and model shown in Fig. 1 are non-computational features of the mind, and consequently cannot be explained using model-based reasoning. Furthermore, we have formulated the bold hypothesis that these connections are given by our phenomenal experience, the qualia: our mental models are connected to objects in the world via our phenomenal experience of them.

Let us now proceed to examine some consequences regarding classical thought experiments in the philosophy of mind that follow from these speculations.

We start with the case of Mary, the color scientist (Jackson, 1986). Mary has the unfortunate fate bestowed upon her that since birth, she has been kept by unknown captors in a black-and-white room, never seeing a single shred of color.

However, she is possessed of considerable scientific acumen, and resumes to unravel every mystery of human color vision. Thanks to her rare talent, soon enough she indeed knows every physical fact that can be known about color vision.

Eventually, through luck or the mercy of her captors, she is released at long last; upon leaving her room, she sees a red rose, experiencing color for the first time. The question is, now, does she learn something new?

If she does—and the intuition that she does seems near impossible to shake—it must follow that the physical facts about color, which she after all possessed, cannot be *all* the facts about color. There must be facts about the world that go 'beyond' the physical in some way.

We now possess the means to attack this conclusion. There are facts pertaining to the connection between model and object that cannot be transported, say, by means of writing or speech. Upon identifying those connections with qualia, with the subjective experience of color, we thus find that Mary could not have learned these facts during her study. There is no model such that knowledge of the physical facts entails (in the sense of the structural entailment in Fig. 1) phenomenal facts pertaining to color vision.

Consequently, it is indeed true that Mary learns something new upon first encountering color. However, in this form, the conclusion poses no threat to naturalist hopes: it indicates merely a certain necessary incompleteness of our models of the world, not a schism within the world itself (say, as between Cartesian *res extensa* and *res cogitans*).

We can proceed similarly with another famous thought experiment aimed at demonstrating the necessity of extra-physical facts: namely, the argument from (philosophical) zombies (Chalmers, 1996), or more generally, the modal argument (Kripke, 1980).

Summed up very briefly, the argument goes as follows: a being physically indistinguishable from you, yet lacking any conscious experience can be conceived; consequently, such a being must be possible in principle. Hence, the physical facts by themselves do not suffice to fix the facts of phenomenal experience.

This may now be countered in much the same way as above. As we have seen, there can be no model explaining the origin of phenomenal facts; but again, this is merely a shortcoming of our models. An analogy may be drawn here to the situation of Gödelian incompleteness (Gödel, 1931): for any formal system F of sufficient strength, there exists a sentence, G, such that G can neither be proven nor disproven from the axioms of F. Consequently, the axioms are consistent both with F's truth, and its falsity.

In a broadly similar way, the phenomenal facts cannot be derived from the physical facts; the latter are consistent both with their presence and absence. However, this does not imply that a world in which all of the physical facts are the same, yet the phenomenal facts are absent (or different, as in the inverted-spectrum thought experiment (Locke, 1700)) is metaphysically possible: after all, since G asserts its own unprovability, we find that it must be true.

In this way, our propositions 1 and 2 thus point towards a clarification of what is often called the *hard problem* of consciousness (Chalmers, 1995): there is indeed no possibility of reducing phenomenal facts to physical facts; however, this is ultimately not a problem, since it is an artifact of how we use models to explain the world, and thus, does not point towards a fundamental difference in kind between the physical and the phenomenal.

This shows the strength of the modeling account as presented above: it allows us to dispel the mysterious nature of consciousness as a mere pseudoproblem. Given these arguments, we should have expected nothing else: if our capacity of modeling is directed at itself, it must fail, on pain of circularity; consequently, the means by which this modeling takes place remains opaque to our reason.

Having thus built up some trust in the resilience of the concepts and ideas developed so far, we may now proceed to turn them to the question of fundamentals.

5 What is Fundamental?

At the moment our first predecessor conceived of itself as a separate, distinguishable entity within a larger world, everything changed. Where previously, there was only the one world acting on itself through various means, now, a second world emerged: the world of models. This, too, with some interpretive license, is anticipated in the *Dàodéjīng*:

> The nameless is the beginning of heaven and earth.
> The named is the mother of ten thousand things.

Once we begin to model (to name) things, the world (heaven and earth) is decomposed into "ten thousand things". Before the world was ever *conceived of as* something, when it just was, there was nothing but the world; afterwards, there are plants and birds and rocks and things, as the song goes.

In more modern terms, information only enters the world as it is being modeled. This can be understood using notions from algorithmic information theory (Chaitin, 1975). Here, the *algorithmic information* of an object is roughly the minimum length of a description needed to specify it. Given a set, or collection, of objects, a subset has the same information content as its complement (i.e. the set of all things not in that subset): that is because a description allowing to pick out all elements of a subset also allows to pick out all elements not in that set—we just take those that are left over.

As an example, take the infinite set of natural numbers: the preceding phrase fully describes it, using vanishingly little information. We can then name a subset of that set, such as that of prime numbers. It takes not much more information—merely a definition of what it means to be 'prime'—in order to specify that set, and with doing so, we have also specified the set of non-primes—all of those numbers not picked out by the definition of 'prime'.

Most sets of natural numbers, however, do not have such a simple description—indeed, on average, their description will not be significantly shorter than merely naming all of the numbers within that set. Thus, somewhat counterintuitively, there may be substructures with a large amount of (algorithmic) information contained within structures of vanishing information.

Consider the world as a set of objects. Call that set 'everything'. Then, 'everything' has no information content at all, since it must have the same information as its complement, i.e. 'nothing'.

Consequently, it is only in specifying a subset of 'everything' that information is introduced into the world. Such a split, however, is an unavoidable feature of every model of the world, since as discussed above, no model can grasp the world as a whole; each merely encompasses some subset.

Thus, with every model of the world, a given minimum of information is associated: this is what is minimally needed to specify that model—a set of axioms, e.g., from which every fact about the system follows. This is what is 'fundamental', according to that model.

The world itself, however, cannot be reduced to such a set of fundamental facts. The salient fact here is the necessary incompleteness of every model of the world: information is introduced by specifying a subset. Picture a uniformly white piece of paper ripped in two: the paper as a whole has very little information, but the rip takes a large amount of information to describe exactly.

It is thus only the necessarily incomplete nature of modeling that introduces 'fundamental' information into the world, by only ever partially capturing it.

6 Conclusion

In the preceding, I have proposed and partially defended the following line of argumentation:

- We apprehend the world—in perception, planning, and explanation—via models.
- Models are essentially computation: every model corresponds to a certain computation, and every computation is a model of a certain abstract structure (e.g. a computable function)—with us thus inhabiting a sort of 'virtual reality'.
- On pain of vicious regress, modeling itself cannot be fully modeled—in particular, hence, the connection between model and object cannot be computational in nature.
- Consequently, this connection is opaque to reason, cannot be communicated, and puts us in contact with objects in the world—properties it shares with qualia, or phenomenal experience.
- If we thus identify the connection between model and object with phenomenal experience, we find that this experience cannot be derived from the physical facts alone, since no model can exist that supports this derivation—thus accounting for the apparent 'gap' between physics and subjective experience.
- Hence, every model of the world as such must necessarily be incomplete.
- Due to their partial nature, every model is associated with a certain minimal information necessary to specify a part of the whole.
- This information is 'fundamental' to the model in the sense that it suffices to completely specify it (take, for instance, a set of axioms defining a formal mathematical system).
- 'Fundamental' information is thus a feature of models, but not of the world they model, that is introduced only by the necessary incompleteness of every model.

If the above is substantially correct, then the question "What is fundamental?" is already misguided: our instinct for searching for the fundamental is simply due to our model-based reasoning. Imagination works via modeling, and consequently, all that we imagine reduces to some fundamental layer, some set of axioms, some base facts from which everything else follows. But this does not mean the world has to follow suit. Indeed, it is by virtue of their necessary incompleteness that every model contains some irreducible fundamental information.

This might at first seem to be a pessimistic conclusion. After all, if the above is right, there is no single unified picture, no grand overarching story to be told of the world such that all the pieces fit—as this would necessitate a model of the world as a whole. Do we thus have to give up the dream of making sense of the world? Is science still possible, if it remains partial?

I believe that there is no cause for alarm. Science, like mathematics after the discovery of the Gödelian incompleteness theorems, will continue just fine. After all, quantum mechanics already tells us that there is no single, unified picture within which to describe the world—wave or particle, momentum- or position-space, each complete set of commuting observables gives a partial picture, and all these pictures are complementary to one another—irreconcilable, and nevertheless indispensable. Indeed, it can be argued that limitations to predictability analogous to the Gödelian limitations are ultimately at the root of quantum complementarity (Szangolies, 2016, Chap. 5).

In fact, these results should provide some measure of relief. After all, any answer to the question "What is fundamental?" seems to immediately invite the followup "Why this?". When we follow Thales in stipulating that "everything is water", we immediately wonder, "but why water?".

No model can reflect the world as it is. In consequence, since all our contact with the world is mediated by models, we are never in contact with the world as it is. Mistaking our models for the world itself—what Whitehead called the 'fallacy of misplaced concreteness' (Whitehead, 1920)—carries dire philosophical consequences: it makes a spooky mystery of conscious experience, and causes us to look for fundamentals that, were they ever found, immediately would become suspect themselves.

But what if, just for a moment, we could remove the glasses through which we view the world? What would the world without models look like? Well, first of all, it would not look like anything: 'looking like' already presupposes the existence of models. Rather, we should be like the sphex again, and enter a state in which we do not conceive of ourselves as separate beings within the world, but instead, would become simple extensions of the world acting on itself.

With again a little poetic license, one might think of such a state as not unlike the Buddhist notion of 'no mind' or 'no self', in which the 'emptiness' of the world—i.e. its lack of any fundamental nature—is manifest. Everything that we perceive then arises in a process of 'dependent origination': just as the shapes of two pieces of torn apart paper, or two subsets of the natural numbers, depend on one another, so do the model and the thing modeled.

The questions we are ordinarily preoccupied with then simply cease to make sense, becoming impossible to even be formulated. Thus, some of the most difficult philosophical problems might call not so much for their solution, as for their dissolution.

References

Chaitin, Gregory J (1975). "A Theory of Program Size formally identical to Information Theory". In: 22.4, pp. 547–569.

Chalmers, David J (1995). "Facing up to the problem of consciousness". In: *Journal of consciousness studies* 2.3, pp. 200–219.

— (1996). *The conscious mind: In search of a fundamental theory.* Oxford: Oxford University Press.

Church, Alonzo (1936). "An Unsolvable Problem of Elementary Number Theory". In: *American Journal of Mathematics* 58.2, pp. 345–363.

Dennett, Daniel C (1988). "Quining qualia". In: *Consciousness in Contemporary Science.* Ed. by Anthony J. Marcel and Edoardo Bisiach. Oxford University Press.

Gödel, Kurt (1931). "Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme I". In: *Monatshefte für mathematik und physik* 38.1, pp. 173–198.

Jackson, Frank (1986). "What Mary didn't know". In: *The Journal of Philosophy* 83.5, pp. 291–295.

Kripke, Saul (1980). *Naming and Necessity.* Harvard: Harvard University Press.

Laozi (1913). *Tao Te Ching.* Trans. by D. T. Suzuki and Paul Carus. La Salle, Illinois: Open Court.

— (1972). *Tao Te Ching.* Trans. by Gia-Fu Feng and Jane English. New York: Vintage Books/Random House.

Locke, John (1700). *An essay concerning human understanding.* Awnsham, John Churchil, at the Black-Swan in Pater-Noster-Row, and Samuel Manship, at the Ship in Cornhill, near the Royal-Exchange.

Nagel, Thomas (1974). "What is it like to be a bat?" In: *The philosophical review* 83.4, pp. 435–450.

Rosen, Robert (1991). *Life Itself: A Comprehensive Inquiry into the Nature, Origin, and Fabrication of Life.* New York: Columbia University Press.

Szangolies, Jochen (2016). "Quantum Correlations: On their Detection, Applications, and Foundations". PhD thesis. Heinrich-Heine-Universität Düsseldorf.

Turing, Alan M. (2004). "Intelligent Machinery". In: *The Essential Turing: Seminal Writings in Computing, Logic, Philosophy, Artificial Intelligence, and Artificial Life.* Ed. by B. Jack Copeland. Oxford: Oxford University Press, pp. 410–432.

Whitehead, Alfred North (1920). *The Concept of Nature.* Cambridge: Cambridge University Press.