16(4), pp.524-578, 2018 **ISSN 1334-4684**

INTERDISCIPLINARY DESCRIPTION OF COMPLEX SYSTEMS

Scientific Journal

Scientific Journal

INTERDISCIPLINARY DESCRIPTION OF COMPLEX SYSTEMS

INDECS, volume 16, issue 4, pages 524-578, year 2018
Published 28th December 2018 in Zagreb, Croatia
Released online 29th December 2018

Office

Croatian Interdisciplinary Society
c/o Faculty of Mechanical Engineering & Naval Architecture
I. Lučića 1, HR – 10 000 Zagreb, Croatia
E-mails: editor@indecs.eu (for journal), ured@idd.hr (for publisher)

Published by *Croatian Interdisciplinary Society* (http://www.idd.hr) quarterly as printed (ISSN 1334-4684) and online (ISSN 1334-4676) edition. Printed by *Redak d.o.o.* (HR) in 50 pieces. Online edition, http://indecs.eu, contains freely available full texts of published articles.

Journal INDECS is financially supported by Croatian Ministry of Science and Education.

Content of the journal INDECS is included in the DOAJ, EBSCO, EconLit, ERIH PLUS, Ulrich's and Web of Science Core Collection.

INDECS publishes original, peer-reviewed, scientific contributions prepared as reviews, regular articles and conference papers, brief and preliminary reports and comments to published articles. Manuscripts are automatically processed with the system Comet, see details here: http://journal.sdewes.org/indecs.

The accessibility of all URLs in the texts was checked one weak before the publishing date.

TABLE OF CONTENTS

LIST OF REFEREES

The following scholars, listed in alphabetic order, refereed manuscripts for the journal INDECS in period from December 2017 to December 2018:

Boris Čulina	Irena Palić
Jelena Ćosić Lesičar	Mirjana Pejić Bach
Mirjana Čižmešija	Benjamin Perasović
Željko Dobrović	Petar Piljek
Nikša Dubreta	Adnan Pitafi
Amalija Horvatić Novak	Kornelije Rabuzin
Denis Kotarski	Noor Sabah
Andrej Kovačić	Vesna Sesar
Mira Krpan	Armano Srbljinović
Ivana Marić	Domagoj Tolić
Blagica Novkovska	Jovana Zoroja

Their contribution to the quality of the Journal's content is acknowledged.

Zagreb, 27th December 2018 Josip Stepanić

INDECS AWARD

Dear authors of articles published in Vol. 15 of the journal INDECS,

the contest for the INDECS award, INDECSA 2018, choosing of the best article published in INDECS during 2018, i.e. in Vol. 16, is opened.

The voters are you, the authors of articles published in INDECS Vol. 15, i.e. in 2017, and the members of the INDECS' Editorial Board. Each and every voter contributes with one vote.

Propositions for the INDECSA are available from the web site of INDECSA, http://indecs.eu/index.php?s=indecsa.

I would like to ask you to give your vote to the article which you consider to be the best among the articles published in the year 2018.

The votes will be collected till 28th February 2019 and the results will be presented in INDECS 17(1).

Cordially,

Zagreb, 27th December 2018 Josip Stepanić

A NOTE ABOUT THE *COGSCI* THEMATIC ISSUE

Dear readers,

articles in this issue of the journal INDECS are developed versions of the selected papers presented at the *Cognitive Science* conference, a part of the 21st International Multiconference *Information Systems*, held in Ljubljana in October 2018. That makes this issue of the journal INDECS a thematic issue devoted to Cognitive Science – *CogSci issue*.

All published articles obtained two affirmative reviews.

Instead of a separate editorial, the overview of the published articles is a part of the content of the first contribution, that of the guest editors for this issue, Prof. Urban Kordeš, Ph.D. and Florian Klauser, Ph.D.

Cordially,

Zagreb, 28th December 2018 Josip Stepanić

Interdisciplinary Description of Complex Systems 16(4), 524-532, 2018

LOOPS AND RECURSIONS IN COGNITIVE SCIENCE: CROSS-ROADS BETWEEN METHODOLOGY AND EPISTEMOLOGY

Florian Klauser* and Urban Kordeš

University of Ljubljana, Faculty of Education, Center for Cognitive Science
Ljubljana, Slovenia

DOI: 10.7906/indecs.16.4.1
Regular article

Received: 19th November 2018.
Accepted: 20th December 2018.

ABSTRACT

This article addresses the need for cognitive science to loop back and examine its roots and presuppositions, pointing out the three recursive issues: 1.) The observer effect or how observing a phenomenon affects the phenomenon that is being observed, an issue that has been acknowledged by natural science, which cognitive science attempts to emulate, and empirical phenomenology, but not cognitive science itself; 2.) Human kinds or how our research affects us, the researchers, and society (people's self-understandings), an issue which forms a loop with the observer effect – observation thus changing the observed, the observer, as well as itself, and 3.) The dangers of over-eager extrapolation or how complexity is lost during shifts in explanatory level, issues pertaining to using findings from studies of one explanatory level (e.g. experiments with rats) to inform a different explanatory level (issues within human society). Finally, the article presents a fourth recursive loop which presents a potential solution to the above: a self-correcting mechanism that allows science to recursively correct its mistakes and improve on its own work.

KEY WORDS

observer effect, human kinds, recursion, looping effects, philosophy of science

CLASSIFICATION

JEL: D83, D84, Z19

*Corresponding author, η: florian.klauser@pef.uni-lj.si; +386 1 589 2200;
Faculty of Education, Kardeljeva pl. 16, 1000 Ljubljana, Slovenia

INTRODUCTION

Upon reaching a dead end, the only way forward is back. One has to retrace one's steps and investigate the path that has led up to that point, so that a more viable way may be found. This is also true in science, and doubly so in cognitive science, where its constituting disciplines present a myriad of branching, joining and again diverging paths. So let us loop back (a phrase that shall become the leitmotif of this article) to the roots of these paths in order to see where we came from and whether there is perhaps a dead end ahead.

The cognitive revolution [1] that gave rise to cognitive science traces its roots to cybernetics – one of the first modern interdisciplinary fields that attempted to study the mind with the computational methods of the rapidly developing fields of computer science and artificial intelligence. It was then that the idea took root to liken the mind to a computer [2], processing information from an input (stimuli) into an appropriate output (behaviour). Some 25 years after the birth of the information-processing (IP) analogy, the young and upcoming field of neuroscience came to fore, uncovering the hardware behind the software that has been (and still is) laid out by psychology, linguistics and anthropology.

There were (and still are), however, problems that provide quite an explanatory challenge if they are to be phrased in an IP framework (but one of many prominent examples of these problems being illustrated by [3], see also [4-6]). While these problems are not exactly dead ends, they are sufficient to evoke doubt in the explanatory power of the IP metaphor – a metaphor that while providing a relatively simple and vivid illustration of some issues, it obfuscates others. Increased awareness of the IP metaphor's (and through it, cognitive science's) limitations and problems gave birth to new explanatory frameworks such as embodied, embedded, extended, enactive or affective cognition (see e.g. [7, 8]).

However, not all issues have been resolved. The emergence of competing theories does not mean a complete transition away from the IP metaphor. Even if it did, these theories do not address all the problems that cognitive science struggles with. Some of these issues shall be the topic of this special issue of INDECS, wherein we shall focus on select methodological and epistemological challenges, such as those of the recursive character of cognitive systems – started by this article and continued by T. Strle [9] – of different ways of modelling and their validity, from computer modelling – see T. Kolenik [10] – to thought experiment – M. Malec [11] – and the challenges the notion of probability poses for metaphysics – P. Lukan [12].

This particular article shall attempt to serve as connecting tissue between these topics, though its main two focuses shall mainly be: the recursive nature of research in cognitive science, and the importance of being mindful of how the paths we take and investigations we perform loop back towards us (the meaning of this shall, hopefully, become clearer throughout this article). Four recursions or loops shall be addressed, the first three representing challenges in research of complex phenomena (i.e., the mind), the last representing a potential solution – or at least the promise of one – to all challenges:

1) The observer effect – how observing a phenomenon affects the phenomenon that is being observed.
2) Human kinds – how our research affects us, the researchers, and society – in the sense of people's self-understandings.
3) From simplification to extrapolation – how complexity is lost during shifts in explanatory level.
4) The self-correcting mechanism: how science recursively corrects its mistakes and improves on its own work.

THE OBSERVER EFFECT

Natural science holds the prestigious position of the human endeavour that uncovers the secrets of the universe ‘as it is’. Aspiring to likewise uncover the secrets of the mind ‘as it is’, cognitive science aims to emulate natural science by imitating its methods. Thus, cognitive science’s methods (in general) tend towards an experimental design, isolating a single (quantifiable) variable and controlling all other parameters, until finally, through rigorous statistical analysis, relationships between variables can be determined. The quantities that constitute the explanatory apparatus of science must be as few and as well defined as possible to still describe the whole system, while the measurement must be such that it disturbs the system in the least possible amount, thus producing a clean, pristine quantity, without disturbing other important parameters. These are (some of the) ideals of psychological and neuroscientific research.

But apparently unbeknownst to the core of scientific community within cognitive science, some odd 100 years ago, physics (the non plus ultra of natural science) has come to the realisation that the idea of measurement as extracting a quantity out of reality is not always viable. Within the realm of quantum physics, when measuring a phenomenon, it does not behave the same way as it would when it is not being measured (see e.g. [13]) – this has been dubbed the observer effect. Moreover, the very physicist who first articulated this realisation, Niels Bohr [14], postulates that the observer effect also applies to measurement within psychology, all the while psychologists cling to assumptions about the nature of measurement that physicists have long let go.

It needs to be conceded that the basic idea of the observer effect is not completely alien to the field of psychology: a similar effect has been noticed and dubbed “demand characteristics” [15]. A participant in a study is aware that they are in an experimental situation and act accordingly: they try to be a ‘good participant’ in that they willingly perform the tasks as they think it is expected of them. These perceived expectations and the behaviour they elicit differs from participant to participant [15]. Perhaps the most illustrative and recent study showing the consequences of demand characteristics is the one by R. Hurlburt et al. [16], where fMRI images were taken of people during inner speech, with there being difference in brain activations between spontaneous occurrences of inner speech and on-demand performances. Yet studies such as this one are an exception. Demand characteristics are not often acknowledged in psychological or neuroscientific research, and if they are, they are a thing to be minimized [17].

There is, however, a field of research within cognitive science that pays greater attention to the difference between measured and unmeasured phenomena, the young and burgeoning field of experience research, also known as empirical phenomenology. Besides collaborating with other fields to reveal the effects of demand characteristics (as it did in the previous illustration thereof, Hurlburt et al., 2016), it also acknowledges the observer effect within its own field of experience research, speaking of an “excavation fallacy” [18], which describes the issue that exploring experience deforms, distorts or even wholly creates the experience one purports to examine.

How, then, do we deal with the observer effect and the various forms it takes throughout cognitive science? It is yet unclear how (or even whether) it would be possible to be rid of the effect altogether, but ignoring it or trying to minimise the effect are both approaches of questionable potency. Empirical phenomenology seems so far to be the only field (besides quantum mechanics itself) to have properly acknowledged the seriousness of this issue and formulated a response (cf. [19]). U. Kordeš and E. Demšar [20] suggest that the excavation fallacy should be rephrased as an excavation characteristic – meaning that the effect becomes itself something to be studied. We have to leave it to our fourth loop to apply this same approach to the other fields within cognitive science, but as the example of Hurlburt et al. [16] indicates, baby steps are being made.

HUMAN KINDS

An effect similar to the observer effect can be detected between science and society – a relationship that can be caricaturised as, again, between observer and the observed, but this time focused on how observation changes the observer themselves. To better understand this interplay Ian Hacking [21] provides the terms natural and human kinds.

Natural kinds are the typically well defined, context independent, and not (too) interconnected concepts or kinds usually found in the natural sciences – such as gravity, force, electric charge, etc. Human kinds, on the other hand, tend to be understood only within a specific context, appear in groups with interwoven meanings, with shifting or flexible definitions – e.g. virtue, morality, love, decision-making, etc. Our definitions and descriptions of human kinds are influenced by how we experience them, which is reciprocally affected by the definitions and descriptions of the greater societal discourse we are contributing to. Human kinds are found in every-day discourse and the humanities, as well as, despite its efforts, in psychology and cognitive science.

To reiterate: many of the phenomena researched by cognitive science are intrinsically dependent on our intuitions[1], all the while our use of those terms shapes the very intuitions the definitions of those phenomena depend on. The relationship between research, resulting constructs, those constructs' effect on social sense-making, and, in turn, social sense-making's effect on research, is not yet sufficiently understood (though attempts have been made, see e.g. [9, 22]).

The previously described issues concerning the observer effect and human kinds come full circle: measurement (or observation) is not a simple extraction of data, but an act that itself changes the phenomenon being observed. Similarly, the observation's results shape how the phenomenon is understood in the greater societal discourse, which loops back, affecting the observer's intuitions, which shape their observation. To simplify these two loops into one: observation (measurement, research) changes both the observed and the observer (Figure 1).

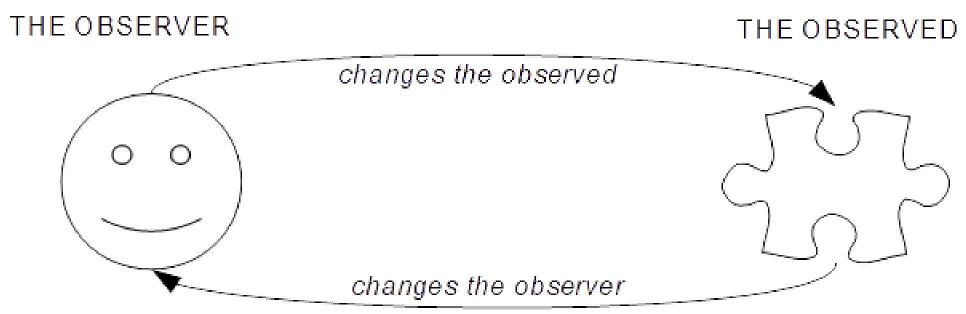

Figure 1. The recursive nature of observation.

The recursive and self-referential nature of research makes a simple transfer of methodologies found in natural science to the fields of research on the mind exceptionally difficult. Even more so as different fields deal in different levels of explanation, such as molecules, cells, individuals, societies and cultures. Methodologies employed to study molecules are questionably effective for studying individuals thus the findings from, for example thermodynamics cannot easily be translated into how the mind works (though respectable attempts have been made, cf. [23]).

FROM SIMPLIFICATION TO EXTRAPOLATION

In trying to study the mind by means of natural science means to transfer information between different levels of explanation and complexity, distorting information along the way. The transfer goes two ways, which shall here be described as ‘simplification’[2] (from a ‘higher’ level to a ‘lower’ one) and ‘extrapolation’ (vice versa). Studying the mind usually involves shifts in both directions: first simplification, then extrapolation.

As mentioned before, the most ‘prestigious’ way of research, which cognitive science is striving towards, is that of natural science. So the first step of researching a given phenomenon is to ‘simplify’ it, distilling it into few, easy to measure (quantifiable) variables. However, while, for example, simplifying water temperature into degrees Celsius is rather straight forward, the same cannot be said for complex cognitive phenomena such as decision making, empathy or morality.

Similar concerns have already been raised in biology under the name of epistemological reduction [24], questioning whether, for example, all of evolution can be explained solely by genes. Our concerns regarding simplification take a similar direction, perhaps going even a bit further. We do not only question whether phenomena such as empathy can be sufficiently explained by, for example, activations of certain neurons, but what (information) is lost when one distils or simplifies empathy into neuronal activity? The question is thus: how does one even go about transforming complex and ambiguous phenomena into simple, well-defined variables? We would argue that it is impossible to do so without losing a significant amount of nuance and complexity – which in turn makes the transfer back from the lower level of explanation onto a higher one (extrapolation) problematic, as one might have lost something integral in the process (an idea not dissimilar to incommensurability [25]). Yet the extrapolative jump to conclusions is well employed in the social system of science: results from studies are framed as relevant in the broader societal context, which changes how society understands itself, how individuals understand themselves, and how some of those individuals – the scientists – understand the phenomena they are researching. From observing how neurons that fire when a monkey eats a banana also fire when the monkey sees another monkey eat a banana [26], we (humanity as a whole) are quick to posit ‘answers’ to ‘big’ questions of cognition such as empathy or even love.

An illustrative example of the dangers of over-eager extrapolation is the now famous experiment with rats whose brains had been wired to a lever so that, upon pressing the lever, their reward centres were stimulated [27]. Once the rig had been figured out, the rats would press the lever continuously, thousands of times per hour, day after day. In similar studies, it has been shown that these more direct forms of stimulation (intravenous cocaine or direct electrical brain stimulation) were preferred even over food, resulting in the rats’ eventual starvation [28].

These findings were kindling for the fires of reckless explanation. Their conclusions were immediately applied to humans: that our other survival instincts bow to the tyrannical rule of our rewards system. There is no hope for wholesome living – for as long as drugs are available, addiction and ruin are inevitable. Such have we divined from rats.

However, the cogs of science ground on. A subsequent study [29] introduced a control group. While one group of rats was kept, just like in the previous studies, in a small cage, the other group had a pen with numerous toys, ample food and lots of room for playing and mating – a so-called ‘rat park’. Both groups started off addicted to morphine and given a free choice of regular water or morphine solution. As by then expected, the caged rats gorged themselves on the drug. But, surprisingly, the park rats seemed to prefer regular water [29], indicating that addiction might be stymied by a sufficiently stimulating environment. The previous extrapolation that the rewards system subjugates all other instincts was thus challenged, thus also pointing out how problematic it is to infer complex environmentally dependent cognitive dynamics from studies accommodating necessarily reductive and 'simplifying' methodology (which tries to negate environmental impact as much as possible).

The here elucidated issue with simplification and (reckless) extrapolation is concerning for the whole of cognitive science. Is the cognition that is researched in the lab the same one that occurs 'in the wild'? If (cognitive) science needs to distil every phenomenon into a measurable variable, can it still be said that it is researching that very phenomenon, or is it but studying a homeopathic solution of questionable potency and ecological validity[3]?

THE SELF-CORRECTING MECHANISM

The heretofore presented issues paint a grim and pessimistic picture. It seems that whatever step we take, however far we loop back, we are surrounded by dead ends and every time we try to move to a different road it turns out to be on a different (explanatory) level. Is there really no hope to someday find our way to the truth about the mind?

It would be only fair to concede at this point that the hitherto used metaphor of branching paths and stepping forward is, though illustrative, somewhat deceiving as far as the nature of scientific progress is concerned. Science is not in the business of discovering reality, but in negotiating with the environment, or as Bertrand Russell [30; p.15] puts it:

> "Science thus encourages the abandonment of the search for absolute truth, and the substitution of what may be called "technical" truth, which belongs to any theory that can be successfully employed in inventions or in predicting the future. "Technical" truth is a matter of degree: a theory from which more successful inventions and predictions spring is truer than one which gives rise to fewer. "Knowledge" ceases to be a mental mirror of the universe, and becomes merely a practical tool in the manipulation of matter."

If we subject the whole endeavour of science to this criterion of "technical" truth – that is, whether its theories can be successfully employed – we find ourselves before a staggering mass of evidence for its efficacy. How then do we console our analyses of its numerous flaws with its resounding success? How can it be riddled with mistakes and still grind on, spewing out better and better theories? The answer is exactly thus: it continues to spew out better and better theories, all the while looping back, correcting past mistakes and improving or discarding and substituting past theories that no longer appear to hold water.

This is science's self-correcting mechanism: a culture of falsification [31], replicability and peer-review – a culture of testing, retesting and gathering of new data to support or disprove. R. Feynman characterised science as "belief in the ignorance of experts" [32; p.315]. All this is done in an attempt to cover all the blind spots, account for biases and rectify past mistakes. The institutionalized self-correcting process was already evident in our previous example with rats: scientists put the work of their peers under close scrutiny, found shortcomings, and developed a better experiment. The new experiment in turn was put under close scrutiny that revealed shortcomings (cf. [33, 34]), and the cycle of re-examination continues.

Through this very same process, cognitive science – even from way before it was called that – evolved from a sole interest in the relationship between stimulus and behaviour (behaviourism), to an interest in the brain and how it processes information (information-processing metaphor) and now to tendencies towards embodied and social aspects of the mind. Finally, in the past few decades, the phenomenological or first-person perspective is being taken into account as well (for example, in the form of enactivism, see [7, 35]).

The self-correcting mechanism is perhaps science’s strongest virtue. It allows science to constantly loop back and improve itself in the face of new data, to question perhaps out-dated metaphors, to re-evaluate the very nature of measurement and observation, to become aware of its own effect on society and society’s effect on scientific work, and of the dangers of shifting between different levels of explanation. What is more, this mechanism built into the scientific process itself, so change need not be brought about by outside intervention. To overthrow a misruling despot within science – perhaps in the form of a common metaphor or a leading theory – there is no need for an armed militia to instigate a violent revolution. Science constantly revolutionises itself. The “governance” of a leading theory is thoroughly and ceaselessly being questioned by the very proponents of that theory and their peers. Thus, we invite the reader to do just that. Peruse this text (and this issue, this journal, other journals, …) and scrutinize the contents, be mindful of the problems described (t)herein in so far as they apply to your work, and do not hesitate to call us out on what we ourselves have missed. The descent into placid naivety can, after all, be ever so seductive.

REMARKS

[1]A prominent example of our intuitions shaping our research is found in computer modeling. The issue known as Pac-Man effect [36] brings to attention that, to simplify, it is always the one making the computer model that determines what is food that should be sought out and what are predators that should be avoided, not the agent itself (see also [10]).

[2]That is not to say that phenomena and their explanations are simple in the colloquial sense (anyone trying to wrap their head around quantum mechanics can attest to that), just that less “other factors” (such as the environment) need to be accounted for or ignored when researching them.

[3]See also Strle [37] for similar claims in the domain of decision-making.

REFERENCES

[1] Gardner, H.: *The mind’s new science: A history of the cognitive revolution*. Basic Books, New York, 1987,

[2] Searle, J.R.: *Cognitive Science and the Computer Metaphor*. In: Negrotti, M., ed.: *Understanding the Artificial: On the Future Shape of Artificial Intelligence*. Artificial Intelligence and Society. Springer, London, 1991, http://dx.doi.org/10.1007/978-1-4471-1776-6_8,

[3] Held, R. and Hein A.: *Movement-produced stimulation in the development of visually guided behavior*. Journal of Comparative and Physiological Psychology **56**(5), 872-876, 1963, http://dx.doi.org/10.1037/h0040546,

[4] Hardcastle, V.G.: *A critique of information processing theories of consciousness*. Minds and Machines **5**(1), 89-107, 1995, http://dx.doi.org/10.1007/BF00974191,

[5] Norman, D.A.: *Twelve issues for cognitive science*. Cognitive Science **4**(1), 1-32, 1980, http://dx.doi.org/10.1207/s15516709cog0401_1,

[6] Thagard, P.: *Cognitive Science*. In: Zalta, E.N., ed.: *The Stanford Encyclopedia of Philosophy*. Winter 2018 Edition, 2018, https://plato.stanford.edu/archives/win2018/entries/cognitive-science,

[7] Varela, F.J.; Thompson, E. and Rosch, E.: *The Embodied Mind: Cognitive Science and Human Experience*. MIT Press , Cambridge, 1991,

[8] Damasio, A.R.: *The somatic marker hypothesis and the possible functions of the prefrontal cortex*.
Philosophical Transactions of the Royal Society of London B: Biological Sciences, **351**(1346), 1413-1420, 1996,
http://dx.doi.org/10.1098/rstb.1996.0125,

[9] Strle, T.: *Looping Minds: How Cognitive Science Exerts Influence on Its Findings*.
Interdisciplinary Description of Complex Systems **16**(4), 533-544, 2018,
http://dx.doi.org/10.7906.indecs.16.4.2,

[10] Kolenik, T.: *Seeking after the Glitter of Intelligence in the Base Metal of Computing: The Scope and Limits of Computational Models in Researching Cognitive Phenomena*.
Interdisciplinary Description of Complex Systems **16**(4), 545-557, 2018,
http://dx.doi.org/10.7906.indecs.16.4.3,

[11] Malec, M.: *Considerations on Scientific Thought Experiments*.
Interdisciplinary Description of Complex Systems **16**(4), 558-566, 2018,
http://dx.doi.org/10.7906.indecs.16.4.4,

[12] Lukan, P.: *Probability and Metaphysics*.
Interdisciplinary Description of Complex Systems **16**(4), 567-578, 2018,
http://dx.doi.org/10.7906.indecs.16.4.5,

[13] Buks, E.; Schuster, R.; Heiblum, M.; Mahalu, D. and Umansky, V.: *Dephasing in electron interference by a 'which-path' detector*.
Nature **391**(6670), 871-875, 1998,
http://dx.doi.org/10.1038/36057,

[14] Bohr, N.: *Atomic Theory and the Description of Nature*.
Cambridge Universtity Press, Cambridge, 1934,

[15] Orne, M.T.: *On the social psychology of the psychological experiment: With particular reference to demand characteristics and their implications*.
American Psychologist **17**(11), 776-786, 1962,
http://dx.doi.org/10.1037/h0043424,

[16] Hurlburt, R.T.; Alderson-Day, B.; Kühn, S. and Fernyhough, C.: *Exploring the ecological validity of thinking on demand: neural correlates of elicited vs. spontaneously occurring inner speech*.
PLoS ONE **11**(2), e0147932, 2016,
http://dx.doi.org/10.1371/journal.pone.0147932,

[17] Roller, M.R. and Lavrakas, P.J.: *Applied qualitative research design: A total quality framework approach*.
Guilford Publications, New York, 2015,

[18] Depraz, N.; Varela, F.J. and Vermersch, P.: *On becoming aware*.
John Benjamins, Amsterdam, 2003,
http://dx.doi.org/10.1075/aicr.43,

[19] Kordeš, U.: *A better metaphor for understanding consciousness?*
Interdisciplinary Description of Complex Systems **13**(4), 525-533, 2015,
http://dx.doi.org/10.7906/indecs.13.4.4,

[20] Kordeš, U. and Demšar, E.: *Excavating Belief about Past Experience: Experiential Dynamics of the Reflective Act*.
Constructivist Foundations **13**(2), 219-229, 2018,

[21] Hacking, I.: *The looping effects of human kinds*.
In: Sperber, D.; Premack, D. and James Premack, A., eds.: *Causal cognition: A multidisciplinary debate*. Oxford University Press, New York, pp.351-394, 1995,

[22] Strle, T. and Markič, O.: *Looping effects of neurolaw, and the precarious marriage between neuroscience and the law*.
Balkan Journal of Philosophy **10**(1), 17-26, 2018,
http://dx.doi.org/10.5840/bjp20181013,

[23] Friston, K.J. and Stephan, K.E.: *Free-energy and the brain*.
Synthese **159**(3), 417-458, 2007,
http://dx.doi.org/10.1007/s11229-007-9237-y,

[24] Brigandt, I. and Love, A.: *Reductionism in Biology*.
In: Zalta, E.N., ed.: The Stanford Encyclopedia of Philosophy. Spring 2017 Edition, 2017,
https://plato.stanford.edu/archives/spr2017/entries/reduction-biology,

[25] Oberheim, E. and Hoyningen-Huene, P.: *The Incommensurability of Scientific Theories*.
In: Zalta, E.N., ed.: *The Stanford Encyclopedia of Philosophy*. Fall 2018 Edition, 2018,
https://plato.stanford.edu/archives/fall2018/entries/incommensurability,

[26] Di Pellegrino, G.; Fadiga, L.; Fogassi, L.; Gallese, V. and Rizzolatti, G.: *Understanding motor events: a neurophysiological study*.
Experimental Brain Research **91**(1), 176-180, 1992,
http://dx.doi.org/10.1007/BF00230027,

[27] Olds, J.: *Self-Stimulation of the Brain; its use to study local effects of hunger, sec and drugs*.
Science **127**(3294), 315-324, 1958,
http://dx.doi.org/10.1126/science.127.3294.315,

[28] Wise, R.A.: *Brain reward circuitry: insights from unsensed incentive*s.
Neuron **36**(2), 229-240, 2002,
http://dx.doi.org/10.1016/S0896-6273(02)00965-0,

[29] Alexander, B.K.; Coambs, R.B. and Hadaway, P.F.: *The effect of housing and gender on morphine self-administration in rats*.
Psychopharmacology **58**(2), 175-179, 1978,
http://dx.doi.org/10.1007/BF00426903,

[30] Russell, B.: *Religion and science*.
Oxford University Press, 1997,

[31] Popper, K.: *The logic of scientific discovery*.
Routledge, London, 2005,
http://dx.doi.org/10.4324/9780203994627,

[32] Feynman, R.: *What is science?*
The Physics Teacher **7**(6), 313-320, 1969,
http://dx.doi.org/10.1119/1.2351388,

[33] Bozarth, M.A.; Murray, A. and Wise, R.A.: *Influence of housing conditions on the acquisition of intravenous heroin and cocaine self-administration in rats*.
Pharmacology Biochemistry and Behavior **33**(4), 903-907, 1989,
http://dx.doi.org/10.1016/0091-3057(89)90490-5,

[34] Petrie, B.F.: *Environment is not the most important variable in determining oral morphine consumption in Wistar rats*.
Psychological Reports **78**(2), 391-400, 1996,
http://dx.doi.org/10.2466/pr0.1996.78.2.391,

[35] Di Paolo, E.A.; Rohde, M. and De Jaegher, H.: *Horizons for the enactive mind: Values, social interaction, and play*.
In: Stewart, J.; Gapenne, O. and Di Paolo, E., eds.: *Enaction: towards a new paradigm for cognitive science*. MIT Press, Cambridge, 2007,

[36] Riegler, A.: *Towards a Radical Constructivist Understanding of Science*.
Foundations of Science **6**(1-3), 1-30, 2001,
http://dx.doi.org/10.1023/A:1011305022115,

[37] Strle, T.: *Embodied, Enacted and Experienced Decision-Making*.
Phainomena **XXV**(98-99), 83-107, 2016.

Interdisciplinary Description of Complex Systems 16(4), 533-544, 2018

LOOPING MINDS: HOW COGNITIVE SCIENCE EXERTS INFLUENCE ON ITS FINDINGS

Toma Strle*

University of Ljubljana, Faculty of Education
Ljubljana, Slovenia

DOI: 10.7906/indecs.16.4.2
Regular article

Received: 20th November 2018.
Accepted: 23rd December 2018.

ABSTRACT

Drawing on the distinction between natural and human kinds, I will discuss the looping effects of human kinds through the lens of contemporary cognitive (neuro)science. I will try to show that cognitive science is mainly in the business of investigating, understanding and explaining human kinds. As new conceptualisations of the human mind, agency and our nature are being created (by, for instance, neuroscience), they open up the possibility for new, different understandings of what it means to be a human being. This, I will argue, can change how people think and behave and thus change the very phenomena cognitive science investigates. Consequently, cognitive science can affect its very (future) findings. This holds especially true when society embraces new conceptualisations of the human mind and new ways of self-understanding become part and parcel of social discourse, activities, and/or structures.

The quest for understanding the human mind, I will claim, is a looping journey, where what we "discover" about the human mind is inherently dependent on how we, as human beings, understand ourselves; and how we understand ourselves is, to a certain degree, dependent on how science understands us and on how we interpret what it has to say about our nature. At the end of the article, this will lead me to consider cognitive science as an intrinsically ethical endeavour.

KEY WORDS

agency, cognitive neuroscience, human kinds, objectivity, self-referentiality, sociality

CLASSIFICATION

APA: 2340, 2520, 2630, 3040

JEL: D91

*Corresponding author, η: toma.strle@pef.uni-lj.si; +386 1 5892 200;
University of Ljubljana, Faculty of Education, Kardeljeva ploščad 16, 1000 Ljubljana, Slovenia

INTRODUCTION

In the article, I will discuss the looping effects of human kinds through the lens of contemporary cognitive (neuro)science. Drawing on Hacking's distinction between natural and human kinds [1, 2], I will try to show that psychology and cognitive (neuro)science are mainly in the business of investigating, understanding and explaining human, and not natural kinds. That is kinds that are dependent on what beliefs we hold about them, what descriptions we apply to them and what kind of actions we perform according to our understanding of them – kinds that are affected by and change according to our self-understandings.

As new conceptualisations of the human mind, agency and our nature are being created by science, this opens up the possibility for new, different understandings of what it means to be a human being and thus change how we think and act. This holds especially true when society embraces new conceptualisations of the human mind on a larger scale and new ways of self-understanding become part and parcel of social discourse, activities, and/or structures. And if we agree with the claim that cognitive science's outlook on the human mind – be it neuro-reductionist or enactivist – can affect how people think, decide, exert self-control, behave in moral situations, etc. and (by definition) agree that thinking, decision-making, self-control, moral behaviour, etc. are, among others, cognitive (neuro)science's intended "objects" of investigation, we should accept the claim that cognitive (neuro)science has an impact on the very phenomena it investigates and thus influences its very (future) findings. Understanding the human mind is a *looping journey*; what we "discover" about the human mind is inherently dependent on how we, as human beings, understand ourselves; and how we understand ourselves is, to a certain degree, dependent on how science understands us and how we interpret what it implies about our nature.

Such looping effects are quite foreign to sciences that strive to be as objective as possible, independent of observers' – researcher's or subject's – perspectives and attitudes towards the phenomenon under investigation, and value-free in their account of the human mind and *nature*. Objectivity, as succinctly defined by Reiss and Sprenger, "expresses the idea that the claims, methods and results of science are not, or should not be influenced by particular perspectives, value commitments, community bias or personal interests, to name a few relevant factors." [3] Objectivity, at least in this regard, is unattainable for cognitive (neuro)science; for, the quest for understanding the human mind is inherently a looping journey. The looping nature of understanding the human mind is, however, not something we could explain away as some sort of (neural?) noise, or remedy by, for instance, better methodology or statistical analysis and utter "Good riddance!".

At the end of the article, the presented view on the human mind will also steer me to consider cognitive science more as an ethical endeavour and not as an endeavour that should reveal what the human mind *really is*.

NATURAL AND HUMAN KINDS, AND THE LOOPING EFFECTS OF HUMAN KINDS

In one of his articles, Hacking [2] (see also [1, 4]) discusses the looping effects of human kinds – how entities ("objects" and categories) studied by psychology exert an influence on themselves. But before I delve into the matter, let me first explain the distinction between human and natural kinds.

According to Bird and Tobin [5], natural kind "corresponds to a grouping that reflects the structure of the natural world rather than the interests and actions of human beings." Crucially, natural kinds can hence be characterised by "intelligibility outside discursive contexts; indifference to the descriptions applied to them; and independence of categories and kinds.” [6; p.773] According to Hacking [2] and Brinkmann [6, 7], examples of natural kinds are quarks, sunsets, the common cold, mud, trees, tigers, water, and gold, to name a few. Take water for instance. Water has several properties: it is tasteless, colourless and boils under specific conditions. But physics and chemistry, in their understanding of what water *is*, do not stop at these properties. They aim at *discovering* the *essential properties* of the phenomenon, which, in turn, also enables them to explain the more superficial properties. In the case of water, its essential property is its molecular structure, not its transparency or colourlessness. And, even though we incorrectly *perceived* these superficial properties – if it, for instance, turned out that our sense of taste is not that accurate and we wrongly perceived water as tasteless –this would not change the *fact* that we call this liquid water (at least in physics and chemistry). The reason being that water’s basic molecular structure remains the same, irrespective of how we perceive it (see also [8]). The aim of natural sciences is thus in this (a little simplified) regard, “to find natural kinds, because, if they succeed in doing so, they can get an overview of the common essential properties of a class of things that allow them to explain other, more superficial, properties of this class of things.” [7; p.1221].

The question that Hacking [1, 2, 4], Brinkmann [6, 7] and others, e.g., [9], pose, is whether psychology is about natural or some other sort of kinds. Do psychology and cognitive science study the same sort of entities (categories, phenomena) as physics, or chemistry? Are their intended objects of investigation – as many would like to believe – independent of particular attitudes, individual perspectives, values, and description applied to them, as is the case in natural sciences? (But confer Kordeš [10] for the discussion on similarities between the study of consciousness and quantum physics.) Some scholars (e.g., Brinkmann and Danziger) would answer with a definite no. Psychology – I will extend some of their claims to cognitive science – does not, in general, study natural but human kinds.

Human kinds can be characterised by “intelligibility only within a discursive context; interaction with the descriptions applied to them; and emergence together of categories and kinds.” [6; p.773]. Human kinds, according to Hacking, are “(i) kinds that are relevant to some of us, (ii), kinds that primarily sort people, their actions, and behaviour, and (iii) kinds that are studied in the human and social sciences, i.e. kinds about which we hope to have knowledge. I add (iv) that kinds of people are paramount; I want to include kinds of human behaviour, action, tendency, etc. only when they are projected to form the idea of a kind of person.” [2; p.354] For Hacking [2], examples of human kinds are homosexuality, multiple-personality-disorder, suicide, teen-age-pregnancy, adolescence, etc. For Hacking, one of the fundamental characteristics of human kinds is their role in classifying and sorting people into certain categories, or kinds of people (see his four-point definition above). However, I will extend the notion of human kinds to aspects of cognition that are “in the minds of people”; to aspects of the human mind that are paramount to people in their self-understanding (e.g., decision-making, self-control, volition, moral agency, mental disorders, mindfulness meditation, etc.), irrespective of whether people are classified as certain kinds because of being understood in this or that way. (See also Thompson’s intriguing article *Looping Effects and the Cognitive Science of Mindfulness Meditation* [11].)

By contrasting natural and human kinds, neither Hacking nor I want to state that human kinds are not natural or part of nature. Human kinds are as real as natural kinds. It is just that they constitute a different “reality” – what they are, how they behave, and what we “discover” about them is partly defined by what beliefs we hold about them, how we characterise them, what we do according to how we understand them, etc. As Brinkmann (paraphrasing A. MacIntyre) crisply puts it: “It makes no difference to water what we call it, but it does make a difference to human kinds. After all, as A. MacIntyre once remarked, molecules do not read chemistry textbooks, whereas humans do read psychology books that affect their self-understandings” [6; p.775].

A very similar attitude towards psychology is held by Danziger in his discussion of what he terms *unreflected naturalism*. He states [9; p.2]:

> "They [psychologists] tend to proceed as though everyday psychological categories, like intelligence, emotion or learning, represented natural kinds, as though the distinctions expressed in such categories accurately reflected the *natural* divisions among psychological phenomena. Psychological discussions typically assume that there really is a distinct kind of entity out there that corresponds exactly to what we refer to as an attitude say, and it is naturally different in kind from other sorts of entities out there for which we have different category names, like motives or emotions. The belief that scientific psychology adds to our knowledge of attitudes, drives, intelligence, etc., involves the implicit assumption that there is a fixed human nature whose natural divisions are reflected in this received network of categories. […] What is certain, however, is that psychological theory requires some pre-understanding of that which it is a theory of. That pre-understanding has generally involved the unspoken conviction that psychological categories constitute historically invariant phenomena of nature, rather than historically determined social constructions."

(Here, a discussion of the notion of naturalism would be in order, but it exceeds the scope of this article; confer [12, 13].) Psychology, for Danziger, is far from being about a fixed human nature that can be studied in an objective way, independent of how we understand it – for him, psychology is not about natural kinds but about kinds that are self-referential and intrinsically bound to social practice. One of the most important features of human kinds is, in this regard, that they exert an influence on themselves. Brinkmann (mostly quoting Hacking's work) nicely expresses the point [6; pp.773-774]:

> "An important feature of human kinds is that they can exert effects on themselves (Martin & Sugarman, 2001). They are affected by their classifications and interact with their classifications, sometimes affecting the classifications themselves. […] This is the looping effect of human kinds: 'People classified in a certain way tend to conform to or grow into the ways they are described; but they also evolve in their own ways, so that the classifications and descriptions have to be constantly revised' (Hacking, 1995a, p. 21). And further: 'Inventing or molding a new kind, a new classification, of people or of behavior may create new ways to be a person, new choices to make, for good or evil. There are new descriptions, and hence new actions under a description. (p. 239)' ".

As said, I will not so much focus on the classificatory or sorting function of human kinds that Hacking emphasises. I shall now rather discuss how science's outlook on the human mind and the way it is perceived by individuals and the public could change how people think and act, and thus change science's very (future) findings. For if a science of the mind investigates human mentality and, at the same time, is in the process of changing people's mentality, it is bound to change what it is "discovering" about it. The quest of understanding the human mind is a looping journey. Choudhury, Nagel and Slaby express the idea in somewhat different manner: "[R]epresentations of the objectified phenomenon influence the phenomenon itself and its subsequent study, and thus, the journey is a loop" [14; p.65].

Let me now turn to cognitive neuroscience's outlook on the human mind, its wide integration into society and to presenting some empirical evidence that indirectly shows how an outlook on the human mind affects people's beliefs, behaviour and mentality. Last, I will discuss how this shows that cognitive (neuro)science is exerting an influence on its very (future) findings.

UBIQUITOUS NEUROSCIENCE: FROM NEURO-TAGGED DISCIPLINES TO SOCIAL STRUCTURE

Neuroscience is becoming better and better in measuring the physical underpinnings of mentality. Its theories are more and more successful in explaining various aspects of the human mind and behaviour in terms of physical processes and mechanisms. It seems the time is ripe that we start understanding the human mind more in terms of its physicality and less in terms of its being separate from it. As Farah puts it [15; p.586]:

> "Neuroscience does not merely give us new tools to be used to the benefit or detriment of humanity; it gives us a new way of thinking about humanity. […] Neuroscience provides an alternative perspective, from which human behavior can also be understood as the result of physical causes. Even for people who do not follow the latest trends in science or spend time thinking about the nature of humanity, the applications of neuroscience […] will provide many reminders that our minds are, at root, physical mechanisms. By making people part of the clockwork universe, neuroscience challenges many assumptions about morality and personhood. […] as the neuroscience of personality, decision making, and impulse control begins to offer a more detailed and specific account of the physical processes leading to irresponsible or criminal behavior, the deterministic viewpoint will probably gain a stronger hold on our intuitions. Whereas the laws of physics are a little too abstract to displace the concept of personal responsibility in our minds, our moral judgments might well be moved by a demonstration of subtle damage to prefrontal inhibitory mechanisms wrought by, for example, past drug abuse or childhood neglect."

It further seems that neuroscience and its mechanistic and sometimes deterministic understanding of the human mind and behaviour are being more and more accepted by the public and discussed in the media [16-18]. In the media, for instance, the brain has been held responsible for many things, adolescents' decisions and purported irresponsible behaviour, among others. To provide one illustration: "Recent studies of brain development in teenagers may finally give parents the scientific authority to say "No you're not!" in answer to the common adolescent complaint, "But I'm old enough to make my own decisions!" That authority comes from brain imaging studies that reveal some surprising features of the adolescent brain" [19].

Neuroscientific methods, findings, and theories are also being more and more integrated into social sciences – traditionally perceived as "softer" sciences in comparison to the more "hard" natural sciences – such as psychology, educational science, economics, business sciences, organizational science, etc., and even into some branches of the humanities (e.g., law, ethics, art). This is well visible from the surge of new neuro-tagged disciplines in the last two decades, such as neuroeducation [20], neuroeconomics [21], neuromarketing [22], consumer neuroscience [23], neuroesthetics [24] and neurolaw [25].

Further, neuroscientific findings, methods, concepts, and theories, are also becoming part and parcel of various social domains, such as the law, education, and policy-making. In the domain of the law, there are, for instance, more and more cases where a "My brain made me do it" defence is being used to diminish the culpability of criminals [26-28], and in the U.S.A. – where the rise in the use of neuroscience in matters of law is most prominent – the supreme court has already partly based its decision to categorically exclude adolescents from sentence of life without the possibility of parole in non-homicide cases upon neuroscientific findings on brain development (in *Graham v. Florida*; see [29]).

In the domain of education, many aspects of learning and teaching have been discussed in light of neuroscientific findings of the human brain and cognition, and various interventions proposed and discussed. For instance in the domains of mathematics, reading, brain training, spaced learning, learning disorders, etc. (see [30] for a review and scope of application across various domains).

Further, it is not only neuroscience that claims that the human mind and behaviour is caused by some sort of unconscious mechanism or mainly guided by unconscious processing. Psychology and behavioural economics of human judgment and decision-making, in large part, make similar claims. They are, even though still predominantly behaviourist, also more and more infused with neuroscience; see [31]. Behavioural sciences of decision-making argue, for instance, that human rationality and self-control are rather limited and that our judgments and choices are many times biased, predominantly lead by unconscious processes, e.g., [32, 33]. Some authors [34, 35] thus propose decision nudges as a solution to flawed human decision-making. Namely, they claim that we should implement changes to background environments (to simplify decision-making, to eliminate the need for choice by default rules, etc.) against which people make choices, to nudge or steer people's choice and behaviour in health, wealth, and happiness-promoting direction. Moreover, more and more governments, policy-making agencies and economic institutions worldwide are trying to base their decisions about policy-making and/or change in social structures upon behavioural insight that is the "result from multidisciplinary research in fields such as economics, psychology and neuroscience, to understand how humans behave and make decisions in everyday life." [36; p.10] For some examples see the following lengthy publications: European Commission's in-house science service (the Joint Research Centre) report *Behavioural Insights Applied to Policy* [36]; the World Bank's report *Mind, Society, and Behavior* [37]; and the OECD's report *Behavioural Insights and Public Policy: Lessons from Around the World* [38].

Together with its growing popularity that is visible through the emergence of new neuro-tagged disciplines, discussions of neuroscientific findings in the media, and neuroscience's growing acceptance in various public domains (such as the law, economics, policy-making, and education), neuroscience seems to not have only reached most scientific communities, but also a large part of the public. This also means that neuroscience is on the way to challenge – if this has not already happened – our self-understanding [18; p.220]:

> "Contemporary neuroscience carries particular social weight. In today's secular societies, the brain is an acutely significant organ, represented as the seat of mind and self (Rose, 2007). Consequently, the production of brain-related knowledge is culturally important, carrying implications for how people see themselves as individuals and human beings. Brain-based information possesses rhetorical power: logically irrelevant neuroscience information imbues an argument with authoritative, scientific credibility (McCabe and Castel, 2008; Weisberg et al., 2008). Thus, the assimilation of neuroscience into public consciousness may have repercussions for beliefs, attitudes, and behavior, and as neuroscience grows in prominence, it is necessary to cultivate awareness of how it is mobilized in society."

HOW SCIENCE'S OUTLOOK ON THE HUMAN MIND AND ITS RECEPTION AFFECT ITS VERY FINDINGS

Let me now present some studies that indirectly show how changing people's beliefs about human agency towards a more deterministic and/or mechanistic conception – which seems similar to how neuroscience is being interpreted, at least in some contexts – changes how people think and behave. I will also consider the question of how cultural beliefs influence the human mind and how this bears upon science's findings of cognitive phenomena.

In their seminal study, Vohs and Schooler [39] demonstrate that changing people's beliefs about free will changes their moral behaviour. They show that weakening people's belief in free will – by presenting them with deterministic and/or reductionistic descriptions of human mentality and action (e.g, a passage from F. Crick's 1994 book [40]) – makes people, compared to the control group, passively (study 1) and actively (study 2) cheat more. Furthermore, studies show that weakening belief in free will increases aggression and decreases willingness to help [41], induces impulsive and antisocial tendencies [42], and even leads to change in basic brain function, such as to the reduction of the readiness potential (linked to voluntary motor action) [43]. According to Rigoni and colleagues, such studies show the following [43; p.617]:

> "[U]ndermining the idea that we are "the masters of our own houses" … presumably reduces the intentional effort we put into action. … Putting less effort into an action might weaken our sense of agency for these actions and lead to a reduced feeling of responsibility. This reduced feeling of responsibility would very likely result in more careless and irresponsible behavior. … From this perspective, we could hypothesize that dismissing the idea that people can control their own actions acts as a nonauthorship indicator, thereby decreasing people's sense of authorship. In sum, … this suggests that abstract belief systems might have a much more fundamental effect than most people would expect."

Similarly, the functioning of self-control – one of the important aspects of human agency psychology and neuroscience avidly investigate – can reveal itself in different ways according to distinct cultural beliefs. Research by Savani and Job [44] shows, for instance, that the ego-depletion effect – exerting willpower leads to worse performance on further mentally demanding task – is quite dependent on people's beliefs about the exertion of will and self-control. They show that Indians' acts of self-control *improved* their subsequent performance on demanding mental tasks, contrary to the American subjects, whose subsequent performance worsened in accordance with the traditional ego-depletion effect. Namely, Indians, contrary to the Americans, exhibited the reverse ego-depletion effect. According to the authors of the study, this is due to their belief that exerting willpower is "energising" – a belief not held by the American subjects (as measured on a questionnaire). The part of the American subjects that held similar beliefs to Indians also exhibited the reverse ego-depletion effect.

The often-naturalised phenomenon of ego-depletion is thus not a universal phenomenon, rooted in human neurobiology alone. Rather, it is to a certain degree dependent on how people understand themselves – be it the consequence of cultural or scientific beliefs about the nature of the human mind. (Scientific beliefs, it must be emphasised, are often also part and parcel of a culture and its values.) The authors of the study [44] state something quite in line with the ideas put forth in this article: "Peoples' cultural background and beliefs contribute considerably to what has been primarily viewed as a biological phenomenon" [44; p.15]. Culture and/or science have the potential to exert an influence on people's beliefs about their minds and that, in turn, changes how their minds function. Hence, the purportedly objective scientific findings of seemingly biological phenomena *loopingly* change according to how people understand themselves; at least in part.

I do not wish to claim that neuroscience *per se* entails that human agency is an illusion or that neuroscience, as such, diminishes human agency (e.g., self-control). Science alone – without someone to interpret, understand, or discuss it – does not have any point of view on the human mind. However, the relation between what neuroscience (or, better put, neuroscientists) states about the human mind and how this is received and interpreted by individuals and the public, is quite important for people's self-understandings. (See also

Dumit's works [45, 46] where he points out that people who see neurobiological evidence for their mental illness, for instance, interpret this evidence in many different ways and act accordingly.) And even though it was not true that neuroscientific findings and theories, as I was in part implying, are being interpreted in the direction of deterministic and mechanistic conception of the human mind, neuroscience seems to have become rather ubiquitous in many aspects of our lives and is, for this reason, probably on the way to redefine how we understand ourselves. In consequence, cognitive (neuro)scientific future findings are bound to be affected by new self-understandings that are continually being brought forth by sciences of the mind, creating a *looping journey* of understanding the human mind.

IN CONCLUSION: LOOPING MINDS AND AN ETHICAL OUTLOOK ON COGNITIVE SCIENCE

When people embrace new (scientific) conceptualisations of the human mind, new, different ways of being a human being become possible. As Hacking puts it: "When new descriptions become available, when they come into circulation, or even when they become the sorts of things that it is all right to say, to think, then there are new things to choose to do. When new intentions become open to me, because new descriptions, new concepts, become available to me, I live in a world of new opportunities" [1; p.236]. And, as new ways of self-understanding become part and parcel of social discourse, activities, and structure, people's behaviour and activities, thoughts and beliefs, feelings and desires, and their self-understandings change as well. Researching, explaining and understanding particular aspects of the human mind are hence not activities of discovery, but, at least to some degree, activities of co-creation and co-construction. That is, what we discover about the human mind is rather dependent on how people *understand themselves* and how people self-understand is partly dependent on science's outlook on the nature of the human mind. Understanding the human mind is thus inherently self-referring. (In cognitive science, the self-referential and looping nature of understanding the human mind is taken seriously, for instance, by the enactive account of cognition put forth by Varela, Thompson, and Rosch [47]; see also [48] for discussions on circularities of the mind.)

This, however, implies that cognitive science (or any science of the human mind) is not in the business of uncovering objective *truths* – devoid of particular perspective and value-free – but in the business of "making up people" [4]. Understanding the human mind and the way we *decide* to understand the human mind (ourselves and others) is thus an endeavour full of personal and cultural values, of what we aspire to become and is, all in all, quite an existential journey. This might remind us that cognitive science – the quest of which is to unravel the mystery of the human mind – is, in its core, essentially an ethical endeavour and not an endeavour that should reveal what the human mind *really is*. To conclude this thought with the beautiful concluding paragraphs of Varela's essay *The Creative Circle: Sketches on the natural History of Circularity* [49; p.320]:

> "[W]hen we follow the guiding thread of circularity and its natural history, we may look at that quandary from a different perspective: that of participation and interpretation, where the subject and the object are inseparably meshed. … It reveals to us a world where 'no-ground,' 'no-foundation' can become the basis for understanding that the age-old ideal of objectivity … is, by its own scientific standards, a chimera. We should do better to fully accept the notoriously different and more difficult situation of existing in a world where no one in particular can have a claim to better understanding in a universal sense. This is indeed interesting: that the empirical world of the living and the logic of self-reference, that the whole of the natural history of circularity should tell us that ethics […] is the very foundation of knowledge, and also its final point."

REFERENCES

[1] Hacking, I.: *Rewriting the soul*.
Princeton University Press, Princeton, 1995,

[2] Hacking, I.: *The looping effect of human kinds*.
In: Sperber, D.; Premack, D. and Premack, A.J., eds.: *Causal cognition: A multidisciplinary debate*. Clarendon Press, Oxford, pp.351-383, 1995,
http://dx.doi.org/10.1093/acprof:oso/9780198524021.003.0012,

[3] Reiss, J. and Sprenger, J.: *Scientific Objectivity*.
In Zalta, E.N., ed.: *The Stanford Encyclopedia of Philosophy*. Winter 2017 Edition, 2017,
https://plato.stanford.edu/archives/win2017/entries/scientific-objectivity,

[4] Hacking, I.: *Making Up People*.
In: Heller, T.C.; Wellbery, D.E. and Sosna, M., eds.: *Reconstructing Individualism: Autonomy, Individuality and the Self in Western Thought*. Stanford University Press, Stanford, pp.222-236, 1986,

[5] Bird, A. and Tobin, E.: *Natural Kinds*.
In: Zalta, E.N., ed.: *The Stanford Encyclopedia of Philosophy*. Spring 2018 Edition, 2018,
https://plato.stanford.edu/archives/spr2018/entries/natural-kinds,

[6] Brinkmann, S.: *Human Kinds and Looping Effects in Psychology*.
Theory & Psychology **15**(6), 769-791, 2005,
http://dx.doi.org/10.1177/0959354305059332,

[7] Brinkmann, S.: *Natural and Human Kinds*.
In: Teo, T., ed.: *Encyclopedia of Critical Psychology*. Springer-Verlag, New York, pp.1220-1225, 2014,
http://dx.doi.org/10.1007/978-1-4614-5583-7_657,

[8] Putnam, H.: *Meaning and reference*.
Journal of Philosophy **70**(19), 699-711, 1973,
http://dx.doi.org/10.2307/2025079,

[9] Danziger, K.: *Natural Kinds, Human Kinds, and Historicity*.
In: Maiers, W. et al., eds.: *Challenges to Theoretical Psychology*. Captus Press, Toronto, pp.24-32, 1999,

[10] Kordeš, U.: *A better Metaphor for Understanding Consciousness?*
Interdisciplinary Description of Complex Systems **13**(4), 525-533, 2015,
http://dx.doi.org/10.7906/indecs.13.4.4,

[11] Thompson, E.: *Looping Effects and the Cognitive Science of Mindfulness Meditation*.
In: McMahan, D.L. and Braun, E., eds.: *Meditation, Buddhism, and Science.* Oxford University Press, New York, pp.47-61, 2017,
http://dx.doi.org/10.31231/osf.io/9kh2y,

[12] Markič, O.: *Naturalism and the experiential perspective*.
Interdisciplinary Description of Complex Systems **13**(4), 535-539, 2015,
http://dx.doi.org/10.7906/indecs.13.4.5,

[13] Papineau, D.: *Naturalism*.
In Zalta, E.N., ed.: *The Stanford Encyclopedia of Philosophy*. Winter 2016 Edition, 2016,
https://plato.stanford.edu/archives/win2016/entries/naturalism,

[14] Choudhury, S.; Nagel, S.K. and Slaby, J.: *Critical Neuroscience: Linking Neuroscience and Society through Critical Practice*.
BioSocieties **4**(1), 61-77, 2009,
http://dx.doi.org/10.1017/S1745855209006437,

[15] Farah, M.J.: *Neuroethics: the ethical, legal and societal impact of neuroscience*.
Annual Review of Psychology **63**, 571-591, 2012,
http://dx.doi.org/10.1146/annurev.psych.093008.100438,

[16] Racine, E.; Bar-Ilan, O. and Illes, J.: *Brain Imaging: A Decade of Coverage in the Print Media*.
Science Communication **28**(1), 122-143, 2006,
http://dx.doi.org/10.1177/1075547006291990,

[17] Racine, E.; Waldman, S.; Rosenberg, J. and Illes, J.: *Contemporary neuroscience in the media.*
Social Science & Medicine **71**(4), 725-733, 2010,
http://dx.doi.org/10.1016/j.socscimed.2010.05.017,
[18] O'Connor, C.; Rees, G. and Joffe, H.: *Neuroscience in the Public Sphere*.
Neuron **74**(2), 220-226, 2012,
http://dx.doi.org/10.1016/j.neuron.2012.04.004,
[19] Talukder, G.: *Decision-making is Still a Work in Progress for Teenagers*.
https://brainconnection.brainhq.com/2013/03/20/decision-making-is-still-a-work-in-progress-for-teenagers,
[20] Howard-Jones., P.: *Introducing Neuroeducational Research: Neuroscience, Education and the Brain from Contexts to Practice*.
Routhledge, New York, 2010,
[21] Camerer, C.; Loewenstein, G. and Prelec, D.: *Neuroeconomics: How Neuroscience Can Inform Economics*.
Journal of Economic Literature **43**(1), 9-64, 2005,
http://dx.doi.org/10.1257/0022051053737843,
[22] Lee, N.; Broderick, A.J. and Chamberlain, L.: *What is 'neuromarketing'? A discussion and agenda for future research*.
International Journal of Psychophysiology **63**(2), 199-204, 2007,
http://dx.doi.org/10.1016/j.ijpsycho.2006.03.007,
[23] Kenning, P. and Linzmajer, M.: *Consumer neuroscience: an overview of an emerging discipline with implications for consumer policy*.
Journal für Verbraucherschutz und Lebensmittelsicherheit **6**(1), 111-125, 2011,
http://dx.doi.org/10.1007/s00003-010-0652-5,
[24] Nalbantian, S.: *Neuroaesthetics: neuroscientific theory and illustration from the arts*.
Interdisciplinary Science Reviews **33**(4), 357-368, 2008,
http://dx.doi.org/10.1179/174327908x392906,
[25] Gazzaniga, M.S. et al.: *The Law and Neuroscience.*
Neuron **60**(3), 412-415, 2008,
http://dx.doi.org/10.1016/j.neuron.2008.10.022,
[26] Catley, P. and Claydon, L.: *The use of neuroscientific evidence in the courtroom by those accused of criminal offenses in England and Wales*.
Journal of Law and the Biosciences **2**(3), 510-549, 2015,
http://dx.doi.org/10.1093/jlb/lsv025,
[27] Farahany, N.A.: *Neuroscience and behavioral genetics in US criminal law: an empirical analysis*.
Journal of Law and the Biosciences **2**(3), 485-509, 2016,
http://dx.doi.org/10.1093/jlb/lsv059,
[28] Strle, T. and Markič, O.: *Looping effects of neurolaw and the precarious marriage between neuroscience and the law*.
Balkan Journal of Philosophy **10**(1), 17-26, 2018,
http://dx.doi.org/10.5840/bjp20181013,
[29] Morse, S.J.: *Criminal Law and Common Sense: An Essay on the Perils and Promise of Neuroscience*.
Marquette Law Review **99**(1), 39-74, 2015,
[30] Howard-Jones, P.: *Neuroscience and Education: A Review of Educational Interventions and Approaches Informed by Neuroscience*.
The Education Endowment Foundation, London, 2014,
[31] Glimcher, P.W. and Fehr, E., eds.: *Neuroeconomics: Decision Making and the Brain.*
Academic Press, 2013,
[32] Haidt, J.: *The emotional dog and its rational tail: a social intuitionist approach to moral judgment*.
Psychological Review **108**(4), 814-834, 2001,
https://www.ncbi.nlm.nih.gov/pubmed/11699120,

[33] Kahneman, D.: *Thinking Fast and Slow*.
Farrar, Straus & Giroux, New York, 2011,

[34] Thaler, R.H. and Sunstein, C.R.: *Nudge: Improving Decisions about Health, Wealth and Happiness*.
Yale University Press, New Haven & London, 2008,

[35] Sunstein, C.R.: *Default Rules Are Better Than Active Choosing (Often).*
Trends in Cognitive Sciences **21**(8), 600-606, 2017,
http://dx.doi.org/10.1016/j.tics.2017.05.003,

[36] Lourenço, J.S.; Ciriolo, E.; Almeida, S.R. and Troussard, X.: *Behavioural insights applied to policy.*
Report EEUR 27726 EN, Publications Office of the European Union, 2016,
http://dx.doi.org/10.2760/903938,

[37] World Bank: *World Development Report 2015: Mind, Society, and Behavior*.
World Bank, Washington, 2015,
http://dx.doi.org/10.1596/978-1-4648-0342-0,

[38] OECD: *Behavioural Insights and Public Policy: Lessons from Around the World*.
OECD Publishing, Paris, 2017,
http://dx.doi.org/10.1787/9789264270480-en,

[39] Vohs, K.D. and Schooler, J.W.: *The Value of Believing in Free Will: Encouraging a Belief in Determinism Increases Cheating*.
Psychological Science **19**(1), 49-54, 2008,
http://dx.doi.org/10.1111/j.1467-9280.2008.02045.x,

[40] Crick, F.: *The Astonishing Hypothesis: The Scientific Search for the Soul.*
Charles Scribner's Sons, New York, 1994,

[41] Baumeister, R.F.; Masicampo, E.J. and DeWall, C.N.: *Prosocial Benefits of Feeling Free: Disbelief in Free Will Increases Aggression and Reduces Helpfulness.*
Personality and Social Psychology Bulletin **35**(2), 260-268, 2009,
http://dx.doi.org/10.1177/0146167208327217,

[42] Rigoni, D.; Kühn, S.; Gaudino, G.; Sartori, G. and Brass, M.: *Reducing self-control by weakening belief in free will*.
Consciousness and Cognition **21**(3), 1482-1490, 2012,
http://dx.doi.org/10.1016/j.concog.2012.04.004,

[43] Rigoni, D.; Kühn, S.; Sartori, G. and Brass, M.: *Inducing Disbelief in Free Will Alters Brain Correlates of Preconscious Motor Preparation*.
Psychological Science **22**(5), 613-618, 2011,
http://dx.doi.org/10.1177/0956797611405680,

[44] Savani, K. and Job, V.: *Reverse Ego-Depletion: Acts of Self-Control Can Improve Subsequent Performance in Indian Cultural Contexts*.
Journal of Personality and Social Psychology **113**(4), 589-607, 2017,
http://dx.doi.org/10.1037/pspi0000099,

[45] Dumit, J.: *Is it me or my brain? Depression and neuroscientific facts.*
Journal of Medical Humanities **24**(1-2), 35-47, 2003,
http://dx.doi.org/10.1023/A:1021353631347,

[46] Dumit, J.: *Picturing Personhood: Brain scans and biomedical identity*.
Princeton University Press, Princeton, 2004,

[47] Varela, F.J.; Thompson, E. and Rosch, E.: *The embodied mind: Cognitive science and human experience.*
The MIT Press, Cambridge, 1991,

[48] Klauser, F. and Kordeš, U.: *Loops and Recursions in Cognitive Science: Cross-Roads between Methodology and Epistemology*.
Interdisciplinary Description of Complex Systems **16**(4), 524-532, 2018,
http://dx.doi.org/10.7906.indecs.16.4.1,

[49] Varela, F.J.: *The creative circle: sketches on the natural history of circularity*. In: Watzlawick, P., ed.: *The invented reality: Contributions to constructivism.* Norton Publishing, New York, pp.309-325, 1984.

Interdisciplinary Description of Complex Systems 16(4), 545-557, 2018

SEEKING AFTER THE GLITTER OF INTELLIGENCE IN THE BASE METAL OF COMPUTING: THE SCOPE AND LIMITS OF COMPUTATIONAL MODELS IN RESEARCHING COGNITIVE PHENOMENA

Tine Kolenik*

Jožef Stefan Institute, Department of Intelligent Systems
Ljubljana, Slovenia

Unversity of Ljubljana, Faculty of Education, Center for Cognitive Science
Ljubljana, Slovenia

DOI: 10.7906/indecs.16.4.3
Regular article

Received: 12th December 2018.
Accepted: 17th December 2018.

ABSTRACT

Computational modelling has a rich history of successful use in researching cognitive phenomena. Its discoveries and applications do not seem to be stopping, yet with the rise of contemporary cognitive science paradigms, its scope and limits have been consistently put into the spotlight. The article reveals the scope of computational modelling by revealing its important role in progressing cognitive science research and helping cause important paradigm shifts as well as being useful at many levels of analysis. The limits are revealed to be some that are not easily solvable, if at all, mostly as they are not dependant on technological advancements. There are two main obstacles for computational modelling of cognitive phenomena: research bias, which manifests through the necessary presence of the designer's epistemological position as well as ideas on the mind and thus unavoidably being included in the model; and autonomy, the impenetrable basic element of living nature, which seems to make living organisms self-determine and thus create their own meaning in the world, something that seems to be unmodellable due to the designer always inputting her own meanings into the model and onto the modelled agents, which has been dubbed the PacMan Syndrome. Computational modelling is discussed in the light of these shortcomings, especially what it means to model living nature.

KEY WORDS

cognitive science paradigms, computational modelling, history of cognitive science, PacMan Syndrome, research bias

CLASSIFICATION

ACM: I.2.0
JEL: D91

*Corresponding author, η: tine.kolenik@ijs.si; -;
Jožef Stefan Institute, Department of Intelligent Systems, Jamova cesta 39, 1000 Ljubljana, Slovenia

INTRODUCTION

Terry Winograd, a computer scientist, mostly known for his work on natural language processing in his SHRDLU program in the 1960s and 1970s [1], became rather quickly disillusioned when he started researching the mind using artificial intelligence (AI) half a decade ago. "Seekers after the glitter of intelligence are misguided in trying to cast it in the base metal of computing," Winograd expressed his scepticism in his acclaimed paper "Thinking Machines: Can there be? Are we?" [2; p.217]. In it, he argues that computer models that are used for researching cognitive phenomena are limited to investigating and exhibiting only a particular kind of intelligence, "one that can usefully be likened to bureaucracy" [2; p.198]. This particular kind of intelligence, the only one computational modelling (CM) and AI are supposed handle, has two bases. This kind of intelligence covers everything that can be represented as a search problem, as some AI researchers say that everything in AI is a search problem [3]. Therefore, if it is a search problem, if it really is a search problem – and few things in life really are search problems for natural cognition – CM and AI can be used to research and replicate it. What is more – a lot more, since this seems to be deeply problematic and wholly unsolvable – is that search problem solving is only meaningful to humans. This is not disconnected from the second basis, which seems to be equally unsolvable. Since researchers in cognitive science usually (even if unknowingly) subscribe to one cognitive science paradigm or another, they have different ideas about cognition and how it works. These more oft than not epistemological presuppositions have their way of creeping into the models and simulations, stay there and manifest in influencing the research and its results profoundly. Removing researchers' bias and presuppositions is impossible as long as there are cognitive science paradigms.

However, CM has historically proven to be a powerful tool for researching cognitive phenomena. The limits of this scope are still debated on, especially in direct relation to natural cognition, but if anything, CM has produced impressing results. It has been one of the, if not the main catalyst for radical progress in cognitive science, whether through its virtues or through its faults, its positive results or its failures. A look through the history of cognitive science and the role of CM in the paradigm shifts is paramount to understanding the scope (and some limits) of CM. First, an overview of cognitive CM and its methods is given, with a few listed examples of how CM helped researchers in understanding cognitive phenomena at different levels of analysis. Second, an overview of the role of CM in paradigm shifts in cognitive science is presented.

THE ROLE OF COMPUTATIONAL MODELLING IN COGNITIVE SCIENCE AND ITS HISTORY

This section overviews how computational modelling can be and has been used in cognitive science. The overview looks at CM's use at different levels of analysis as well as its role in different cognitive science paradigms and their shifts. The role of CM in cognitive science's progress and paradigm shifts is especially important. Its recount is meant to show how certain ideas about the mind limited research on cognitive phenomena, and CM's use helped explore new ideas to fuel progress in how the mind is viewed.

There are numerous domains of phenomena CM can be used in for research. To showcase this wide scope, one example from each level of analysis, arguably pertaining to cognitive science, will be listed (see Table 1): from "the sociological level, the psychological level, the componential level, and the physiological level" [4; p.14].

Table 1. Hierarchy of four levels of analysis, from [4; p.12].

Level	Object of analysis	Type of analysis	Computational model
1	inter-agent processes	social/cultural	collections of agents
2	agents	psychological	individual agents
3	intra-agent processes	componential	modular construction of agents
4	substrates	physiological	biological realisation of modules

Computational research on the sociological level, which includes collective behaviour, group cognition, inter-agent processes and agent-environment interaction, is mostly concerned with emergent phenomena from interactions. Recently, CM of pedestrian behaviour has been successful in determining individual behaviour from emergent properties of crowding. Moussaïd, Helbing and Theraulaz found that “guided by visual information, namely the distance of obstructions in candidate lines of sight, pedestrians apply two simple cognitive procedures to adapt their walking speeds and directions” [5; p.6884]. Their model predicts “individual trajectories and collective patterns of motion in good quantitative agreement with a large variety of empirical and experimental data”, “the emergence of self-organization phenomena, such as the spontaneous formation of unidirectional lanes or stop-and-go waves” and that “the combination of pedestrian heuristics with body collisions generates crowd turbulence at extreme densities” [5; p.6884]. They assert that their model gives insight into spontaneous human behaviour in dynamic systems with many agents, and this insight has already been accepted in the circles of research on human crowd motion [6].

Computational research on the psychological level, which includes individual behaviour, mental states, emotion, perception, action, concepts, development, learning and so on, has been investigated with CM of child development. The endeavour has been largely successful in accounting for and giving insights into developmental processes. One of the big questions in this area is whether development and learning are distinct processes. Sun [4] summarises computational insights on this, which have become important components of contemporary child development theories in psychology [7]:

> “Using constructive learning models also resolves the ‘paradox of development’: It was argued that if learning was done by proposing and testing hypotheses, it was not possible to learn anything that could not already be represented. This argument becomes irrelevant in light of constructive learning models where learning mechanisms that construct representations are separate from the representation of domain-specific knowledge. […] [C]omputational modeling suggests that development is functionally distinct from learning” [4; p.20].

Computational research on the componential level, which includes intra-agent processes or cognitive architectures, produced a number of architectures, such as CLARION, ACT-R and Soar, which succeeded in various cognitive domains. CLARION has been indispensable in developing a comprehensive theory on skill learning. Sun [4] explains CLARION’s role in understanding skill learning:

> “At a theoretical level, this work explicates the interaction between implicit and explicit cognitive processes in skill learning […]. It highlights the interaction between the two types of processes and its various effects on learning […]. At an empirical level, a model centered on such an interaction constructed based on CLARION was used to account for data in a variety of task domains […]. The model was able to explain data in these task domains, shedding light on some apparently contradictory findings […]” [4; pp.20-21].

Sun argues that CLARION “helped to achieve a level of theoretical integration and explanation beyond the previous theorizing” [4; p.21].

Computational research on the physiological level, which includes biological substrates, is mostly applied in the discipline of computational neuroscience. One inspiring example of the use of neuroscientific CM is understanding how blind people perceive after being implemented with technological artefacts. Fine and Boynton [8] use "a computational model of axon fibre trajectories developed using traced nerve fibre bundle trajectories extracted from fundus photographs of 55 human subjects" [8; p.4] to understand "distortions of the perceptual experience" [8; p.1]. Gained insights about visual perception in blind people from this research has been used in developing better sight restoration technologies [9].

The examples on CM usage in studying cognitive phenomena at different levels of analysis show clear contribution to knowledge on cognition, thereby exhibiting CM's wide scope. To understand it in context, CM research that fuelled paradigm change in cognitive science will be discussed. Again, specific examples will be listed for showcasing. The paradigm changes that occurred in cognitive science shifted core epistemological presuppositions and ideas about the mind and cognition, removing some limits older paradigms unwittingly imposed onto research. Consecutive paradigms therefore trod the path to new knowledge, as new research was able to be produced on previously unsolvable obstacles.

Three paradigm shifts can be discerned in cognitive science and its history, each with its own scientific tool, in no coincidence belonging to computational methodology (see Figure 1): cognitivism [10] with symbolic information processing, connectionism [11] with artificial neural networks and embodied-embedded-enactive approaches [12] with mobile robots [13].

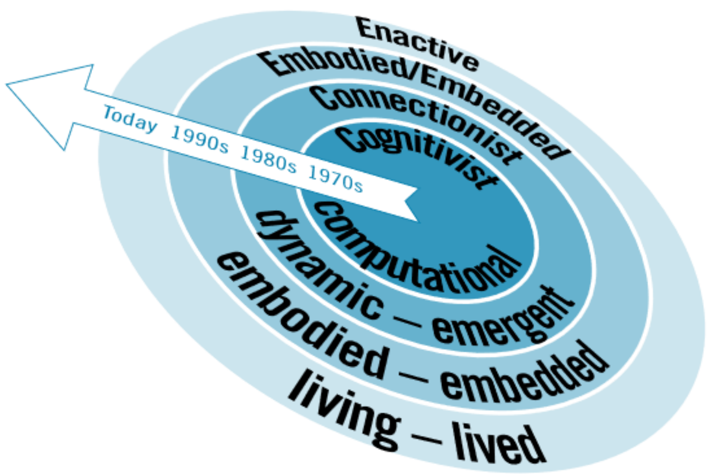

Figure 1. Illustration of one interpretation of the evolution of paradigms in cognitive science, from [13; p.76].

The first paradigm, the paradigm that gave birth to cognitive science, is cognitivism, which claims that cognition is computation with arbitrary symbols which represent the world [10]. Cognitivism can be said to have started with one important insight. This insight can be traced back to the Pythagorean school, which tried to describe reality with abstract descriptions – e.g., they figured out a "harmonious relationship between strings on an instrument which have certain simple mathematical relationships to each other (e.g. if one string is precisely double the length of another, its pitch is an octave lower)" [14; p.14]. The insight that parts of the world can be usefully represented by manipulation of arbitrary symbols has found home in thinking about mind and cognition as well, mostly due to many (initial) successes in applying this insight. The most well-known example may be a computer model of a chess player, which was first created by D. Prinz (Turing's colleague) in 1951. Progress in the modelling of a chess player, which was supposed to represent higher cognitive functions such as abstract thinking, escalated very quickly, pinnacled by IBM's Deep Blue 1996 winning against the then greatest chess player, G. Kasparov. Computational models of cognitive phenomena that showed results, like the artificial chess player, signalled to cognitive science researchers that natural cognition is computational as well.

However, cognitivist CM has had a lot of issues due to its ideas on how cognition works. One stark example is when dealing with large domains, where the number and complexity of rules are too high to model top-down, computational models are unable to output desirable behaviours. A shift in focus was needed for solving problems such as generalisation. Rumelhart and McClelland [15] were able to predict past tense forms of English verbs, a feat previously unsolved, when they dealt with computations not in the sequential, centralised, top-down and symbolic way, but rather in a parallel, distributed and bottom-up way:

> "The task is interesting because although most of the verbs in English (the regular verbs) form the past tense by adding the suffix '-ed', many of the most frequently verbs are irregular ('is' / 'was', 'come' / 'came', 'go' / 'went'). The net was first trained on a set containing a large number of irregular verbs, and later on a set of 460 verbs containing mostly regulars. The net learned the past tenses of the 460 verbs in about 200 rounds of training, and it generalized fairly well to verbs not in the training set [11; §.11]".

The researchers managed to solve the task using artificial neural networks, and the design and successful application of this method itself was influential in a paradigm shift that produced a new set of fundamental and epistemological ideas on cognition as an emergent phenomenon from a network of distributed, interconnected basic elements [11]. However, many researchers, such as Varela [12], felt that connectionism was still too focused on the brain, as connectionism classified it as the sole source of cognition, and everything, including the world – in the form of representations – was crammed in it. Opposition[1] to such views thought that the connectionist account of cognition was incomplete, and consequently could not completely explain how organisms perceive and act in the world so elegantly, effectively and efficiently.

The roboticist R. Brooks, who used CM to explore cognition, was part of the opposition. Brooks "pointed out that the bulk of evolution had been spent getting organisms to the stage where they had useful sensory and motor systems; phenomena such as tool use, agriculture, literature, and calculus represent only the most recent few 'seconds' in the evolutionary clock" [16; p.20]. He therefore focused his efforts "on the hard job of building systems which have sound basic sensorimotor capacities" [16; p.20], which were until his time in the 1980s still an unattainable goal for artificial design. Brooks' idea was that sensorimotor abilities and the organism's direct interaction with the world were crucial to understanding and unlocking the mystery of cognition, and that thinking about the latter as if it is building complex representations of the world in the central cognition unit (the brain), is wrong. These ideas were later applied to many phenomena thought to be in the domain of higher cognition, such as mental imagery [17]. This was a fairly obscure view in mainstream cognitive science circles, but a growing movement started to champion such ideas in that same time (although not all came to the same conclusions, neither furthered the cause by using computational methods [12]). Brooks demonstrated the value of his ideas on cognition by building a robot that could act intelligently in his laboratory by moving and performing simple tasks [18], a feat not achieved before. The relatively simple, multi-layered subsumption architecture was key for the robot to demonstrate Brooks' notion that "the world is its own best model" [19; p.15].

Such computational works played a role in spawning a new way of thinking about cognition in terms of its relationship with the world. The embodied-embedded-enactive view sees cognition as situated (cognition does not deal with abstract descriptions, but rather the world here and now), embodied (bodies non-trivially constitute cognition) and emergent (a central concept of connectionism as well, but here cognition is seen as emerging from the brain-body-environment triad). The organism supposedly enacts its world, constructs it. The role of CM is unclear in this last, yet no less central bit of the paradigm. The many problems, rather than solutions, in using CM in researching cognition seem to have had a much bigger, if not an essential role in paving the way towards the paradigm shift itself, as they strengthened the idea that the mind is not working as a computer is [12]. There are some emerging areas of investigation, very close to the enactive part of the paradigm, that use CM in their studying of how cognition constructs the world, such as predictive processing [20]. However, their impact, scope and limits are, as of yet, unclear.

UNDERSTANDING THE LIMITS TO THE SCOPE OF COMPUTER MODELLING IN COGNITIVE SCIENCE

Despite the important role of CM in cognitive science, there seem to be significant limits to its use – and most are not well understood at all. The necessary consequence of not understanding the limits of CM is that its scope and reach in researching cognitive phenomena are ambiguous as well. Two issues that seem to largely contribute to the limits and problems of CM are outlined in this section: the problem of meaning in artificial agents and a specific kind of research bias, associated with cognitive science paradigms and their presuppositions. To be able to outline these issues, the term ‘computational’, which is the cause of many misunderstandings when trying to theorise on CM, will be examined first.

The history of cognitive science and especially the rein of cognitivism have obscured the meaning of ‘computational’ to the degree where it may have become unusable in theoretical contemplations. Riegler, Stewart and Ziemke [21] agree that “the concept of ‘computational model’ is ambiguous and depends on whether it is used in the sense of the computational(ist) theory of mind or simply as a tool” [21; p.1]. They present two different senses in which ‘computational’ is used and which cause theoretical confusion:

> “The first sense of the term is that the processes being modeled are conceived of as being themselves computational in nature. This is the sense that is deliberately, honestly and openly taken on board by the proponents of the classical ‘computational theory of mind’ (CtM) – notably by Jerry Fodor and Noam Chomsky, who were historically the prime advocates of this view. […]
>
> The second sense of the term is that the author has used a computer as a tool to express a vision, to render certain aspects or properties of his model salient, to make them graspable by relatively direct intuition. This second sense of the term is quite different from the first. It is not (just) ‘wider’ or ‘looser’: that would invite an assimilation of the two, which would lead us back to the very conflation and confusion that we are trying to avoid” [21; p.2].

This differentiation is only the first step in uncovering the ambiguity of ‘computational’. Even if a researcher subscribes to CtM, the question remains whether any model where the world is simulated can be thought of as computational in the first sense – as being computational in nature. This means that even if a part of the model, the agent, is faithfully modelled, a large part, the world around the agent, is not. Why is that? Many natural phenomena, other than those related to living organisms, can be, to different degrees, described with mathematical formulas. They can therefore be modelled – e.g., the behaviour of winds can be described with differential equations, simulated and the result will be a useful approximation [22]. However, as opposed to the mind, no one believes that the winds are doing computations in and for their behaviours [21]. The idea is therefore that living organisms perform computations which are expressed by their behaviour in the world, while non-living nature does not. What happens then when the two are interacting – the modelled living organism and the modelled environment? If one is supposed to represent the true nature of the world (computational workings of living beings), while the other is supposed to represent approximate description of how the world works through mathematics, the ordeal seems to work against the first sense in which Riegler et al. [21] describe 'computational' – "processes being modeled are conceived of as being themselves computational in nature" [21; p.2]. But there is much deeper issue in this turmoil of modelling the mind and the world in different ways while using the same ambiguous concept. It is firmly connected to the confusion the term 'computational' has caused, and it has become thoroughly ingrained in theory and practice. Riegler [23] labelled this issue the PacMan Syndrome:

> "[I]t is useless to implement agents in artificial intelligence or artificial life with an a priori defined set of concepts, and to claim they were 'intelligent'. […] Artificial agents interact with anthropomorphically defined entities, such as 'food' and 'enemy', which make sense only to the programmer of the system. No attention is paid to questions like: How have organisms arrived at the idea that something is a source of food? How do they 'know' that another creature is a dangerous opponent?" [23; pp.4-5].

Riegler describes how agents are being modelled in the same way as the world is – by approximating observable behaviour from a third-person point of view rather than creating minds that would truly conform to the first sense of 'computational'. The problem is therefore not in creating agents through paradigms that conform to CtM, but rather a much more puzzling enigma. The enigma seems to be stronger than Riegler et al.'s delineation between the two senses of 'computational'. It is, as Füllsack [24] implies, that living organisms self-create knowledge and therefore behaviour, while the non-living natural world does not. He is somewhat pessimistic in his thoughts on modelling self-creation of knowledge by artificial agents instead of their creators, saying that CM can support the case for such ideas, but can never succeed in definitively producing it [24], therefore going by the second definition of 'computational' by Riegler et al. Peschl [25] argues similarly, claiming that "it is the designer who plays the role of evolution if he/she designs the network architecture and the structure of the artificial cognitive system" [25; p.2213].

Undertaking 'computational' matters means that one is describing systems from a third-person point of view, and it is this that causes core predicaments in using CM to investigate cognition[2]. Even if the mind performs symbolic computations with arbitrary symbols to function, the mind cannot be computationally described by a third-person observer by inputting knowledge, goals, the rules of the game of living and so on into the agent[3], as what defines living agents is, biologically speaking, their inherent self-determination [31]. The researcher as the ultimate creator cannot faithfully create agents, even if she can approximately describe behaviours.

But what if the researcher has different ideas about the mind, different from the ones CtM offers? The first definition of 'computational' by Riegler et al. does not seem to apply in that case at all. And what if the researcher's theories on cognition, her epistemological presuppositions and her decisions on which aspects to include in CM have significant impact on the results of her research? In his doctoral dissertation, Kjellman [32; p.i] concluded that "the efforts of computer modelling and simulation analysis revealed a pronounced observer-dependency regarding investigation." A considerable number of conditions of the modelled system is at the whim of its creator. Computational methods themselves carry a certain bias as well, as optional parameters and other features of various algorithms that are arbitrarily set make many theoretical issues overlooked or trivial and therefore removed from the phenomena as such [33]. Even CM of solely behavioural approximations carries a huge risk of observer bias, of the researcher's ideas on the mind influencing the research itself.

One such case of presuppositions and ideas about the mind affecting the results was investigated by Kolenik and Kordeš [34]. They investigated the relationship between mind and world by using CM with genetic algorithms to study whether it is isomorphic or non-isomorphic perception that is more evolutionary beneficial for modelled organisms. They introduced two computational models – a model by Hoffman, Singh and Prakash [35], which possessed cognitivist presuppositions, and a newly designed model, which built on Hoffman et al.'s model by replacing certain cognitivist presuppositions for enactivist ones, mostly focusing on the addition of a sensorimotor loop [36]. Both models produced the same final result, as they showed that non-isomorphic perception is evolutionary more beneficial than isomorphic. However, the sensorimotor loop caused the newly designed model to evolve faster, a feat that had not been anticipated in advance. Different ideas about cognition therefore made a significant impact on the results, which makes exploring the influence of different presuppositions on models' behaviour relevant in general, especially as this is an issue researchers are mostly not aware of. However, awareness seems to be the hard limit to trying to solve this – one cannot model cognition without starting from a set of epistemological and other presuppositions on the mind and how it works. Inevitably, these will be different amongst researchers, manifesting in different models and different outcomes even in research concerning the same cognitive phenomena.

Being aware of the presence of hard problems in CM, Peschl and Riegler [37] try to formulate an answer to what CM offers. They argue that CM and computer simulations have to be taken in a certain way, that what is important and insightful "are not so much results about details, but concern conceptual knowledge which can be used as input and stimulation for both empirical and epistemological investigations" [37; p.15]. They distinguish CM from other empirical investigations in cognitive science and its constituents, where most of the approaches to cognition "were more or less speculative and common-sense interpretations of cognitive phenomena" [37; p.15], as progress "in empirical sciences is based on a continuous process of construction, negotiation, and adaptation to the 'empirical data' " [37; p.15]. They point out the downsides of such an approach, as often "the complexity of cognitive processes and their substratum does not match the comparably poor empirical approaches and understanding of cognitive phenomena …" [37; p.15]. They feel that the more speculative areas of cognitive science open the door to CM and simulation:

> "Fortunately, the simulation method introduces a new dimension to cognitive science […]. Simulation models are especially interesting in the context of cognitive neuroscience, as its empirical results and theories are sometimes so rich in detail […] that it is almost impossible to relate them to cognitive phenomena. In other words, there is an explanatory gap and a strong tension between the epistemologically inspired questions on cognition […] and the empirical and highly detailed results from neuroscience. In this context the connectionist approach – in the broadest sense – plays a crucial role as mediator: it stands between the two poles of the rather speculative epistemological theories and the empirically grounded neuroscientific details and – in many cases – makes them compatible. This compatibility is achieved by the trick of focusing on the conceptual level of neural processes" [37; p.15].

Peschl's and Riegler's faith in computer modelling is limited, and precise in that limitation. They feel it is an extremely important method, not for "the technical details of simulation which we are interested in, but rather in the conceptual implications which these models have on the problem of knowledge representation" [37; p.15]. They claim that that conceptual level "can bring about both an empirically and epistemologically sound understanding of the ancient problem of representation in cognitive systems" [37; p.16] and "guide empirical research not only on the level of technical details, but – and this seems to be even more important—on a conceptual level (e.g., concerning the assumptions/premises of a research strategy, the epistemological framework and foundations, etc.)" [37; p.16]. Peschl [38] sums it up: "Computers are playing an important role as simulating instruments for artificial neural networks in order to achieve a deeper understanding of cognitive processes in an interdisciplinary context" [38; p.192].

It is apparent that the term 'computational' comes with a conceptually ambiguous burden, which has theoretical and practical consequences when it relates to being used to investigate natural phenomena. This is one of the reasons that some approaches have branded themselves or have been branded as anti-computational, which begs the question whether non-computational accounts of the mind can be computationally modelled. Like wind, which does not perform computations itself, yet can be usefully computationally modelled, non-computationalist accounts of the mind may follow suit. What is more, Riegler et al. [21] question the sentiment that enactivist approaches, for example, are necessarily anti- or non-computationalist. They are not completely clear on why that is, but a strong case is to be made for computational methods being used in regards to enactive approaches. What is clear is that enactive approaches have made considerable progress partly due to the use of computational and robotic models – and vice versa, enactive approaches definitely inspire CM [17, 29, 39]. Without presupposing that particular phenomena are computational in nature as such means that everything can be (descriptively) modelled and simulated [4, 40] – and many phenomena have been. This means that even non-computationalist approaches can benefit from being modelled and vice versa – CM can benefit from non-computationalist approaches. The deeper question of further addressing what the differences between living and non-living nature mean for modelling, however, remains wide open.

DISCUSSION AND CONCLUSION

Computational modelling as an investigative method in cognitive science has a long and rich history of successes and failures. Both have been essential in furthering the role of CM as well as furthering our understanding on the mind. However, after a good number of decades of research in cognitive science, CM has persistently encountered (at least) two monumental obstacles, described in this article, which do not seem to be solvable. Unlike most other obstacles CM has faced in that time, the ones described in this article do not seem to be connected to the lack of technological advancements.

Research bias in terms of CM of the mind is not something that can be avoided. To model the mind, some initial position on what the mind is has to be taken, inevitably contaminating the modelling and influencing the outcome. This is not inherently good or bad – it is simply the nature of such investigation. The PacMan Syndrome, the problem of designing models where the designer determines what is meaningful for her creations instead of designing artificial agents that self-determine, which seems to be an essential feature of living organisms, seems unavoidable as well. It may therefore be beneficial to rethink whether the question of what the scope and limits of CM for researching cognitive phenomena are is the correct one. Since CM clearly has limits, albeit only somewhat identified, the attention has to be turned to cognitive science for answers on what parts of cognitive phenomena behave in a way that can be faithfully modelled. H. von Foerster, the famous cyberneticist, may be one of the few thinkers that began exploring this important question precisely and non-abstractly. He argued that phenomena can be divided into trivial and non-trivial forms (or machines, as von

Foerster termed it) [41]. A trivial form is one that is independent from previous and other operations, analytically determinable, repeatable and predictable. A non-trivial form is one where this is not true. A non-trivial form is unpredictable, its workings cannot be deduced from outcomes, and past outcomes do not allow for prediction of future outcomes. What are then the conditions under which cognitive phenomena are trivial, or become trivial? Which cognitive phenomena are trivial in the first place? What happens when the non-trivial becomes computationally described? What gets lost in the process of trivialisation? How is the trivialised form of a phenomenon connected to its non-trivialised form? And last, but not least – is von Foerster's idea on the trivial and non-trivial a complete account of how to think about this issue? One thing remains certain – computational modelling is not going away, and its successes in describing cognitive behaviour and mimicking it keep on coming [42]. Truly gaining knowledge on the nature of (human) cognition from these successes is another issue.

ACKNOWLEDGEMENTS

I would like to express gratitude to Urban Kordeš, who helped spawn and chisel many ideas in this article. Without him, this work would not have happened.

REMARKS

[1]Opposition is not meant to be taken as a coherent, organised movement, but many voices from different areas of cognitive science that mainly converged in their critique.

[2]This insight is very similar to the insights of second-order cybernetics, a discipline born out of the discipline of cybernetics, the study of systems. Second-order cybernetics studies systems that study systems. For more on second-order cybernetics, see [26-27].

[3]This conclusion is echoed in the enactive paradigm in cognitive science with its concept of autonomy, which is why CM may not be so important in enactivism, as was described in the previous section on the history of cognitive science. Autonomy is highly relevant to the limits of CM this article is tackling, but is, due to the specific nature of enactivism as a biologically-inclined paradigm, outside this article's immediate scope. To find out more on enactive autonomy, see [28-30].

REFERENCES

[1] Winograd, T.: *Understanding natural language*.
Academic Press, New York, 1972,

[2] Winograd, T.: *Thinking machines: Can there be? Are we?*
In: Sheehan, J. and Sosna, M., eds.: The Boundaries of Humanity: Humans, Animals, Machines. University of California Press, Berkeley, pp.198-223, 1991,

[3] Wallingford, E.: *Session 23: Turning One Problem Into Another*.
https://www.cs.uni.edu/~wallingf/teaching/cs3530/sessions/session23.html, accessed 30th November 2018,

[4] Sun, R.: *Introduction to computational cognitive modeling*.
In: Sun, R., ed.: *Cambridge handbook of computational psychology*. Cambridge University Press, New York, pp.3-19, 2008,
http://dx.doi.org/10.1017/CBO9780511816772.003,

[5] Moussaïd, M.; Helbing, D. and Theraulaz, G.: *How simple rules determine pedestrian behavior and crowd disasters*.
Proceedings of the National Academy of Sciences of the United States of America **108**(17), 6884-6888, 2011,
http://dx.doi.org/10.1073/pnas.1016507108,

[6] Duives, D.C.; Daamen, W. and Hoogendoorn, S.P.: *State-of-the-art crowd motion simulation models.*
Transportation Research Part C: Emerging Technologies **37**, 193-209, 2013,
http://dx.doi.org/10.1016/j.trc.2013.02.005,

[7] Shultz, T.R. and Sirois, S.: *Computational models of developmental psychology*.
In: Sun, R., ed.: *The Cambridge handbook of computational psychology*. Cambridge University Press, New York, pp.451-476, 2008,
http://dx.doi.org/10.1017/CBO9780511816772.020,

[8] Fine, I. and Boynton, G.M.: *Pulse trains to percepts: the challenge of creating a perceptually intelligible world with sight recovery technologies*.
Philosophical Transactions of the Royal Society B: Biological Sciences **370**(1677), No. 20140208, 2015,
http://dx.doi.org/10.1098/rstb.2014.0208,

[9] Beyeler, M.; Rokem, A.; Boyton, G.M. and Fine, I.: *Learning to see again: biological constraints on cortical plasticity and the implications for sight restoration technologies*.
Journal of Neural Engineering **14**(5), No. 051003, 2017,
http://dx.doi.org/10.1088/1741-2552/aa795e,

[10] Putnam, H.: *Brains and behaviour*.
In: Butler, R.J., ed.: *Analytical Philosophy*. Vol. II. Basil Blackwell, Oxford, pp.1-19, 1968,

[11] Garson, J.: *Connectionism*.
https://plato.stanford.edu/entries/connectionism, accessed 27th November 2018,

[12] Varela, F.J.; Thompson, E. and Rosch, E.: *The embodied mind: cognitive science and human experience.*
The MIT Press, Cambridge, 2016 (original work published 1991),

[13] Froese, T.: *From Cybernetics to Second-Order Cybernetics: A Comparative Analysis of Their Central Ideas*.
Constructivist Foundations **5**(2), 75-85, 2010,

[14] Ó Nualláin, S.: *The Search For Mind: A new foundation for Cognitive Science*.
Intellect Books, Bristol, 2002,

[15] Rumelhart, D.E. and McClelland, J.L.: *On learning the past tense of English verbs*.
In: Rumelhart, D.E. and McClelland, J.L., eds.: *Parallel Distributed Processing*. Vol. 2. MIT Press, Cambridge, pp.216-271, 1986,

[16] Elman, J.L.: *Connectionism, artificial life, and dynamical systems: new approaches to old questions*.
In: Bechtel, W. and Graham G., eds.: *A companion to cognitive science*. Blackwell Science, Hoboken, pp.488-505, 1998,

[17] Jug, J.; Kolenik, T.; Ofner, A. and Farkaš, I.: *Computational model of enactive visuospatial mental imagery using saccadic perceptual actions*.
Cognitive Systems Research **49**, 157-177, 2018,
http://dx.doi.org/10.1016/j.cogsys.2018.01.005,

[18] Brooks, R.: *A robust layered control system for a mobile robot*.
IEEE Journal on Robotics and Automation **2**(1), 14-23, 1986,
http://dx.doi.org/10.1109/jra.1986.1087032,

[19] Brooks, R.A.: *Intelligence without Representation*.
Artificial Intelligence **47**(1-3), 139-159, 1991,
http://dx.doi.org/10.1016/0004-3702(91)90053-m,

[20] Clark, A.: *Whatever next? Predictive brains, situated agents, and the future of cognitive science*.
Behavioral and Brain Sciences **36**(3), 181-204, 2013,
http://dx.doi.org/10.1017/S0140525X12000477,

[21] Riegler, A.; Stewart, J. and Ziemke, T.: *Computational Approaches to Constructivism*.
Constructivist Foundations **9**(1), 1-6, 2013,

[22] Zárate-Miñano, R.; Anghel, M. and Milano, F.: *Continuous wind speed models based on stochastic differential equations*.
Applied Energy **104**, 42-44, 2013,
http://dx.doi.org/10.1016/j.apenergy.2012.10.064,

[23] Riegler, A.: *Towards a Radical Constructivist Understanding of Science*.
Foundations of Science **6**(1), 1-30, 2001,
http://dx.doi.org/10.1023/A:1011305022115,

[24] Füllsack, M.: *Author's Response: Constructivism as Possibility?*
Constructivist Foundations **9**(1), 23-25, 2013,

[25] Peschl, M.F.: *A simulation system for the investigation of cognitive processes in artificial cognitive systems-Radical connectionism and computational neuroepistemology*.
In: Proceedings of the 1991 IEEE International Joint Conference on Neural Networks. Vol. 3, pp.2210-2215, 1991,
http://dx.doi.org/10.1109/IJCNN.1991.170716,

[26] Von Foerster, H.: *Ethics and second-order cybernetics*.
Stanford Humanities Review **4**(2), 308-319, 1995,

[27] Mancilla, R.: *Introduction to Sociocybernetics (Part 1): Third Order Cybernetics and a Basic Framework for Society*.
Journal of Sociocybernetics **42**(9), 35-56, 2011,

[28] Barandiaran, X.E.: *Autonomy and Enactivism: Towards a Theory of Sensorimotor Autonomous Agency*.
Topoi **36**(3), 409-430, 2017,
http://dx.doi.org/10.1007/s11245-016-9365-4,

[29] Froese, T. and Ziemke, T.: *Enactive artificial intelligence: Investigating the systemic organization of life and mind*.
Artificial Intelligence **173**(3-4), 466-500, 2009,
http://dx.doi.org/10.1016/j.artint.2008.12.001,

[30] Kolenik, T.: *Embodied Cognitive Robotics, the Question of Meaning and the Necessity of Non-Trivial Autonomy*.
In: Markič, O.; Gams, M.; Strle T. and Kordeš, U., eds.: *Cognitive science: proceedings of the 19th International Multiconference Information Society* – IS 2016. Institut Jožef Stefan, Ljubljana, pp.24-27, 2016,

[31] Maturana, H.R. and Varela, F.J.: *Tree of Knowledge: The Biological Roots of Human Understanding*.
Shambala, 1988,

[32] Kjellman, K.A.: *Constructive systems science – the only remaining alternative? A contribution to science and human epistemology*. Ph.D. Thesis.
http://www.diva-portal.org/smash/get/diva2:9397/FULLTEXT01.pdf, accessed 25th November 2018,

[33] Green, C.D.: *Are Connectionist Models Theories of Cognition?*
Psycoloquy **9**(4), No.2, 1998,

[34] Kolenik, T. and Kordeš, U.: *The Genetic Algorithm Model of the Interface Theory of Perception under Paradigm Cross-Examination: A Response to Hoffman, Singh and Prakash.*
in preparation,

[35] Hoffman, D.D.; Singh, M. and Prakash, C.: *The interface theory of perception*.
Psychonomic Bulletin & Review **22**(6), 1480-1506, 2015,
http://dx.doi.org/10.3758/s13423-014-0776-1,

[36] Di Paolo, E.; Buhrmann, T. and Barandiaran, X.E.: *Sensorimotor life: an enactive proposal.*
Oxford University Press, Oxford, 2017,
http://dx.doi.org/10.1093/acprof:oso/9780198786849.001.0001,

[37] Peschl, M.F. and Riegler, A.: *Does Representation Need Reality*?
In: Riegler, A.; Peschl, M. and von Stein, A., eds.: *Understanding Representation in the Cognitive Sciences*. Kluwer Academic & Plenum Publishers, New York, pp.9-17, 1999,
http://dx.doi.org/10.1007/978-0-585-29605-0_1,

[38] Peschl, M.F.: *Interdisciplinarity and Self-Organization in Computational Neuroepistemology*. In: Jackson, M.C. et al., eds.: *Systems Thinking in Europe*. Springer, Boston, pp.189-195, 1991, http://dx.doi.org/10.1007/978-1-4615-3748-9_23,

[39] Allen, M. and Friston, K.J.: *From cognitivism to autopoiesis: towards a computational framework for the embodied mind*. Synthese **195**(6), 2459-2482, 2018, http://dx.doi.org/10.1007/s11229-016-1288-5,

[40] Farkaš, I.: *Indispensability of Computational Modeling in Cognitive Science*. Journal of Cognitive Science **13**(4), 401-429, 2012, http://dx.doi.org/10.17791/jcs.2012.13.4.401,

[41] Von Foerster, H.: *Understanding Understanding: Essays on Cybernetics and Cognition*. Springer-Verlag, New York, 2003,

[42] Bansal, T. et al.: *Emergent Complexity via Multi-Agent Competition*. preprint arXiv:1710.03748v4 [cs.AI], https://arxiv.org/abs/1710.03748v3.

Interdisciplinary Description of Complex Systems 16(4), 558-566, 2018

CONSIDERATIONS ON SCIENTIFIC THOUGHT EXPERIMENTS

Maja Malec*

University of Ljubljana, Faculty of Arts
Ljubljana, Slovenia

DOI: 10.7906/indecs.16.4.4
Regular article

Received: 3rd November 2018.
Accepted: 21st December 2018.

ABSTRACT

I provide some considerations on scientific thought experiments, focusing on their epistemic value. First, I outline the distinctive features of scientific thought experiments, provide some historical background and, as an illustration, describe two thought experiments: Galileo's on falling bodies and Stevin's on inclined plane. I take thought experiments in physics as an example from which more general conclusions can be drawn – about thought experiments in other natural sciences, but also in philosophy, mathematics, and other sciences. Further, I present Kuhn's epistemic puzzle as well as some proposed solutions. This closely relates to the question what kind of processes are involved in scientific thought experimenting. The satisfactory answer must consider scientific discoveries. I outline the mental model account as the most promising account since it best incorporates the findings of cognitive science. I conclude with two issues that the account needs to resolve and the role cognitive science can play in this.

KEY WORDS

philosophy of science, philosophy of physics, thought experiments, naturalism, mental model

CLASSIFICATION

APA: 2630, 2340

JEL: D83, D91

PACS: 01.70.+w, 01.65.+g

*Corresponding author, η: maja.malec@ff.uni-lj.si; +386 (1) 241 1102;
Faculty of Arts, Aškerčeva 2, SI-1000 Ljubljana

INTRODUCTION

Thought experiments are employed in very diverse areas. They play a central role in philosophy which is in big part an ‘armchair pursuit’, as Williamson aptly puts it. By this he refers to the fact that traditional methods of philosophy ‘consist of thinking, without any special interaction with the world beyond the chair, such as measurement, observation or experiment would typically involve’ [1; p.1]. However, thought experiments play a significant role in most empirical natural sciences as well. Physics, for example, includes an impressive number of them. Its birth in the 17th century is characterized by some ingenious thought experiments by Galileo, Newton and Leibniz (Galileo’s falling bodies, Newton’s bucket, and Leibniz’s space shift), Einstein relied heavily on them (e.g. chasing a light beam, elevator or moving train) in the creation of relativistic physics, and they were crucial in the development of quantum mechanics (e.g. Heisenberg’s gamma microscope, Schrödinger’s cat, and EPR experiment). We often encounter them in mathematics, sociology, political science and history, but also partly in art [2]. Some find them in everyday life as well. Nersessian, for example, posits that experimenting in thought is involved in such mundane cognitive tasks as figuring out whether a large sofa would fit through the door, or what the best route from one place to another would be [3]. Others even recognize their usefulness in therapy. Gendler, for instance, suggests that fear of flying can be overcome by repeatedly imagining oneself flying safely [4].

Given the diversity, it is impossible to provide an all-encompassing definition of a ‘thought experiment’. There are, however, certain common characteristics that all thought experiments share. Namely, conducting a thought experiment involves contemplating an imaginary scenario with the aim of acquiring a piece of information. Specifics then depend on the area in which the thought experiment is conducted. Most work has been done on thought experimenting in philosophy and natural science (esp. physics), and researchers mostly agree that the differences between them are not in kind, but in degree. Moreover, thought experimenting in social sciences seems closer to such activity in philosophy than in natural sciences, so a unified account is undoubtedly preferable. Thus, in this paper I take thought experimenting in physics as an example from which more general conclusions can be drawn – about such activity in other natural sciences, but also in philosophy, mathematics, and other sciences.

First, I will outline the distinctive characteristics of scientific thought experiments, provide some historical background and, as an illustration, describe two thought experiments: Galileo’s on falling bodies and Stevin’s on inclined plane. Further, I will present the epistemic puzzle, which was first formulated by Kuhn, as well as some proposed solutions. This closely relates to the question what kind of processes are involved in scientific thought experimenting. I will conclude by presenting the mental model account which is, in my opinion, the most promising account since it best incorporates the findings of cognitive science.

THE STRUCTURE OF SCIENTIFIC THOUGHT EXPERIMENTS

Conducting a scientific thought experiment involves contemplating an imaginary scenario with the aim of illustrating, supporting or refuting some scientific hypothesis or theory. One distinctive feature of thought experimenting is that one reasons about a particular set of circumstances in order to reach a conclusion which does not pertain only to the described situation, but it is applicable more generally. That is, the conclusion, derived from this particular situation, has repercussions for certain hypothesis or theory. In the case of physical thought experiments, a certain hypothesis or theory concerning physical phenomena is confirmed or disconfirmed. (In less spectacular cases, thought experiments serve as heuristic devices, providing a graphic illustration of a scientific theory.) Another distinctive feature of thought experimenting is that the means by which we reach a scenario is imagination – we more or less vividly imagine a certain situation – and not observation.

Thought experiments, as their name already indicates, are liken to actual experiments, the main difference being that the latter are not conducted in thought, but in the real world. In both cases, we are supposed to draw from a particular situation some general conclusion concerning the actual world. A major difference, however, is that actual experiments can fail to confirm our predictions, while it is not immediately clear how thought experiments can fail.

This is why P. Duhem objects to the justificatory use of thought experiments in physics. He argues that by invoking a 'fictitious experiment' a physicist 'is justifying a principle not by means of facts observed but by means of facts whose existence is predicted, and this prediction has no other foundation than the belief in the principle supported by alleged experiment'. This is a vicious circle, Duhem argues, and concludes that thought experiments cannot teach us anything about the world, at most they can suggest hypotheses which can be tested only indirectly after the whole theory is constructed and its consequences compared with an actual experiment [5; pp.202-205].

In general, Duhem considers physical theory as "a system of mathematical propositions, deduced from small number of principles, which aim to represent as simply, as completely, and as exactly as possible a set of experimental laws" [5; p.19]. As such, it is not explanatory, but descriptive, a logical construction of given physical phenomena. It is therefore not surprising that thought experiments as well as models in his account merely serve as illustrations of the theory, thus facilitating our understanding. They are not a means of discovery.

Certainly, Galileo, one of the founding fathers of modern physics, would disagree with Duhem's disregard for thought experiments. His falling bodies thought experiment results in a rejection of the Aristotelian theory that heavier bodies fall faster than lighter ones, but also in a new proposal that all bodies, regardless of their weight, fall with an equal speed. Concretely, Galileo imagines two stones of different weights being tied together and asks about the speed of their fall. Obviously, the heavier stone, which on its own falls faster than the lighter one, will now, tied to the latter, fall a bit slower, while the latter will now fall a bit faster due to the faster speed of the first. The speed of the fall of the two bodies tied together is therefore in between the speeds of the two bodies falling separately. However, the two bodies tied together form a unity as well and this new body is the heaviest of them all and it should therefore fall the fastest. Contradiction! The new body cannot fall with two different speeds at the same time, and Galileo argues that the only way out of this paradox is to posit that the speed by which a body falls is independent of its weight [6].

When we, together with Galileo, imagine the proposed scenario and let it run, the falsity of Aristotelian theory becomes immediately obvious to us, and we do not need to perform any actual experiment to make sure as Duhem would have it. His contemporary, E. Mach, who was first to systematically study scientific thought experiments, has a more favourable opinion of them. In general, thought experiments help to prepare actual experiments, but some deliver results so convincingly that they make an actual experiment redundant. Thought experiments, such as Galileo's, are crucial for scientific progress since they lead 'to enormous changes in our thinking and to an opening up of most important new paths of inquiry' [7; p.138]. He believes that their value resides in their ability to access unarticulated knowledge that we acquired in our previous contacts with the world: "Everything which we observe imprints itself uncomprehended and unanalysed in our precepts and ideas … In these accumulated experiences we possess a treasure-store which is ever close at hand, and of which only the smallest portion is embodied in clear articulate thought" [8; p.28].

Mach describes this kind of knowledge as instinctive and points out that we feel at once that the result of a thought experiment is correct. The thought experiment he analyses is Stevin's inclined plane thought experiment by which he established the force needed to prevent an object to slide down a frictionless inclined plane. Stevin imagines an endless chain wrapped around a triangular prism with a horizontal base. There are two possibilities: the chain is in equilibrium or not. If it is not, then it must move continuously, thus constituting perpetual motion, which we immediately recognise as absurd. Therefore, the chain must be in equilibrium. Next, he imagines that the lower part of the chain, hanging over the prism, is cut off. Thus, only the links along the two diagonal planes remain, and clearly, the system is still in equilibrium. From this, Stevin infers that two bodies on two different inclined planes are in equilibrium if their weights are proportional to the lengths of the two planes. The description of the thought experiment is accompanied by two diagrams of the two main stages in the experiment – the prism with the endless chain wrapped around it and the prism with shortened chain – which adds to the persuasiveness of the experiment [8].

Stevin did not have to actually perform the experiment in order to confirm his finding. The thought experiment is persuasive enough to justify the conclusion. In fact, the experiment, which requires frictionless plane, could not be performed at all – a reminder that also real experiments rely not only on observation, but also on abstracting away certain conditions and on predicting what would the results in idealized circumstances be.

Now, it is difficult to be a sceptic about physical thought experiments since they clearly work – the history of physics is a proof of that. There are many examples of thought experiments by which physicists confirmed or disconfirmed a certain physical hypothesis or theory. The real puzzle occupying philosophers of science is rather whether one really acquires new knowledge of the real world by simply contemplating imaginary scenarios. If the answer is affirmative, then one needs to explain what kind of knowledge is so acquired and by what process in particular. And if it is negative, then one needs to explain why it seems to us that we learned something new and how this comes about.

THE EPISTEMIC PUZZLE

The formulation of the puzzle is due to Kuhn, who belongs to the next stage of systematic research of thought experiments from the 1950s through the 1970s, another important researchers from these stage being Popper and Koyré. He states it in the following way: "How, then, relying exclusively upon familiar data, can a thought experiment lead to new knowledge or to new understanding of nature?" [9; p.241].

As we have seen above, Duhem denies that thought experiments can be a source of knowledge and allows them only as a source of hypotheses, which are later, if incorporated in the theory, compared with observation. Mach, on the other hand, acknowledges their epistemic value – some thought experiments are crucial for scientific progress since they give access to unorganized, non-propositional knowledge that we previously acquired through experience. The acquired knowledge is therefore empirical knowledge of our physical world, which we actually already possessed, but was before a thought experiment unavailable in propositional form.

Kuhn himself views thought experiments in the context of scientific practice and recognizes their importance in times of scientific revolutions. They help scientists to recognize the limitations of the old paradigm and facilitate acceptance of the new. This is not a simple case of uncovering inconsistency in the scientist's conceptual apparatus. Kuhn points out that our concepts always come with physical implications, expectations about their use. Consequently, a certain concept can be in a perfect logical order and yet its use does not correctly reflect what the world is like. The function of thought experiments is to reveal such cases. The imagined situation explicitly presents already available, but not yet assimilated empirical information which the current scientific concept cannot capture. This shows that the conceptual apparatus does not fit all situations and it must be revised.

Kuhn solves the epistemic puzzle in such a way that thought experiments without any new empirical data teach us something new about our scientific concepts and their physical implications, and by this about the world itself. As Kuhn puts it, thought experiments can 'disclose nature's failure to confirm to previously held set of expectations', and second, 'they can suggest particular ways in which both expectation and theory must henceforth be revised' [9; p.261].

From 1990s onwards, the presented solutions were further developed, new ones were proposed, and research on scientific thought experimenting was enriched by forming closer relations with research on thought experimenting in other domains. One proposal is platonic rationalism defended by Brown [10]. He argues that there are some special cases of thought experiments which at the same time destroy an old theory and establish a new one; in other words, these experiments are simultaneously destructive and constructive. Such is, for example, Galileo's falling bodies thought experiment in which Aristotelian theory that heavier bodies fall faster is rejected and replaced by the theory that all bodies fall at the same speed. He calls them Platonic thought experiments since they afford us a direct insight into the abstract realm of universals. By grasping, or 'non-sensorily perceiving', relations between universals we acquire *a priori* knowledge of truths regarding the laws of nature, which are in his account necessitation relations between independently existing universals. In this he follows the Dretske-Tooley-Armstrong account of the laws of nature.

Brown argues that in conducting scientific thought experiments, we sometimes grasp the laws of nature *a priori,* solely by intuition. An opposite view, thoroughly empiricist in character, is defended by Norton [11]. According to him all knowledge about our world ultimately derives from experience, so Brown's interpretation of thought experiments as a source of *a priori* knowledge of the laws of nature must be rejected. He strenuously objects to intuition, or quasi-perceptual rational insight being a means by which we acquire knowledge since its working is utterly mysterious. On the other hand, thought experiments are not real experiments since they are only conducted in the head, and therefore cannot provide new empirical data either. This leaves only one option, namely, as Norton puts it, thought experiments 'can only reorganize or generalize what we already know about the physical world and make it explicit' [11; p.335]. Thought experiments thus provide new information about the physical world based on the rearrangement of old data. Moreover, he argues that thought experiments are nothing more than arguments which take us from our assumptions to the conclusion, namely to the outcome of the thought experiment. Now, if assumptions are only reorganized, then a thought experiment is a deductive argument, and if generalization is involved, then it is an inductive argument. Either way, thought experiments as epistemic devices are to be evaluated as arguments.

Brown and Norton propose diametrically opposed solutions to Kuhn's puzzle. They agree that thought experiments provide new empirical knowledge in the absence of any new empirical data, but draw different conclusions from this. Brown concludes that new data must come from some other source and that in a thought experiment we gain a rational insight into this source which is an abstract realm of universals and consequently of the laws of nature. Norton, on the other hand, concludes that new knowledge is elicited from old empirical data by deductive or inductive arguments.

Both solutions are quite unappealing. Norton's 'argument view' totally disregards the special nature of thought experiments. Why would we bother devising an imaginary scenario by dressing arguments up if a basic argument does the job? There must be some function that is particular to a thought experiment and that an argument cannot perform on its own. Moreover, a thought experiment is a twin of real experiment, and if the first is just a disguised argument, then so is the latter. Indeed, a real experiment can also be stated as an argument, but does this mean that it is nothing more than an argument? Surely not, but why would then be any different in the case of a thought experiment?

Brown's 'Platonistic view' fares even worse since the proposed epistemic method is unfamiliar and in need of explanation. Brown tries to strengthen his position by drawing parallels with Platonism in mathematics, but physics and mathematics differ significantly in character and aims. Mathematics is highly abstract, and while mathematical reasoning can provide predictions about nature, this is not its main purpose. On the other hand, providing such predictions is the main task of physics. Physical theory describes and explains the physical world, and it must be supported by observations and experiments. Why would then physical laws or the laws of nature, as Brown calls them, be about some mysterious universals inhabiting the Platonic, abstract realm? If it does not make much sense metaphysically, it makes even less sense epistemically. How do we come to know such laws directly? How does this special 'sense' work? It would be easier to claim that we discern them from the regularities that we observe in the physical world, but this option is unavailable to Brown. In Platonic thought experiments, the old theory is rejected and replaced by a new one without any new empirical information being provided, solely based on rational insight into abstract realm.

Brown and Norton both understood Kuhn's puzzle as wondering how new knowledge can be gained given that thought experiments are conducted entirely in one's head without any new empirical data being provided. However, besides knowledge, Kuhn also mentions a new understanding of nature, and his own solution is that thought experiments provide a better understanding of physical concepts and their applications to the physical world. If novelty is not limited to knowledge, the number of possible solutions increases. For example, Elgin [12] and Stuart [13] further explore the connection of thought experiments with understanding. One possibility is also Mach's suggestion that thought experiments provide access to non-propositional knowledge. This idea is further developed in the mental model account [14, 15]. Additionally, its proponents Nersessian and Miščević claim that thought experiments can also result in new experiences, concepts, beliefs, or abilities etc.

Another question that an epistemological account of scientific thought experiments must answer is what is involved in this experimenting, namely, what processes are involved. Norton denies that there is anything epistemically distinctive about it – we merely conduct ordinary deductive or inductive reasoning, we execute an argument. True, an argument is 'disguised in some vivid picturesque or narrative form', which gives it 'special rhetorical powers', but epistemically, it is irrelevant [16; p.1139]. Brown, on the other hand, posits intuition as a special faculty by which we access the abstract realm and acquire *a priori* knowledge of the laws of nature. However, he cannot plausibly explain how it works, and can only claim resemblance with mathematics where the use of intuition is quite established.

While it seems more plausible that thought experimenting is epistemically distinctive, as Brown claims, it is not very likely that we possess some special faculty reserved exclusively for this task. It is far more likely that we employ a set of various cognitive processes that primarily evolved in order to perform some more mundane tasks, and were only later adopted for this more sophisticated use. And because of this, I believe that scientific thought experimenting can be satisfactorily explained only if we turn to cognitive science and its findings for support.

One such example is the naturalistic version of the intuition based account. For example, Gendler claims that in scientific thought experimenting, due to focus on a specific scenario (and not on a general schema), quasi-sensory intuitions are evoked, which results in new beliefs on contingent features of the natural world [4]. Consequently, the distinctive feature of intuitions involved in scientific thought experimenting is not their subject matter – similarly as observations, they deal with the contingent features of the natural world – but the way in which we come to possess them – they are not the result of some explicit reasoning process that we are conscious of, but they appear immediately in our mind while contemplating an imaginary scenario.

In explaining the nature and functioning of intuitions as well as of the imagination which is at the core of thought experimenting, the naturalist account relies on the findings of cognitive science. One difficulty to be addressed is the evolutionary explanation of the capacity to intuit, which seems to cast doubts on the reliability of intuitions. Much work still needs to be done, and both could profit from this interaction.

MENTAL MODEL ACCOUNT

Another promising naturalistic account explains thought experimenting as a form of mental modelling [14, 15]. The mental model was first developed and used in cognitive science. Nersessian further developed the original idea, supplementing it with findings from cognitive science, concluding that the relevant form is a form of 'simulative model-based reasoning, where inferences are made through constructing and manipulating models, whether conceptual, physical or computational' [17; p.310, 18].

She bases her account upon the idea that mental modelling is a fundamental form of human reasoning, which can be inferred from numerous experimental findings. The notion of 'mental model' was first used by psychologist K. Craik in 1943. He posits that on many occasions people reason by conducting thought experiments on internal models of real phenomena and that each model is structurally, functionally or behaviourally analogous to the real phenomenon. His idea was popularized in the early 1980s. Today one can distinguish at least three independent research streams within cognitive science employing his idea. The first uses the term to explain the effects of semantic information in logical reasoning. The second uses it to explain empirical findings according to which people, when reasoning about a certain situation described to them, first built a representation of the structure of a situation as a basis for their reasoning. The third does not focus on representation in working memory as the first two, but rather on the long-term representation of knowledge. It attempts to explain experimental results showing that people use organized knowledge structures related to physical systems when trying to figure out manual control systems and devices [17].

Nersessian believes these empirical findings show that mental modelling is a fundamental form of human reasoning, and further supports this hypothesis with findings from research on mental imagery, mental spatial simulation, mental animation as well as embodied mental representation. In her opinion, the main ingredient of mental modelling is simulation, which was already pointed out by Craik. Today proponents of this view speculate that the capacity evolved as a means for simulating possible ways of manoeuvring within the physical environment and for solving problems affecting the survival. The ability to anticipate future events and possible outcomes of actions is undoubtedly highly adaptive and should be possessed by many animals. Human beings are different in that they can create mental models not only from perception, but also from description. Now, in scientific thought experimenting this ordinary capability is simply more refined and adapted to scientific problem-solving or , as Nersessian says, it is put "in service of creative reasoning about nature" [17; p.312].

In conducting a physical experiment, for example, a mental model of a certain physical situation is built, which complies with the constraints related to the representation of physical phenomena. They are determined by experience and current understanding; for instance, by explicit as well as implicit knowledge of the spatiotemporal relations, entities and processes involved, the causal structure and so forth. Such restrictions are causal coherence, spatial structure and mathematical consistency. And the function of stories and diagrams is not only ornamental, but it also conveys restrictions which are consequently incorporated in the model. This enables faster reasoning that often could not even be performed in a propositional form. In addition, epistemic access to the features of representations which are not available in propositional form explains why thought experimenting can lead to new discoveries about the world. Finally, its reliability is plausibly explained given that thought experimenting is based on the mundane cognitive capacity for mental modelling [17].

Nersessian concludes her paper by pointing out that the cognitive-scientific findings on which she bases her account relate to much simpler forms of reasoning than those used in scientific thought experimenting. Additionally, they are mostly about the use of mental models in working memory, but in thought experimenting also information from long-term memory is invoked. Therefore, the next step would be for cognitive scientists to investigate if her hypothesis can be experimentally supported [18].

I agree with Nersessian that research on embodied mental representations looks promising in explaining how stored, tacit information from long-term memory connects to the working memory representation, and that it should be further developed. In general, the functioning of the mind can only be understood in the context of its relationships to the physical body interacting with the world. The importance of embodied cognition for the mental model account of thought experiments is highlighted elsewhere [19].

In that paper, we also ask whether the mental model account can be successfully applied to thought experiments conducted in other domains. We focus on philosophical thought experiments, more specifically, on those by which we are trying to determine whether something is metaphysically possible or necessary. The most common such experiments elicit essentialist intuitions concerning the nature of things, which is not something that we easily glean from our everyday interactions with the world. Characteristically, in conducting such experiments, we do not rely so much on imagination, but on the capacity to conceive, which is more abstract and thus seems more detached from perception than from imagination. The question is whether in these cases the mental model account provides adequate explanation. This is pertinent because it is preferable to have one unified account, but also because within physics itself some experiments are very abstract, and it seems impossible that the proposed situations could be imagined. Such experiments are found in quantum mechanics and the theory of relativity [19].

Nersessian’s tentative answer is that the mental models represent demonstratively, but that this does not mean that vivid, picture-like or movie-like imagery is required. Research in neuropsychology shows that perceptual systems play an important role in imaginative thinking, which is also supported by evolutionary considerations. On the other hand, research also shows that internal representations are sketchier than external pictorial representations. Therefore, in some thought experiments, constructed models can be even more schematic and symbolic. This is the beginning, but further details must be provided in order for the answer to be satisfactory – a task for philosophers and cognitive scientists alike.

REFERENCES

[1] Williamson, T.: *The Philosophy of Philosophy.* Blackwell Publishing Ltd, Oxford, 2007, http://dx.doi.org/10.1002/9780470696675,

[2] Elgin, C.Z.: *Fiction as Thought Experiment*.
Perspectives on Science **22**(2), 221-241, 2014,
http://dx.doi.org/10.1162/POSC_a_00128,
[3] Nersessian, N.J.: *Thought Experimenting as Mental Modeling.*
Croatian Journal of Philosophy **7**(20), 125-161, 2007,
[4] Gendler, T.S.: *Thought Experiments Rethought – and Reperceived*.
In: Gendler, T.S.: *Intuition, Imagination, and Philosophical Methodology*. Oxford University Press, New York, 2010,
http://dx.doi.org/10.1093/acprof:oso/9780199589760.003.0003,
[5] Duhem, P.: *The Aim and Structure of Physical Theory*.
Translated by Wiener, P.P. Princeton University Press, Princeton, 1991,
[6] Galilei, G.: *Dialogues Concerning Two New Sciences*.
Translated by Crew, H.and de Salvio, A. Dover Publication Ltd, Dover, 1954,
[7] Mach, E.: *On Thought Experiments*.
In: Mach, E.: *Knowledge and Error: Sketches on the Psychology of Enquiry*. D. Reidel Publishing Company, Dordrecht & Boston, 1976,
http://dx.doi.org/10.1007/978-94-010-1428-1_11,
[8] Mach, E.: *The Science of Mechanics: A Critical and Historical Account of Its Development*.
The Open Court Publishing Co, Chicago & London, 1919,
[9] Kuhn, T.S.: *A Function for Thought Experiments*.
In: Kuhn, T.S.: *The Essential Tension: Selected Studies in Scientific Tradition and Change*. The Chicago University Press, Chicago & London, 1977,
[10] Brown, J.R.: *The Laboratory of the Mind: Thought Experiments in Natural Sciences*. 2nd edition.
Routledge, New York & London, 2011,
http://dx.doi.org/10.4324/9780203847794,
[11] Norton, J.D.: *Are Thought Experiments Just What You Thought?*
Canadian Journal of Philosophy **26**(3), 333-366, 1996,
http://dx.doi.org/10.1080/00455091.1996.10717457,
[12] Elgin, C.Z.: *Understanding: Art and Science*.
Synthese **95**(1), 13-28, 1993,
http://dx.doi.org/10.1007/BF01064665,
[13] Stuart, M.T.: *How Thought Experiments Increase Understanding*.
In: Brown, J.R.; Fehige, Y. and Stuart, M.T., eds.: *The Routledge Companion to Thought Experiments*. Routledge, London & New York, 2018,
[14] Nersessian, N.J.: *In the Theoretician's Laboratory: Thought Experimenting as Mental Modeling*.
Proceedings of the Biennial Meeting of the Philosophy of Science Association **2**, pp.291-301, 1992,
[15] Miščević, N.: *Mental Models and Thought Experiments*.
International Studies in the Philosophy of Science **6**(3), 215-226, 1992,
http://dx.doi.org/10.1080/02698599208573432,
[16] Norton, J.D.: *On Thought Experiements: Is There More to the Argument?*
Philosophy of Science **71**(5), 1139-1151, 2004,
http://dx.doi.org/10.1086/425238,
[17] Nersessian, N.J.: *Cognitive Science, Mental Modeling, and Thought Experiments.*
In: Brown, J.R.; Fehige, Y. and Stuart, M.T., eds.: *The Routledge Companion to Thought Experiments*. Routledge, London & New York, 2018,
[18] Nersessian, N.J.: *Creating Scientific Concepts*.
The MIT Press, Cambridge & London, 2008,
http://dx.doi.org/10.7551/mitpress/7967.001.0001,
[19] Malec, M. and Markič, O.: *Miščević on Intuitions and Thought Experiments*.
In: Borstner, B. and Gartner, S., eds.: *Thought Experiments between Nature and Society: A Festschrift for Nenad Miščević*. Cambridge Scholars Publishing, Newcastle upon Tyne, 2017.

Interdisciplinary Description of Complex Systems 16(4), 567-578, 2018

PROBABILITY AND METAPHYSICS

Peter Lukan*

University of Ljubljana, Faculty of Arts
Ljubljana, Slovenia

DOI: 10.7906/indecs.16.4.5
Regular article

Received: 10th December 2018.
Accepted: 19th December 2018.

ABSTRACT

In this article I explore the relationship between the concepts of probability and metaphysics. I start by presenting first Popper's metaphysical interpretation of probability and then Suppes' probabilistic metaphysics. Their views are examples of two rare modern attempts to explore metaphysical ramifications of probability. Whereas Popper's approach is less analytic and does not invest any effort in analysing the concept of metaphysics itself, Suppes' approach encompasses both of these aspects. As much as a clarification of the concept of probability employed also a clarification of the concept of metaphysics is needed. For this purpose I first give a short account of Hume's and Kant's positions on metaphysics as these are early instances of the impact that probabilistic inductive reasoning made on metaphysical thought with its newly acquired evidential weight. Then I conclude with Suppes' explicated metaphysical position, indicating the main lines along which a probabilistic metaphysics could be developed.

KEY WORDS

probability, metaphysics, Karl Popper, Patrick Suppes

CLASSIFICATION

JEL: D91

PACS: 01.70.+w, 02.50.Cw

*Corresponding author, η: 5erlukan@gmail.com; -;
Faculty of Arts, Aškerčeva 2, SI-1000 Ljubljana

INTRODUCTION

Within modern science, the concept of probability gained substantial importance at the beginning of the 20th century although its reflection is still catching up with some delay. It became central to an independent science of statistics, which evolved from Galton's heredity studies, and to statistical hypothesis testing established by Pearson, Gosset and Fisher[1], which nowadays forms the base of the modern scientific approach to data analysis. In the 1920's, probability further became of central importance with the advent of quantum mechanics, whose predictions are intrinsically probabilistic in nature. The so-called modern evolutionary synthesis, which took place in the 1930's and 1940's and merged Darwin's and Mendeleev's understanding of genetics into evolutionary genetics and was based on Fisher's statistical analysis of heredity, resulted in a probabilistic-statistical concept of gene and thus also employs the concept of probability [1; p.4]. Von Plato in his work on modern concepts of probability even claims that the modernity of what we understand as modern science (and modern physics specifically) consists in its employing (and developing) the concept(s) of probability: "… the shift from classical to modern probability appears as part of a great movement, the very change from classical to modern science itself" [2; p.26].

As far as philosophy is regarded, the concept of probability was introduced into modern philosophical discourse by empiricist philosophers like Berkeley and Hume, about a century after probability theory emerged with the work of Pascal, Huygens and others[2] [3; pp.57-72, 3; pp.92-121]. Especially Hume can be credited as the one who employed the concept of probability as an integral part of his philosophical thought, but his position was understood more as rejecting metaphysics altogether and relying on habits and probability as a kind of best – but insufficient – surrogates.

Probability is often used as a logical concept like in Bayesian confirmation theory, as a broader epistemological concept like in Bayesian statistics, decision theory and statistics in general, or as a purely mathematical concept. This is in accordance with the variants of probability one can find with interpreters of this concept. Interpretations of probability are usually classified into subjective probability, which entails the so-called logical and personalist variant, and objective probability, which entails the so-called frequentist interpretation and an additional variant, which figures under different names: Guttmann writes about physicalist interpretation [4], Lucas writes about axiomatic interpretation [5], von Plato about measure-theoretic interpretation [2]. The peculiarity of this variant is that it can be applied to continuous mathematical spaces that we meet in physics.

Although there are many interpretations of the concept of probability, one rarely encounters an interpretation of this concept that would have explicit metaphysical implications. It seems that some notion of physical probability is a precondition to develop a metaphysical concept of probability, at least in the modern scientific era. This path of development of understanding can be clearly seen with K. Popper (1902-1994) and his propensity interpretation of probability, which not all authors on interpretations of probability find relevant to study. Popper is one of the most noticeable authors that attributed to the concept of probability a metaphysical dimension. Another such author was P. Suppes (1922-2014). It is mainly these two views that I want to present. Interestingly enough, the authors do not mention each other in their main works on probability and metaphysics, despite being one of the few authors to see the need for such step in philosophical-scientific thought.

POPPER'S PROPENSITY INTERPRETATION OF PROBABILITY

Karl Popper, best known for his work on scientific theories, first presented his propensity theory in his work *The Propensity Interpretation of Probability* (1959) and further developed it in his late work *A World of Propensities* (1994). As some other members of the Vienna circle, with which Popper was affiliated at the time, he too was first adherent of the frequentist interpretation of probability, but he gradually developed a critical position towards it. The main problem he saw with frequentism was its inability to endorse a concept of probability of single events. Frequentists like e.g. von Mises held that probability should be understood as a limit of relative frequencies and therefore only be applicable to a series of events or to collectives, as he named them [6; pp.114-115]. Coherent with his definition in *Probability, Statistics and Truth* (1928), von Mises explicitly rejected objective probabilities of single cases. Talk about objective probabilities of single event is nonsensical from the frequentist point of view while it is pretty common with subjective interpretations of probability.

Popper understood this as a drawback, especially when trying to interpret the wave function employed in quantum mechanics. Since Heisenberg's introduction of the principle of uncertainty Popper felt the need to introduce an adequate objective theory of probability instead of the then prevailing epistemic understanding:

> "But starting with Werner Heisenberg in 1927, a great change occurred in quantum physics. It became clear that minute processes made the clockwork imprecise: there were *objective indeterminacies*. Physics had to bring in probabilities. It was here that I had some severe disagreement with Heisenberg and other physicists, even with my hero, Einstein. For most of them adopted the view that the probabilities had to do with our *lack of knowledge* and, therefore, with our state of mind: they adopted a *subjectivist* theory of probability. In opposition to this, I wished to adopt an *objectivist* theory" [7; pp.7-8].

This is also the reason why Popper departed from von Mises' viewpoint at some stage of his development.

Similarly, he started to break up with R. Carnap, with whom he had collaborated. At first, they agreed they would strictly differ between probability as was used in physical theories and especially quantum mechanics and probability of hypotheses as degrees of their confirmation or corroboration. With respect to this he writes: "we agreed *not* to assume, without strong arguments, that the degree of confirmation or corroboration of a hypothesis satisfies the calculus of probabilities, but to regard this question as … open – indeed, as the central problem" [7; p.5]. Popper thus strictly discerned subjective probability as applied to hypotheses and objective probability as used in physics.

SHIFT FROM COLLECTIVES TO CONDITIONS

It might be beneficial to take a closer look at how Popper introduced his propensity concept. The main shift in his understanding consisted in viewing the outcomes of an experiment as results of repeatable conditions and not of collectives. He illustrates this with throws of two dice, of which one gives equiprobable results and the other does not, i.e. is not a fair die. If by some procedure we always first select among one of the two dice and then throw the selected die, we will collect a long series of results, which von Mises would attribute to a collective or maybe would group them as two collectives. Popper maintains that we can mark such series with a set of generative conditions, in this case the two dice, their properties and properties of the surface on which they land. This is a shift from a multitude to a small set of conditions. A similar proposal was given already by Kolmogorov, but he did not give arguments for this [2; p.220].

In the described sense probabilities are dependent on experimental conditions: "Since the probabilities turn out to depend upon the experimental arrangement, they may be looked upon as *properties of this arrangement*. *They characterize the disposition, or the propensity,* of the experimental arrangement to give rise to certain characteristic frequencies *when the* *experiment is often repeated*" (from Popper, K.: *The Propensity Interpretation of the Calculus of Probability, and the Quantum Theory*. p.67, 1957, as cited in [6; p.116]). This implies that probabilities are a result of the inner structure of the observed object and of the way the environment that acts upon it. Popper stresses this point very clearly in retrospective: "I had stressed that propensities should not be regarded as properties inherent in an object, such as a die or a penny, but that they should be regarded as inherent in a situation (of which, of course, the object was a part). I asserted that the situational aspect of the propensity theory was important, and decisively important for a realist interpretation of quantum theory" [7; p.14]. Borrowing a popular term from quantum mechanics we can say that probabilities are expressions of the entanglement between experimental conditions and the object.

Popper gives the example of throwing a coin on different types of surfaces. When we throw a coin on a perfectly flat surface, the probabilities for both outcomes are 1/2. This implicitly takes into account the properties of the landing surface – this is assumed to be flat, hard, non-reactive, etc. If we design a surface full of vertical slots, in which a coin can land vertically (i.e. on none of the two faces), we get a third possible outcome with its corresponding probability. With coin tossing, probabilities are always also a function of the properties of the surface on which they land, but because we usually throw it so that it lands on a flat and horizontal surface, we abstract the probabilities to represent a property of the coin itself. This is the reason why Popper views the introduction of probability in relation to the set of experimental conditions as legitimate, although the trials are not repeated many times.

Popper thus stresses that even when we have only one trial, there is a propensity for a certain result because of objective experimental conditions entangled with the object of inquiry. It does make sense to talk about the probability of one event as something objective, despite not knowing the outcome in the single case and being able to verify the probabilities only after many trials. The event we are trying to predict inherits the status of objectivity from the objectivity of the circumstances.

In this sense Popper breaks down probabilities of events to objective conditions. Propensities are then weighted possibilities, because they are not equally probable. He tries to conceptually discern the contributions of individual impacting conditions to the overall probability, which he writes in the form of so-called probability of total event, which entails the use of conditional probabilities; for an event B and n conditions A_i we have:

$$P(B) = \sum_{i=1}^{n} P(B|A_i)P(A_i). \qquad (1)$$

If I illustrate this using Popper's example, we can imagine tossing a coin on a surface, which is partially flat and partially carved in the sense described above. For each of the two surface sections the probabilities for the three outcomes are different: the conditional probability of the coin landing vertically on the flat part of surface is, for example, close to 0, while this is not true for the carved section of the surface. The two sections of the surface have different propensities of outcomes.

Frequentists and objective probabilists in general do not regard conditional probability as central in any sense. Conditional probabilities in frequentism are an expression of the so-called reference classes, i.e. subclasses of events. To give an example, in statistical analyses we can improve our probability assignment for mortality of an individual if we take into account additional information about him or her like sex, age, social status etc. This is disturbing for frequentists from a foundational point of view as we may introduce an unlimited number of reference classes and we cannot claim that for a certain object or phenomenon a single objective probability assignment exists (for this see e.g. [6; p.113]). What is a foundational problem for the frequentist interpretation is on the other hand a means of learning and organizing data in statistical inquiry. Recognizing reference classes in a population or data set is a way to discern relevant factors that impact a certain phenomenon.

Bayes' theorem, which introduced conditional probability, is regarded mostly as a pragmatic formula – even Fisher's Maximum likelihood method, which formally employs it, understands it merely as a heuristic tool[3]. Popper's approach is one of the rare ones that offer a proper possible objectivist interpretation of conditional probability as a set of objective conditions intertwined with the object of inquiry. His position is that additional information helps to sharpen our probability assignments for probabilities of single events, which are objectively existent although unclear.

Another author on interpretations of probability, J.R. Lucas, is also of the view that in some cases it makes sense to talk about objective probabilities of individual objects or events, but he stresses that in such cases we view the event as representative of a type of events or objects, not as an individual one. "It is not because we do not know whether an individual has a certain factor present or not, but because we want to consider him not just as an individual, but *qua* member of a population, that we use probabilistic rather than black-and-white, Yes-or-No, true-or-false language" [5; p.188]. We can view the individual event/object as a perfectly determined individual object or as a representative of a wider population. The language of probability is such that it can view the same object as specified to a lesser or greater extent, which is illustrated exactly by the reference classes. From this point of view there is no need to maintain that probabilities of single events do not exist, the question might just be whether they are, strictly speaking, objective. Or alternatively, is objectivity of types of events/objects the same as objectivity of individual events/objects?

It is to be mentioned that different variants of the propensity interpretation of probability evolved. D. Gillies, one of the rare authors who pay special attention to propensity interpretations of probability, mentions besides Popper also D.W. Miller and J.H. Fetzer. The first held a similar view as Popper while the latter developed a theory in which the probability is a function of the complete set of relevant conditions and not of the state of the world, which is a position that Popper later adopted. In this sense, both Fetzer's theory and later Popper's theory have metaphysical pretensions, which is exactly the aspect Gillies criticizes about them [6; p.127]. Propensity interpretations of probability have not become a distinct branch of probability very likely because they did not add any technical insight to the corpus of probability theories, their merits are more of philosophical nature.

THE CONCEPT OF PROPENSITY

The concept of propensity as developed by the late Popper, of which probability is a measure, gets its relevance and strength in conjunction to the concept of causality. Popper was of the view that ordinary causality is a special type of propensity, more precisely, the propensity with probability 1: "Causation is just a special case of propensity: the case of a propensity equal to 1, a *determining* demand, of force, for realization" [7; p.20].

In his book *A World of Propensities*, published during the last year of his life, Popper developed his propensity interpretation to further metaphysical dimensions, of which he was admittedly unaware before: "it was only in the last year that I realized its cosmological significance. I mean the fact that we live in a *world of propensities*" [7; p.9]. In this graduated interpretation Popper does not understand probabilities in an analytical sense anymore, that is, as conditions that generate certain relative frequencies, but as a current state of the world. He ascribes to propensities a fully equivalent status to physical concepts: "Propensities, it is assumed, are not mere possibilities but are physical realities. They are as real as forces, or fields of forces. And vice versa: forces are propensities. … Fields of forces are fields of propensities" [7; p.12].

It might thus be more precise to claim that Popper views propensities as a fundamental concept which underlies also other physical concepts like forces and fields. They are to him what is ultimately ontologically given[4]. Propensities are inherent in situations and in this sense they bridge also the modern gap between subject and object as they mirror the entanglement of a knowing subject with the world. The subject, behaving actively or passively, will always be part of a situation, which he or she can influence with his or her understanding. As Popper puts it:

> "Now, in our real changing world the situation and, with it, the possibilities, and thus the propensities, change all the time. They certainly may change if we, or any other organisms, prefer one possibility to another; or if we discover a possibility where we have not seen one before. Our very understanding of the world changes the conditions of the changing world; and so do our wishes, our preferences, our motivations, our hopes, our dreams, our phantasies, our hypotheses, our theories" [7; p.17].

The picture Popper is painting here is very interactive and also very dynamic. It is not past situations that determine future situations but changing propensities, which influence future situations without uniquely determining them; the future is not objectively determined, but objectively open: "The future is open: objectively open. … The present can be described as the continuing process of the actualization[5] of propensities" [7; p.18].

From this point of view, the present is nothing simple – it becomes a complex entanglement of possibilities which extend into the past and the future. "And in so far as these possibilities can, and partly will, realize themselves in time, the open future is, in some way, already present, with its many competing possibilities, almost as a promise, as a temptation, as a lure. The future is, in this way, actively present in every moment" [7; p.19]. Popper's world of propensities postulates a blurred concept of the present. In such a world, as already mentioned, probabilities are not an expression of human ignorance. "Changing propensities are objective processes, and they have nothing to do with our lack of knowledge" [7; p.18]. Our knowledge itself evolves[6] and probabilities mirror not the impossibility of objective knowledge in general but the impossibility of fully abstract objective knowledge which dismisses conditions and/or the knowing subject, because they are always inherent in the situation.

Last but not least, the concept of propensity introduces also a kind of neutral teleology without any explicit mentioning of this term. Popper rather makes a parallel to Newtonian attractive forces: "It is not the kicks from the back, from the past, that impel us but the attraction, the lure of the future and its competing possibilities, that attract us, that entice us. This is what keeps life – and, indeed, the world – unfolding. (Remember that Newtonian forces too are attractive forces!)" [7; pp.20-21].

Popper's late concept of propensity is not analytical and overshadows the concept of probability, which is basically a measure. In this sense his account is not an interpretation of probability anymore, but a metaphysics of propensities, which has an analytical correlative in the concept of probability and is coherent with it. If we now try to sum up the main postulates of Popper's ontology, we can say that propensities:

- are the ultimate ontological reality,
- are the basis of the entanglement between subject and object,
- are intertwined among them and constitute a blurred present,
- do not determine future states uniquely, but openly,
- are temporally stretched entities that entail a kind of neutral teleology,
- can generate empirically observed relative frequencies and can be measured by a frequentist concept of probability,

- are inherent in situations and preclude the possibility of traditional abstract objective knowledge.

SUPPES' PROBABILISTIC METAPHYSICS

Patrick Suppes' main motivation in his *Probabilistic Metaphysics* (1984) is his dissatisfaction with how probability is incorporated into the western thought in general, not only into science in particular. This state of inadequate accommodation is what has brought by confusion in understanding the concept of probability, which has been subject to the split between subjective and objective interpretations. The classical concept of probability advocated for example by Laplace was compatible with determinism because it entailed the so-called epistemic understanding of probability, i.e. that probability is merely an expression of our ignorance. Suppes sees the need for a probabilistic metaphysics, by which he means a paradigm that would relax the tensions associated with the understanding of the concept of probability.

Suppes criticizes the postulates of what he calls neotraditional metaphysics, which can be found in much of post-Kantian philosophy, in much of contemporary analytic philosophy and in contemporary versions of logical empiricism or logical positivism. He summarizes them as follows [8; p.2]:

N1. The future is determined by the past.
N2. Every event has a sufficient determinant cause.
N3. Knowledge must be grounded in certainty.
N4. Scientific knowledge can in principle be made complete.
N5. Scientific knowledge and method can in principle be unified.

Instead of logical empiricism Suppes advocates something which he calls probabilistic empiricism. In his words:

> "I use concepts of probability to deal with metaphysical and epistemological matters, and I argue for replacing the concept of logical empiricism by that of probabilistic empiricism. But probabilistic empiricism is not meant to have a reductive bias as I conceive it. I shall claim that it is probabilistic rather than merely logical concepts that provide a rich enough framework to justify both our ordinary ways of thinking about the world and our scientific methods of investigation"[7] [8; p.2].

In the introduction, Suppes puts forward a set of metaphysical propositions which form the core of his probabilistic metaphysics [8; p.10]:

P1. The fundamental laws of natural phenomena are essentially probabilistic rather than deterministic in character.
P2. Our conception of matter must contain an intrinsic probabilistic element.
P3. Causality is probabilistic, not deterministic, in character. Consequently, no inconsistency exists between randomness in nature and the existence of valid causal laws.
P4. Certainty of knowledge – either in the sense of psychological immediacy, in the sense of logical truth, or in the sense of complete precision of measurements – is unachievable.
P5. The collection of past, present, and future scientific theories is not converging to some bounded fixed result that will in the limit give us complete knowledge of the universe.
P6. The sciences are characteristically pluralistic, rather than unified, in language, subject matter, and method.
P7. Language learning and performance in their phonological, grammatical, semantical and prosodic aspects are intrinsically probabilistic in character.
P8. The theory of rationality is intrinsically probabilistic in character.

Tenets P4 and P5 stand in explicit opposition to the tenets N3 and N4, respectively, they represent their negation. The rest of the neoclassical tenets are criticized and the main concepts are modified and put into a different context during the various chapters of the book.

It is not my aim to go into details of all the points above, at this point I would rather acknowledge that common traits with Popper's understanding are evident, both in rejecting N1-N5 and in accepting the majority of Ps, certainly P1, P3, P4 and P5. What stands out as a major addition and is of central importance for Suppes is P2. He demands that the concept of matter be changed in a way to contain an intrinsic probabilistic element. He points out that in modern atomism it was recognized that, contrary to ancient ideas about atoms, matter is not indestructible, it can turn into energy. Moreover, particle physics introduced a myriad of particles, that have very short lifetimes and, so to speak, constantly come in and out of existence. Conceptually speaking, the fundamental importance of fields has become more and more prominent, physics is done in terms of continuous properties of fields rather than singular properties of individual particles. Matter is thus continually changing, there is no reason to think that one form, that is, one set of particles is necessarily more fundamental or ultimate than another.

For the mentioned reasons Suppes proposes to introduce Aristotle's concept of matter as pure potentiality without qualification and attributes, as introduced in his *Physics* and *Metaphysics*. As we observe change, there must be an underlying substratum that does not change, but it has no definite shape. In other words, there is no principle of individuation for matter *qua* matter. Viewing matter as pure potentiality, Suppes holds, is not a guidance on the laws of physical phenomena, rather it is a valuable alternative to the atomistic conceptual view. For Suppes this tenet is of central conceptual concern: "my conception of probabilistic metaphysics should be thought of as an extension of the Aristotelian metaphysics of matter and substance" [8; p.7].

Suppes does not view his metaphysical tenets as an exhaustive and exclusive list, he just wants to stress "general propositions that depend on probabilistic concepts, but this is because of their neglect or, in many cases, rejection, as metaphysical assertions by many philosophers just because of their probabilistic character" [8; pp.10-11]. He can accept other metaphysical propositions as sound and important which do not involve probability in any way. Traditional metaphysics has focused on the nature of being, substance, space and time and similar concepts. One way of formulating Suppes' claim is "that the probabilistic character of phenomena is almost as ubiquitous as their spatial or temporal character" [8; p.11].

Suppes is aware that it is uncommon to claim that probability should be regarded as a fundamental metaphysical concept and that he is running counter many philosophical and scientific thinkers. He sees his main predecessor in C.S. Peirce, who put explicit stress on chance phenomena in nature.

WHAT KIND OF METAPHYSICS?

I have reviewed two metaphysical expositions of probability. Perhaps it can be sensed already from what has been explicated so far that the authors have a different conception of what metaphysics is, although there are many common traits. Before discussing this point I will give a short account of the two earliest views on metaphysics impacted by an emerging concept of probability – Hume's and Kant's.

It was already mentioned that the concept of probability became an integral part of philosophy with Hume. Probability is for him the central epistemological concept. When we want to apply our knowledge, due to our imperfect faculties "all knowledge degenerates into probability" [9; p.121]. Probability with Hume does not have any metaphysical dimensions, it is rather an epistemological concept. By his time the concept of probability had gained a reputation of evidential impact compared to its previous role of mere opinion as opposed to knowledge, that is, *doxa versus episteme*, respectively. In thomistic philosophy, knowledge was the domain of demonstrative proof and was applied to questions about eternity and ontology, while probability had a quite different object – opinion; this line could never be crossed [3; pp.21-22]. In Hume's philosophy demonstration applies only to mathematics, the other evidential impact comes from facts, to which we apply probability. Probability thus has evidential impact, but does not have ontological dimensions yet.

Because of his rejection of the notion of substance, Hume is sometimes labelled as anti-metaphysician. It is recognized that his attitude towards metaphysics *via* his rejection of the concept of substance is more disproving in his earlier *A treatise on human nature* (1739) than in his later *Enquiry concerning human understanding* (1748). While at the end of the former he expresses his despair about the possibilities to build a solid foundation for human understanding in the latter he explicitly expresses that it is necessary to cultivate metaphysics and get rid of the old one, which is full of superstition and prejudice: "And must cultivate true metaphysics with some care, in order to destroy the false and adulterate" [10; p.8]. This is strong indication that Hume was at least in search of some kind of metaphysics. He stands at the dawn of the introduction of a new evidential tool and what metaphysics can rely on it is an open question.

An indirect attempt to answer this question was given by Kant with his critical philosophy. He defends the position of transcendental metaphysics as the science of the *a priori* conditions of possible experience and refrains from making judgements about the nature of things *per se*. His aim is still to build an ultimate scientifically grounded epistemological system and with this goal in mind he explicitly rejects the use of probability in metaphysics: "Nothing can be more absurd, than in metaphysics, a philosophy from pure reason to think of grounding our judgements upon probability and conjecture" [11; p.144].

Kant's critical turn, influenced by Hume, changed the way he had tried to incorporate natural sciences into his philosophy. The main work presenting his view on natural sciences within his critical philosophy is his *Metaphysical Foundations of Natural Science* (1786). Before the turn, natural sciences, at the time called natural philosophy, formed the basis of his ontology, while after the turn his metaphysics is based on principles of pure reason, which themselves determine the requirements of natural science proper. These are: scientific cognition must (i) be systematically ordered (ii) according to rational principles and (iii) be known a priori with apodictic certainty, i.e., with "consciousness of their necessity" [12; p.4]. Because properly scientific cognition must satisfy these strict conditions, it requires "a pure part on which the apodictic certainty that reason seeks can be based" [12; p.4]. Since Kant identifies pure rational cognition that is generated from concepts with metaphysics, it follows that science proper requires a metaphysics of nature, i.e. a special metaphysics.

He then specifies that such a metaphysics of nature could consist in either a transcendental part, which discusses the laws that make possible the concept of a nature in general, "even without relation to any determinate object of experience" [12; p.4], or a "special metaphysical" part, which concerns a "particular nature of this or that kind of things" for which an empirical concept is given [12; p.4]. Opting for the latter, his special metaphysics is thus part of his larger system of metaphysics, adopts its main principles but has a special subject matter: it focuses on a few main ideas of pure reason – space, time, force and matter. In this aspect it is aligned with the main ontological concepts that form the backbone of natural sciences.

Kant's exposition of the concepts of space and time is known from his *Critique of Pure Reason*, while his *Metaphysical Foundations of Natural Science* is the main source for understanding his concept of force and matter. Kant had developed a dynamical theory of matter in which properties of matter like solidity and impenetrability are not taken as primitive and self-explanatory, but as derived from the dynamics of particles and attractive and repulsive forces between them. For this purpose he devised in his *Physical Monadology* (1756) pointlike atoms acting in a continuous space, much like Boscovich did just two years later in his very influential *Theory of Natural Philosophy* (1758).

In his *Metaphysical Foundations* Kant incorporated his theory of matter as part of his special metaphysics. The shift that he makes is that matter, in contrast to his *Physical Monadology*, is no longer viewed as simple and indivisible, but as a genuine continuum occupying all geometrical points of the space it fills. The problem of infinite divisibility is solved by explicitly invoking the transcendental idealism of the *Critique*, more precisely the Second Antinomy. Matter is infinitely divisible, but in experience never actually divided. It is only by viewing matter as a thing in itself that we obtain a genuine contradiction. The concept of matter, like the other three, is subject to his conceptual apparatus of the pure concepts of understanding – the categories of quantity, quality, relation and modality. In this sense, the conceptualization of his particular metaphysics is explicitly aligned with his *Critique of Pure Reason*.

Despite its effort to keep up with natural science, Kant's position is not a metaphysics many natural scientists would adhere to nowadays. In fact, it triggered a line of philosophical thought that claimed to have overcome classical metaphysics, attributing to the latter a somehow negative connotation. Also analytic philosophers from the first half of the 20th century, especially the logical empiricists, understood metaphysics as a negative term without real ground and turned rather to a linguistic analysis as basis for ontology. In this perspective, it is not trivial to maintain a relevant modern conception of metaphysics. Indeed, there have been many authors since Hume who maintained either a strong or weak form of the thesis, that metaphysics is not possible[8] [13].

Popper and Suppes are examples of authors who think some form of metaphysics, of course aligned with modern science, is possible. But while Popper limits himself to presenting us a vision of his ontology without reflecting upon it, Suppes explicitly comments on his concept of metaphysics. He understands his attempt as establishing a kind of descriptive metaphysics, which depends upon contemporary science, as opposed to speculative metaphysics.

> "It is sometimes said that Aristotle's *Metaphysics* is a model of descriptive metaphysics, an attempt to organize the most general and at the same time most significant aspects of experience. Such descriptive metaphysics is contrasted with the kind of speculative metaphysics that Kant was so concerned to criticize and eliminate from philosophy. The kind of probabilistic metaphysics I try to develop in these pages is meant to be descriptive rather than speculative. The conclusions I want to reach depend upon the science of the day …" [8; p.3].

Suppes criticizes Kant's opinion, expressed in an open letter on Fichte's *Wissenschaftslehre*, that his system of critical metaphysics rests on a fully secured foundation, established forever. In this respect he is in line with R.J. Collingwood, whose view of metaphysics he presents as one of the most influential 20th century expositions of metaphysics. They are both sceptical that an adequate metaphysics can be developed once and for all and they both hold that what is a metaphysical presupposition changes as science changes. Kant, although he somehow opened a backdoor for empirical concepts via a special metaphysics of natural science, still remained faithful to his ideal of science.

Suppes also does not share the view that metaphysics has its own subject matter and its own methodology. In this respect he differs from both Kant and Collingwood, who understood metaphysics as a kind of science of absolute presuppositions that cannot be regarded as either true or false. Suppes holds no special status for metaphysical statements: "Collingwood represents a retreat from Kant's position by recognizing the absurdity of trying to establish the metaphysical foundations of science once and for all. … But he retains the Kantian view of metaphysics as having a special subject matter and a special methodology. … My position is that there is no sharp delineation of the class of metaphysical assertions" [8; p.9].

It is not within the scope of this article to search for an adequate conception of metaphysics that would encompass the concept of probability. Rather, my aim was to point at some of the important problems that will likely emerge when one attempts to do so. For this purpose I drew on expositions of metaphysics which were influenced by an emerging concept of probability and expositions that rely on a modern physical interpretation of probability, which has been vital for modern scientific development. Moreover, my aim was to investigate approaches that incorporate the views of quantum theory into a scientifically aligned metaphysics.

REMARKS

[1]For an account of the transformation of statistics into an independent science see e.g. Stigler, J.: *History of Statistics: The Measurement of Uncertainty before 1900*. Belknap Press of Harvard University Press, Cambridge & London, 1986.

[2]For an account of the emergence of the concept of probability and its philosophical extensions see e.g. [3].

[3]For this see e.g. Bolstad, W.: 'Comparing the Bayesian and Likelihood approaches to inference: a graphical approach.' In: Reading, C., ed.: *Data and context in statistics education: Towards an evidence-based society*. Proceedings of the Eighth International Conference on Teaching Statistics ICOTS8. The Netherlands: International Statistical Institute, Voorburg.

[4]In this sense, Popper's concept of propensity is very similar to the concept of *vis* employed by the 18th century natural philosopher Roger Boscovich, best known for his 'curve of forces' and for his influence on physicists and the genealogy of the concept of field. Boscovich did not employ a Newtonian concept of force, but understands it more in the sense of tendency or propensity. For this see the article Lukan, P.: *Roger Boscovich and the quantum mechanical combination of mechanical and statistical laws*. Almagest **6**(1), 65-78, 2015.

[5]This formulation is again very similar to Boscovich's understanding of the *vis*, which acts as potential guidance to the actualization of material points with concrete positions.

[6]This is very well in accordance also with the Bayesian probabilistic point of view, which is permanently open to updating belief.

[7]We encounter this commitment to pursue a common basis of ordinary and scientific knowledge also with geophysicist and philosopher of science H. Jeffreys, who was responsible for the revival of Bayesianism in science.

[8]"Let us call the thesis that all metaphysical statements are meaningless 'the strong form' of the thesis that metaphysics is impossible. (At one time, an enemy of metaphysics might have been content to say that all metaphysical statements were false. But this is obviously not a possible thesis if the denial of a metaphysical statement must itself be a metaphysical statement) And let us call the following statement the 'weak form' of the thesis that metaphysics is impossible: metaphysical statements are meaningful, but human beings can never discover whether any metaphysical statement is true or false (or probable or improbable or warranted or unwarranted)" [13].

REFERENCES

[1] Howie, D.: *Interpreting Probability: Controversies and developments in the early twentieth century*.
Cambridge University Press, Cambridge, 2002,

[2] von Plato, J.: *Creating Modern Probability*.
Cambridge University Press, Cambridge, 1994,

[3] Hacking, I.: *The emergence of probability: a philosophical study of early ideas about probability, induction and statistical inference*.
Cambridge University Press, Cambridge, 1975,

[4] Guttmann, Y.A.: *The Concept of Probability in Statistical Physics*.
Cambridge University Press, Cambridge, 1999,

[5] Lucas, J.R.: *The Concept of Probability*.
Oxford University Press, Glasgow, 1970,

[6] Gillies, D.: *Philosophical Theories of Probability*.
Routledge, New York, 2000,

[7] Popper, K.: *A World of Propensities*.
Thoemmes Press, Bristol, 1995,

[8] Suppes, P.: *Probabilistic Metaphysics*.
Basil Blackwell Publisher Ltd., Oxford & New York, 1984,

[9] Hume, D.: *A Treatise of Human Nature*.
Oxford University Press, Oxford, 2007,

[10] Hume, D.: *An Enquiry Concerning Human Understanding*.
Oxford University Press, Oxford, 2007,

[11] Kant, I. and Carus, P.: *Kant's Prolegomena to Any Future Metaphysics*.
Translated by Carus, P. The Open Court Publishing Company, Chicago, 1912,

[12] Kant, I.: *Metaphysical Foundations of Natural Science*.
Translated by Friedman, M. Cambridge University Press, Cambridge, 2004,

[13] van Inwagen, P. and Sullivan, M.: *Metaphysics*.
In: Zalta, E.N., ed.: *The Stanford Encyclopedia of Philosophy*. Spring 2018 Edition, 2018, https://plato.stanford.edu/archives/spr2018/entries/metaphysics, accessed 10th October 2018.

Interdisciplinary Description of Complex Systems 16(4), v, 2018

POVRATNE VEZE I REKURZIJE U KOGNITIVNOJ ZNANOSTI: POVEZNICE METODOLOGIJE I EPISTEMIOLOGIJE

F. Klauser i U. Kordeš

Sveučilište u Ljubljani, Učiteljski fakultet, Centar za kognitivnu znanost
Ljubljana, Slovenija

SAŽETAK

Rad upućuje potrebu da kognitivna znanost istraži svoje korijene i pretpostavke, navodeći tri rekurzivne cjeline: 1.) efekt promatrača, tj. kako promatranje neke pojave utječe na nju, što je cjelina raspoznata u prirodnim znanostima, koju nastojji uključiti empirijska fenomenologija kognitivne znanosti, ali ne izravno sama kognitivna znanost; 2.) ljudska vrsta, tj. kako istraživanja utječu na nas, istraživače i društvo (ljudsko samorazumijevanje), cjelina koja je povezana s efektom promatrača – jer promatranje mijenja promatrano i promatrača, te 3.) opasnosti od pretjerane ekstrapolacije, tj kako se kompleksnost gubi pri prijelazu na razinu objašnjavanja, što su cjeline koje rezutate istraživanja na jednoj razini (npr. pokusi sa štakorima) koriste za objašnjavanje na drugoj razini (pojave unutar ljudskog društva). Na kraju, rad predstavlja četvrtu rekurzivnu cjelinu kao moguće rješenje predhodno nabrojenih: auto-korektivni mehanizam koji omogućava znanosti da rekurzivno ispravlja vlastite pogreške i unaprijeđuje svoj rad.

KLJUČNE RIJEČI

efekt promatrača, ljudska bića, rekurzija, povratne veze, filozofija znanosti

Interdisciplinary Description of Complex Systems 16(4), vi, 2018

POVEZANI UMOVI: KAKO KOGNITIVNA ZNANOST UTJEČE NA REZULTATE SVOJIH ISTRAŽIVANJA

T. Strle

Sveučilište u Ljubljani, Učiteljski fakultet
Ljubljana, Slovenija

SAŽETAK

Polazeći od razlike između ljudske vrste i ostalih živih vrsta, razmatram zatvorene cjeline djelovanja ljudi kroz leću suvremene kognitivne (neuro)znanosti. Nastojat ću pokazati kako je kognitivna znanost koncentrirana na istraživanje, razumijevanja i objašnjavanje ljudske vrste. Nove konceptualizacije ljudskog uma, djelovanja i naše prirode koje se stvaraju (npr. pomoću neuroznanosti), otvaraju mogućnosti za nova tumačenja što to znači biti ljudsko biće. Navedeno, kako obrazlažem, može promijeniti način razmišljanja i ponašanja ljudi i tako promijeniti niz pojava koje kognitivna znanost proučava. Slijedom navedenog, kognitivna znanost može promijeniti svoja (buduća) saznanja. Ovo je posebno točno kad društvo prihvati nove konceptualizacije ljudskog uma te novi načini razumijevanja sebe postaju dio društvenih razmatranja, aktivnosti i strukture.

Potraga za razumijevanjem ljudskog uma, tvrdim, je putovanje u krug, jer ono što možemo „otkriti“ o ljudskom umu ovisi o tome kako kao ljudska bića razumijemo sebe, što u određenoj mjeri ovisi o tome kako znanost razumije nas i kako mi interpretiramo ono što nam znanost iskazuje o našoj prirodi. To me, na kraju rada, vodi do razmatranja kognitivne znanosti kao u bitnome etičkog poduhvata.

KLJUČNE RIJEČI

agenti, kognitivna neuroznanost, ljudska bića, objektivnost, autoreferentnost, društvenost

Interdisciplinary Description of Complex Systems 16(4), vii, 2018

POTRAGA ZA ISKROM INTELIGENCIJE U OSNOVI RAČUNALA: OPSEG I DOSEG RAČUNALNIH MODELA U ISTRAŽIVANJU KOGNITIVNIH POJAVA

T. Kolenik

Institut Jožef Stefan, Odsjek za inteligentne sustave
Ljubljana, Slovenija

Sveučilište u Ljubljani, Učiteljski fakultet, Centar za kognitivnu znanost
Ljubljana, Slovenija

SAŽETAK

Računalno modeliranje ima bogatu povijest uspješne primjene u istraživanju kognitivnih pojava. Otkrića i nove primjene računalnog modeliranja ne prestaju. No, zbog suvremenih paradigmi kognitivne znanosti njegov opseg i doseg stalno su pod posebnom pažnjom. Rad iskazuje opseg računalnog modeliranja navođenjem njegove značajne uloge u napretku kognitivne znanosti i u doprinosu promjene paradigme, kao i navođenjem korisnosti na različitim stupnjevima analize. Iskazana su i ograničenja, neka od kojih nije moguće lako ukloniti, ako ih je uopće mogue ukloniti budući da većinom ne ovise o napretku tehnologije. Dvije su glavne prepreke računalnom modeliranju kognitivnih pojava: istraživačke predrasude, koje se očituju u nužnom prisustvu epistemiološke pozicije istraživača kao i u razmatranim idejama koje se neizostavno uključuju u model; autonomija, neprobojni element žive prirode koji unosi samo-određenost živih bića čime ona stvaraju vlastito značenje u svijetu, što djeluje nemoguće modelirati zato što istraživači uvijek uključuju vlastita značenja u model i modelirane agente, tzv. *PacMan*-sindrom. Računalno modeliranje razmatrano je u svijetlu navedenih ograničenja, posebno kako bi se razmotrilo modeliranje žive prirode.

KLJUČNE RIJEČI

paradigme kognitivne znanosti, računalno modeliranje, povijest kognitivne znanosti, *PacMan*-sindrom, istraživačka predrasuda

Interdisciplinary Description of Complex Systems 16(4), viii, 2018

RAZMATRANJE MISAONIH ZNANSTVENIH POKUSA

M. Malec

Sveučilište u Ljubljani, Filozofski fakultet
Ljubljana, Slovenia

SAŽETAK

U radu razmatram misaone znanstvene pokuse, fokusirajući se na njihovu epistemiološku vrijednost. Prvo naznačujem karakteristike misaonih znanstvenih pokusa, a zatim navodim pripadnu povijesnu podlogu. Naposljetku, kao ilustraciju opisujem dva misaona eksperimenta: Galilejev misaoni pokus o tijelima koja padaju i Stevinov o koso postavljenoj ravnini. Misaone pokuse u fizici uzimam za primjer na temelju kojega se mogu izvući općenitiji zaključci – o misaonim pokusima u drugim prirodnim znanostima, ali također i u filozofiji, matematici i ostalim znanstvenim područjima. Zatim, izlažem Kuhnovu epistemiološku zagonetku kao i neka predložena rješenja. To je blisko povezano s pitanjem o vrsti procesa koji su uključeni u misaonom znanstvenom eksperimentiranju. Zadovoljavajući odgovor mora uzimati razmotriti znanstvena otkrića. Navodim kako je model misaonog računanja najperspektivnije računanje budući da uključuje kognitivnu znanost. Rad zaključujem s dva problema koje računanje treba razriješiti i razmatranjem uloge koju kognitivna znanost pritom može imati.

KLJUČNE RIJEČI

filozofija znanosti, filozofija fizike, misaoni eksperimenti, naturalizam, misaoni model

Interdisciplinary Description of Complex Systems 16(4), ix, 2018

VJEROJATNOST I METAFIZKA

P. Lukan

Sveučiilište u Ljubljani, Filozofski fakultet
Ljubljana, Slovenija

SAŽETAK

U radu istražujem relaciju između koncepta vjerojatnosti i koncepta metafizike. Prvo izlažem Popperovu metafizičku interpretaciju vjerojatnosti, a zatim Suppesovu vjerojatnosnu metafiziku. Njihova su gledišta primjeri rijetkih modernih pokušaja istraživanja metafizičkih grana vjerojatnosti. Popperov pristup manje je analitički te ne razmatra sam koncept metafizike, dok Suppesov pristup obuhvaća oba koncepta. Razjašnjenje koncepta metafizike potrebno je kao i razjašnjenje koncepta vjerojatnosti. U tu svrhu prvo naznačujem Humeov i Kantov stav o metafizici, što su dva rana primjera utjecaja vjerojatnosnog induktivnog promišljanja na metafizičku misao, kojeg potkrijepljuju noviji rezultati. Rad zaključujem Suppesovim razjašnjenjem metafizičke pozicije, naznačujući moguće glavne smjerove razvoja vjerojatnosne metafizike.

KLJUČNE RIJEČI

vjerojatnost, metafizika, Karl Popper, Patrick Suppes

MANUSCRIPT PREPARATION GUIDELINES

Manuscript sent should contain these elements in the following order: title, name(s) and surname(s) of author(s), affiliation(s), summary, key words, classification, manuscript text, references. Sections acknowledgments and remarks are optional. If present, position them right before the references.

ABSTRACT Concisely and clearly written, approx. 250 words.

KEY WORDS Not more than 5 key words, as accurate and precise as possible.

CLASSIFICATION Suggest at least one classification using documented schemes, e.g., ACM, APA, JEL, PACS.

TEXT Write using UK spelling of English. Preferred file format is Microsoft Word. Provide manuscripts in grey tone. For online version, manuscripts with coloured textual and graphic material are admissible. Consult editors for details.

Use Arial font for titles: 14pt bold capital letters for titles of sections, 12pt bold capitals for titles of subsections and 12pt bold letters for those of sub-subsections. Include 12pt space before these titles.

Include figures and tables in the preferred position in text. Alternatively, put them in different locations, but state where a particular figure or table should be included. Enumerate them separately using Arabic numerals, strictly following the order they are introduced in the text. Reference figures and tables completely, e.g., "as is shown in Figure 1, *y* depends on x …", or in shortened form using parentheses, e.g., "the *y* dependence on *x* shows (Fig. 1) that…", or "… shows (Figs. 1-3) that …".

Enumerate formulas consecutively using Arabic numerals. In text, refer to a formula by noting its number in parentheses, e.g. expression (1). Use regular font to write names of functions, particular symbols and indices (i.e. sin and not *sin*, differential as d not as *d*, imaginary unit as i and not as *i*, base of natural logarithms as e and not as *e*, x_n and not x_n). Use italics for symbols introduced, e.g. $f(x)$. Use brackets and parentheses, e.g. {[( )]}. Use bold letters for vectors and matrices. Put 3pt of space above and below the formulas.

Symbols, abbreviations and other notation that requires explanation should be described in the text, close to the place of first use. Avoid separate lists for that purpose.

Denote footnotes in the text by using Arabic numerals as superscripts. Provide their description in separate section after the concluding section.

References are listed at the end of the article in order of appearance in the text, in formats described below. Data for printed and electronic references is required. Quote references using brackets, e.g. [1], and include multiple references in a single bracket, e.g. [1-3], or [1, 3]. If a part of the reference is used, separate it with semi-colon, e.g. [3; p.4], [3; pp.4-8], [3; p.4, 5; Ch.3]. Mention all authors if there are not more than five of them, starting with surname, and followed with initial(s), as shown below. In other cases mention only the first author and refer to others using et al. If there are two or more authors, separate the last one with the word "and"; for other separations use semicolon. Indicate the titles of all articles, books and other material in italics. Indicate if language is not English. For other data use 11pt font. If both printed version and the Internet source exist, mention them in separate lines. For printed journal articles include journal title, volume, issue (in parentheses), starting and ending page, and year of publication. For other materials include all data enabling one to locate the source. Use the following forms:

[1] Surname, Initial1.Initial2.; Surname, Initial1.Initial2. and Surname, Initial1.Initial2.: *Article title*.
Journal name **Vol**(issue), from-to, year,
http://www.address,

[2] Surname, Initial1.Initial2. and Surname, Initial1.Initial2.: *Book title*.
Publisher, city, year,

[3] Surname, Initial1.Initial2.; Surname, Initial1.Initial2., eds.: *Title*.
Editor(s) listed similarly as authors, ed(s).: *Proceedings title*. Publisher, city, year.

If possible, utilise the template available from the INDECS web page.

CORRESPONDENCE Write the corresponding author's **e**-mail address, **t**elephone and **a**ddress (i.e., η).

ISSN 1334-4684 (printed)
http://indecs.eu