Serendipity & Big Data Symposium

February 22-26, 2021

'Sometimes you just have to look up' by Joshua Sortino (Unsplash)

Book of Abstracts

Serendipity & Big Data Symposium

February 22-26, 2021

Book of Abstracts

Serendipity & Big Data Symposium

February 22-26, 2021

The concept of *Serendipity* describes discoveries made by accident and sagacity in which knowledge is produced unexpectedly while in the pursuit of an unrelated finding (Copeland, 2019). Serendipitous discoveries have attracted increasing attention from areas such as philosophy of science and computer science because of the extensive use of Big Data analytics brought about by advances in statistics and artificial intelligence. The possibility of mining and analysing massive amounts of a wide variety of data, from different sources, generated and processed in high speed, has contributed to a growing automation of the process of scientific discovery. In this scenario, relevant questions are: What could be the possible epistemological, pragmatic, and ethical implications of employing Big Data analytics to the process of serendipitous discoveries? Would Big Data analytics enable or constrain serendipitous discoveries?

We have invited people to reflect upon these questions in the **Serendipity & Big Data** symposium. In this book of abstracts you can have a glimpse of what they came up with! We hope we can have fun, learn new things, make new connections, and, who knows, spread the seeds for future friendships from our interaction in this online meeting. Afterall, *wouldn´t friendship be one of the greatest Serendipities of our lives?*

Symposium Program Chair: *Mariana Vitti Rodrigues*
Serendipity Society Symposia Chair: *Wendy Ross*

Sumário

1. Noisy Games we play with AI´s: Critical aesthetics notes on play of musement (Peirce), in algorithmic rationalities, AI.DA and Digital Twins
Alexander Gerner 1
2. A Contractualist Objection to the Use of Filter Bubbles
Bram Medelli 2
3. Big Data and Great Friends playing with Digital Humanities
Claus Emmeche 3
4. Panning for gold: Organizational Problem Solving, Crowdsourcing, and Serendipity
David C. Thompson 5
5. Serendipity and scientific explanation: a philosophical perspective
Edna Alves Souza 6
6. Unexpected Data and the Unknown Theory
Franklin R. Jacoby 7
7. Scientific understanding in the era of Big data
Giovanni Galli 8
8. Inducing serendipity in a big local digital library via hidden agenda keyphrases
Ian G. Kennedy 9
9. The making of "Modelling serendipity in a computational context"
Joseph Corneli 11
10. From serendipity and ignorance to knowledge and understanding in big data practices
María del Rosario Martínez-Ordaz 12
11. Big Data and mechanicism in science: New horizons in the understanding of serendipitous discoveries?
Maria Eunice Gonzalez 14
12. Applied serendipity and information technologies: The corrective role of casual discoveries in judicial activities
Mariana C. Broens & Everaldo T. Q. Gonzalez 16
13. An informational approach to Serendipity: epistemological implications of Big Data for scientific discovery
Mariana Vitti Rodrigues 17
14. Zemblanity and Big Data: The ugly truths the algorithms remind us of
Ricardo Peraça Cavassane 18
15. Enabling data serendipity: The role of conceptual modelling in collecting and analyzing serendipitous data
Ryan Murphy 20

Noisy Games we play with AI´s: Critical aesthetics notes on play of musement (Peirce), in algorithmic rationalities, AI.DA and Digital Twins

Alexander Gerner

amgerner@fc.ul.pt

This exploratory talk first introduces notes on C.S. Peirce´s notion of aesthetics and abduction in which a tension between *play* of musement and AI and Big Data *games* will be discussed. Since the German playwright Schiller, *play* has become an ontological-existential category between uncertainty and contingency, that I treat as noise in between spheres of freedom, purposelessness, passivity, thinking, work (performance), game-rules and calculation. This epistemic and aesthetic notion of noisy and mimetic (Caillois) play proposes a critic of algorithmic rationalities that tend to become invisible by heeding ethical questions of cognitive mimetics that simulate sameness and in a manipulative manner instead of underlining otherness and difference by distinguishing AI agents & AI systems, or aesthetic judgement and reckoning. I will exemplify these thoughts on artificial and Virtual doubles such as *AI.DA,* the artificial artist, and the personalized Avatar doubles of AI Foundation´s Digital Twins.

Keywords: aesthetics, play, abduction, musement, AI.DA, algorithmic rationalities.

Alexander Matthias Gerner, PhD in History and Philosophy of Science (University of Lisbon) Researcher of the R&D Unit CFCUL Centro de Filosofia das Ciências University of Lisbon, where he is Member its Scientific Council and Vice-coordinator of the GI3 Group - Philosophy of Technology, Human Sciences, Art and Society. Gerner teaches "Philosophy of Technology" (PhD Level). He is author of several peer reviewed articles, book chapters and author of the book *Strata: Geophilosophical Notes on Sérgio Costa*. He initiated the former transversal research line at the CFCUL "Philosophy of Human Technology" and researches on interdisciplinary aspects of philosophy of technology, media and policy, green ethics and a new social contract & human aesthetic experience in the age of AI/ML/VR/Biotech for a critic of algorithmic rationality. Since 7/2019 he researches at the FCUL on **Hacking Humans: Dramaturgies and Technologies of becoming other** (HUM+DRAMATECH 12343/2018 position: 2404) that joins questions of human technology, ethics, policy and dramaturgies.

Alexander Gerner's research is financed by the Portuguese national funds via FCT - Fundação para a Ciência e a Tecnologia, I.P., within the scope of the Transitional Standard - DL57/2016/CP CT [12343/2018- in the scientific field of History and Philosophy of Science and Technology, FCUL Position 2404, project by Alexander Gerner: "Hacking Humans. Dramaturgies and Technologies of Becoming Other."
CFCUL strategic project: [UIDB/00678/2020 e UIDP/00678/2020]
Affiliation: Alexander Gerner, Centro de Filosofia das Ciências, Departamento de História e Filosofia das Ciências Faculdade de Ciências, Universidade de Lisboa, Campo Grande, Lisboa, Portugal.

A Contractualist Objection to the Use of Filter Bubbles

Bram Medelli

i.r.medelli@tilburguniversity.edu

O, Brave New World Wide Web!

In this essay, I examine whether remaining in one's online filter bubble is objectionable based on the ethical system of contractualism as explicated by philosopher Thomas Scanlon. I do so by providing philosophical objections based on empirical research of the effects of (big) data collection. As described by activist Eli Pariser, filter bubbles are carefully crafted individual online spaces on platforms such as Facebook, Twitter or Amazon. These companies have created personalized streams of information through extensive data collection on users, granting programmers the ability to sort results based on their users' interests or likes. I argue that these disconnected bubbles of information currently pose a danger to privacy, serendipity and freedom and are therefore reasonably objectionable for contractualists. Firstly, online platforms are often free, but only if users forfeit their right to online privacy so the companies behind it can collect data for advertising agencies. Secondly, limiting online results to mere interests inhibits people from discovering information they did not know they were looking for. Without unpredictability, users can no longer be pleasantly surprised or discover conflicting values and diverse perspectives on the world. The online world becomes limited to what they want to know. Finally, withholding information that is deemed 'irrelevant' for users based on their personal data limits their freedom to act on a

diverse collection of knowledge. I illustrate these objectionable dangers posed by filter bubbles by describing surveillance capitalism, the new logic of accumulation as described by philosopher Shoshana Zuboff. Heeding media ecologist Neil Postman's warnings about entertainment culture, I compare the consequences of a society ruled by surveillance capitalism to writer Aldous Huxley's novel *Brave New World* and to philosopher Robert Nozick's pleasure machine to illustrate its incompatibility with human desires. These capitalists try to control users' behavior for profit. Online businesses aim to extract a behavioral surplus, namely data, from their users in order to anticipate the needs of their users and to keep them engaged on their platforms. Their goal is to create the most enjoyable experience for their users by filtering information into each individual's perfect bubble; not for their users' sakes, but for profit. This process can eventually endanger democracy if it supersedes the social contract as the guiding principle of politics, substituting it with an institutional regime based on (big) data. Finally, as a counterargument, I consider more hopeful aspects of filter bubbles and online platforms as a justification for their use. However, these justifications are insufficient in rebutting the contractualist objections I discuss. Rather, they illustrate the importance of resistance against these unethical elements of filter bubbles so that its potential for good can be realised. I conclude that entering filter bubbles is objectionable from a contractualist perspective.
Keywords: Filter bubbles, Contractualism, Surveillance capitalism, *Brave New World*

Bram Medelli is following the Master studies 'Philosophy of Contemporary Challenges' at Tilburg University in the Netherlands. His interests lie in the philosophy of (online) culture and media, existentialism, and phenomenology, with a general focus on politics. He is currently writing his thesis on how Jean Baudrillard's concept of hyperreality can help us understand the post-truth, post-modern society many of us live in.

Big Data and Great Friends playing with Digital Humanities

Claus Emmeche

cemmeche.ku.dk@gmail.com

As part of a larger project investigating philosophy of science-related questions about the effects of mono- and inter-disciplinary research on ephemeral and under-investigated

phenomena that does not appear to one of 'the great societal challenges' – such as friendship – this presentation will inquire into the use of data-intensive approaches to discover friendship patterns by discussing two very different cases: (1) a study within social cognitive neuroscience on the relation between friendship and brain activity measured as degrees of 'neural similarity' when subjects are exposed to the same experiences of a sequence of different video clips, the result of which seemed to be that the distance between the subjects in a social network 'predicts' occurrence of friendship; and (2) a study, or rather a play, within cultural linguistics using Google's ngram viewer to access trends in the frequency of culture-specific words and locutions related to friendship. The later case is not published but will rather be presented as serendipitous work in progress, the first case is peer-reviewed and published. This first case exemplifies the potential of combining network theory with brain scanning techniques that generate massive amounts of data that need extensive and complicated computational processing before one can begin to look for patterns in the data that may confirm or reject a pre-given hypothesis. So, this approach is not completely data-driven, but rather combining abductive, inductive, and deductive aspects of research, but its operationalisation of the hypothesis to be tested can be critically questioned. The second case brings to the fore some of the challenges of interpreting patterns easily found via the google ngram viewer tool, an option for testing hypothesis about changing prevalence of friendship-related terms (as a kind of 'cultural keywords') and locutions due to cultural change and/or linguistic evolution. Such 'tests', however, need much circumspection and care, as the patterns found may be related to external factors, like the slow increase in the proportion of scientific literature in the overall number of books.

Keywords: friendship, philosophy of science, discovery, social cognitive neuroscience, linguistics.

Claus Emmeche is a theoretical biologist and philosopher of science and associate professor in the Section for the History and Philosophy of Science at the Department of Science Education, University of Copenhagen, Denmark. He teaches philosophy of science for students in chemistry, biochemistry and biology programmes. He has worked with Jesper Hoffmeyer on launching and developing the field of biosemiotics. His current research interest is the philosophy of interdisciplinarity with friendship studies as a case.

Panning for gold: Organizational Problem Solving, Crowdsourcing, and Serendipity

David C. Thompson

info@interstitio.us

The topics of crowdsourcing and open innovation will be presented as tools for organizational problem solving. A specific experience executing a (small) data science crowdsourcing competition in the life sciences will be described [1]. A network-centered framework will then be introduced to orient open innovation activities with respect to each other; such a framework allows for the rationalization of 'crowd-powered' work, across organizational scales, with high fidelity [2]. With this prior work described, a series of observations and general themes relating to coherent organizational problem solving at scale, using open innovation, will be discussed. Specifically, the topics of inclusive problem framing (problem isomorphism and community adjacency) and measurement (including opportunity and context cost assessment). Finally, this work will be explored in the context of the scholarship of serendipity [3–7]. Specifically, if open innovation practices provide potentially novel insights with respect to problem solution, what might this mean, if anything, for our expectation of serendipitous outcomes?

Keywords: Open Innovation, Crowdsourcing, problem solving, organizations

[1] Bentzien, J.; Muegge, I.; Hamner, B.; Thompson, D. C. Crowd Computing: Using Competitive Dynamics to Develop and Refine Highly Predictive Models. Drug Discov. Today 2013, 18 (9), 472–478.

[2] Bentzien, J.; Bharadwaj, R.; Thompson, D. C. Crowdsourcing in Pharma: A Strategic Framework. Drug Discov. Today 2015, 20 (7), 874–883. https://doi.org/10.1016/j.drudis.2015.01.011.

[3] Copeland, S. On Serendipity in Science: Discovery at the Intersection of Chance and Wisdom. Synthese 2019, 196 (6), 2385–2406. https://doi.org/10.1007/s11229-017-1544-3.

[4] Andriani, P.; Ali, A.; Mastrogiorgio, M. Measuring Exaptation and Its Impact on Innovation, Search, and Problem Solving. Organ. Sci. 2017, 28 (2), 320–338. https://doi.org/10.1287/orsc.2017.1116.

[5] Garud, R.; Gehman, J.; Giuliani, A. P. Serendipity Arrangements for Exapting Science-Based Innovations. Acad. Manag. Perspect. 2018, 32 (1), 125–140. https://doi.org/10.5465/amp.2016.0138.

[6] Ban, T. A. The Role of Serendipity in Drug Discovery. Clin. Res. 2006, 8 (3), 10.

[7] Yaqub, O. Serendipity: Towards a Taxonomy and a Theory. Res. Policy 2018, 47 (1), 169–179. https://doi.org/10.1016/j.respol.2017.10.007.

David C. Thompson has made a career out of first making science and then designing tools that make science. His published work in both the physical and social sciences reflect his love of the scientific method as a way of making sense of the world. David has a PhD in Theoretical Chemistry from the University of Cambridge and an MSci in Chemical Physics from University College London.

Serendipity and scientific explanation: a philosophical perspective

Edna Alves de Souza

ednalves@alumni.usp.br

Our objective in this paper is to analyze the concept of Serendipity against the backdrop of the classic debate about scientific explanation in the context of the philosophy of science. An explanation can be understood as a set of well-structured propositions that affirm or reformulate the observations about an object, fact or situation in reality, in a conceptual system adopted by a community that also shares a validation system for what can be admitted as true or false. But what about the 'scientific' epithet? It marks the specificity of the way in which an explanation is given, justified and accepted by the scientific community. Some conditions must be met for a rational explanation to be called 'scientific' and these conditions constitute the method of investigation accepted by the scientific community. Scientific Serendipity, in its turn, commonly understood as involving happy chance, opportunity, coincidence, luck, came to be considered a part or type of creativity. However, its etymological origin, as well as the famous examples of serendipity in the history of science, invites us to review both its common conception and its reduction to creativity. We understand that serendipity is a borderline concept: it refers to the lucky discovery of an agent, apparently by chance, when, in fact, its rational reconstruction reveals the nexus or coherence of an appropriate causal chain. In this sense, we argue that Serendipity is an ability (not a "happy accident") to creatively perceive meaningful links between seemingly disconnected events.

Keywords: serendipity, scientific explanation, scientific method, causality.

Edna Alves de Souza obtained the title of Doctor in Philosophy from the Faculty of Philosophy, Letters and Human Sciences of the University of São Paulo (FFLCH - USP) in 2014. She is currently a postdoctoral researcher in the Postgraduate Program in Philosophy of University of São Paulo State "Júlio de Mesquita Filho" (UNESP, *Campus* de Marília), with financial support from CAPES. She participates in the project "Understanding opinion and language dynamics using massive data", which has financial support from FAPESP. Her main research interest is on the topics of scientific methodology, scientific realism, cognitive relativism, rationality, truth, information, complexity, and Big Data. She has been a member of the Academic Group for Cognitive Studies (GAEC - UNESP) since 1999 and the Interdisciplinary Group of the Centre for Logic, Epistemology and History of Science (CLE - UNICAMP). She has articles and book chapters published in the areas of Philosophy of Science, Philosophy of Technology and Philosophy of Information.

Unexpected Data and the Unknown Theory

Franklin R Jacoby

frjacoby@fastmail.com

Science is full of unexpected discoveries and a driving force behind many of these discoveries is data. Crucial to the functioning of data in this respect is the ability of researchers to recognize anomalies, particularly unexpected results from experiment and observation. But when are data anomalous and what makes it so difficult to identify them? To answer this question, I use contrastive examples from 18th century chemistry, in particular the work of Henry Cavendish (1731-1810) and Antoine Lavoisier (1743-1794). What makes their work particularly interesting is the difference in how they analyzed and discussed a series of similar experiments on calcination, which they conducted independently. I outline some characteristics of data and of researchers that are jointly necessary for our recognition of, and appreciation for, anomalous data. A crucial finding of this work, I will suggest, is that anomalous data, and indeed open minded researchers (a point that Copeland 2019 explored), are important, but not on their own sufficient for bringing about theoretical change. Two further conditions are necessary: 1) the existence of an alternative and available theory that could provide an explanation for the data and 2) researchers who have an awareness of, and appreciation for, the alternative theory. Without these conditions, anomalous data are just

that, anomalies. We can see the importance of these conditions in the work of Cavendish and Lavoisier. Cavendish operated within the theoretical framework of phlogiston theory and lacked an alternative mode of explanation for his experimental data. Lavoisier, working just a few years later, had alternative explanations available and was able to recognize anomalous data as a significant problem for phlogiston theory. Because of this difference in theoretical context, Lavoisier was able to use his experimental work to inform theoretical choices that were unavailable to Cavendish. This shows that scientific discoveries based on unexpected data require the right theoretical context, as well as the right social context.

Keywords: Data, Evidence, History of Chemistry, Theory-Choice

Franklin Jacoby is a Junior Research Fellow at the Institute for Cross-Disciplinary Engagement at Dartmouth College. He works on topics in the history and philosophy of science, in particular on issues of realism, evidence, theory-choice, data, and on understanding the relationship between social and epistemic factors in science. Chemistry and Biology provide especially fruitful case studies for his research.

Scientific understanding in the era of Big data

Giovanni Galli

g.galli13@campus.uniurb.it

It is well known that big data have been reshaping our life. We do business thanks to big data, we interact on social media platforms and generate flow of big data and also the scientific activities are now developed with big data generating and processing tools. It is urgent to present an epistemology of big data in order 1) to warn scientific activities to be careful to the occurring risks using big data, 2) to state whether big data have a role within the theories of scientific understanding and 3) to find out the implications big data analysis sets to scientific realism. I propose some conditions an epistemology of big data must satisfy to answer these tasks and I will address also the conditions required for a theory of scientific understanding including big data technologies in scientific inquiry. According to De Regt (2017) the notion of intelligibility of a theory is central to the epistemic aims of sciences. He offers a Criterion for Understanding Phenomena which need to be tested in the context of scientific enquiry using big data. Some authors deny that big data will be useful to gain intelligibility of theory

and processes involved in understanding specific phenomena. But if science enhances our understanding of empirical phenomena, we must show what this understanding involves. While serendipity conditions seem to be at work in a phase behind the setting-up of a theory, intelligibility conditions have to be satisfied in the subsequent phase, in the process of explaining and understanding. Many authors claim that the use of big data analysis leads to a lack of understanding and they would claim that big data technologies constrain serendipitous discoveries, while I claim for an open view which allows to reconsider big data technologies within the boundaries of scientific activities leading to the development of our understanding of empirical phenomena. Within this view big data analysis could be considered as a tool coherent with scientific practices in which serendipitous discoveries have still a place.

Keywords: philosophy of science, epistemology of big data, scientific understanding, scientific realism

Giovanni Galli obtained his M.A. from the University of Bologna (2015), with Professor Eva Picardi, and he is currently a Ph.D. candidate at the University of Urbino (Research Methods in Science and Technology program). His research interests are connected to the epistemology of big data, scientific realism and scientific understanding. He is also interested in the scientific communication and, as local administrator, he also focuses on the role sciences play for policy making processes.

Inducing serendipity in a big local digital library via hidden-agenda keyphrases

Ian G. Kennedy

dr.iankennedy@gmail.com

This is a report on empirical research conducted on serendipity, using a digital diary. Serendipity is finding some **thing** you desperately need but did know existed. We teach Serendipity and how to induce it. Students will receive our digital library of >300 documents, called Serendip, on the topic of serendipity. It deliberately contains too much data for students to absorb as a traditional reading list. So they will be taught to *use* serendipity and are taught modern software tools, which the prepared courseware teaches initially. The first "attitude" division of the courseware introduces the history and theory via a studying of

Serendip, our library. The second division is an exercise which is a simulation in deliberately doing creative information-encountering with Serendip. The exercise is done on purpose, to *increase* the probability of finding some **thing** that would have been classed as serendipitous. The technique taught relies on including documented hidden-agenda key keyphrases. These are creative phrases which a "recommender panel" consisting of a group of wiser advisers might brainstorm as suggested directions to serendipitously explore. Naturally this requires uncovering implicit knowledge of what students (and research supervisors!) might not have internalised as being an important part of research-business-as-usual. For example the student may have realised that one can and should trace the roots or references back in time. But does the student know the import of following citation bushes forwards in time?? Serendipity key number 3 / 7 opens the door! This is part of our courseware. For students without this knowledge it seems to be miraculous when they accidentally bump into papers citing their keypapers! Yet, it could and should have been research-business-as-usual. So our courseware teaches the importance of making all opportunities for useful "magical", "mystic" events to happen. Then they will "accidentally" find their valuable, unknown-about, key **things**: chapter, informant, or variable. We can also add to this list of things: data, reference, document, theory, technique, technology, component, co-worker, pointers, links, people's names, keywords and other items from Kennedy's hidden agenda list.

Keywords: inducing information encountering, researchers' hidden agenda. key keyphrases, recommender systems.

Ian G.Kennedy PhD MSc(Eng) BSc(Eng) (Wits) is retired from the University of the Witwatersrand. He started researching serendipity in 2012. For more than two decades, he has lectured on how to do research to industrial, business, commerce, accounting, engineering, mining, financial, management, medical CPD, and academic audiences. He has supervised students to successful doctoral degrees – in more than one faculty – and has advised master's students in most faculties. He has six dozen publications to his name and has presented in a dozen countries for more than two dozen academic organizations. He has served as guest lecturer, program committee member, paper referee, book editor, guest editor, and keynote speaker. His ResearchGate Score was 326.88 on 2020-10-29, which only means that he is very helpful and friendly.

The making of "Modelling serendipity in a computational context"

Joseph Corneli

joseph.corneli@hyperreal.enterprises

Together with colleagues, I have been working on and off since 2014 to complete a paper on "Modelling serendipity in a computational context". Here, I will describe the process that our manuscript has been through to date, touching on points of confusion raised by various reviewers over the years, and reprising some of their helpful remarks. I distill certain central conceptual challenges, and outline our latest strategy for addressing them. My presentation includes a minimal amount of technical detail, and focuses instead on framing certain central theoretical issues. One of the novel aspects of the approach in our preprint is that it asks what it would mean for serendipity to be enacted by machines, rather than by humans. This raises a number of potential stumbling blocks. What would it mean to build a system with the capacity for serendipity? Is serendipity an 'experience' or is it ultimately an intersubjectively verifiable phenomenon? Critics have rightly pointed out that serendipity cannot be 'programmed'. However, contemporary computational systems are not simply programmed once and for all and let loose in the world. To the contrary, they are able to learn from their environments, and adjust their behaviour accordingly. This allows these systems to make discoveries that can surprise their programmers. Do any of these discoveries count as serendipity? And are there foreseeable advantages to giving machines the capacity for serendipity, if it was indeed possible? As a first step towards tackling these questions, our preprint begins by aligning various definitions of serendipity from the literature. At this point, it is useful to ask what the category of serendipity is useful for. It is distinct from luck and error: what is it in practical terms? For the purposes of our inquiry a related question is, then, which empirical settings we might look towards to study serendipity in machine terms. For example, have there been working computational models of serendipity around for years already, or is it only starting to become possible to envision serendipitous systems in light of new approaches to AI? How do we wish to approach the possibility of serendipitous discoveries being made by machines: will we approach them with metrics, with cost/benefit analysis, or something else? Clearly, many questions arise in relation to "Modelling serendipity in a computational context". I will connect my answers to the theme of AI ethics.

My presentation will have lessons both for how we think about serendipity, and how we communicate these understandings.

Keywords: artificial intelligence, autonomy, discovery science, ethics, science communication.

Joe Corneli earned a PhD in Computing at the Knowledge Media Institute of The Open University for a thesis on "Peer Produced Peer Learning" (2014). He subsequently held research roles at Goldsmiths and the University of Edinburgh. He is now based at Oxford Brookes University's Institute for Ethical AI, as a Research Fellow. His research work brings AI methods together with social machines.

From serendipity and ignorance to knowledge and understanding in big data practices

María del Rosario Martínez-Ordaz

martinezordazm@gmail.com

The incorporation of big data into the empirical sciences has modified the ways in which scientific knowledge was pursued and scientific methodology was followed. On the one hand, this incorporation has an impressively positive side providing scientists with extremely novel information. Nowadays, scientists are able to, for instance, reach objects that were initially inaccessible to them – either because of their size or because of their remoteness. On the other hand, the outcomes of big data applications often involve high degrees of epistemic opacity about how such outcomes were generated (Cf. Humphreys 2009, Floridi 2012, Leonelli 2014). This leaves scientists having to choose between rejecting the evidential role that these outputs may play or, at the risk of being irrational, relying on them even when ignoring where they come from and how they were obtained. Here I aim at explaining how we can make sense of the continued trust placed by scientists in discoveries (made) by accidents while also ascribing rationality to them. I focus on the cases in which, in big data practices, scientists accidentally arrive at a game-changing novel result but they also, irredeemably, ignore the ways in which this result was actually constructed. I argue that there is an extremely close relation between cases of serendipity in big data practices and a very particular type of ignorance: ignorance of theoretical structure with reliable consequences (Cf. Martinez-Ordaz 2020). I explain how this reliability of the outcomes makes possible the achievement of epis-

temic goods such as objectual knowledge of initially inaccessible objects as well as modal understanding of how these objects (could) behave and relate to one another, all this even if being ignorant of the inference patterns that govern the datasets from which the access to these objects is constructed. I proceed in four steps. First, I address the connections between serendipity and ignorance, and to do so, I adopt an inferentialist account of ignorance (Cf. Magnani 2009; Arfini 2016; Arfini, Bertolotti and Magnani 2018; Martinez-Ordaz 2020). Here, I also address the importance of the game-changing character of serendipity and the possibility of interpreting and explaining, posthoc, the discovery (Cf. Copeland 2015, Baumeister et al. 2010). Second, I argue that some of the epistemic opacities that emerge in big data practices cause that, in some cases of discoveries (made) by accidents, scientists, irremediably, ignore the ways in which such outcomes were actually obtained. Here I am concerned with processes associated to the realization of algorithms in code as well as to the ways in which programs are actually run in particular instances (Cf. Humphreys 2009, Creel 2020). Third, I contend that the ignorance that underlies big data practices in the empirical sciences is ignorance of theoretical structure with reliable consequences; I explain the role that the game-changing character of serendipitous discoveries plays in assuring the reliability of these consequences. Finally, I explain how the reliability of serendipitous discoveries warrants the achievement of certain types of knowledge and understanding in big data practices. I illustrate this with a case study from observational cosmology.

Keywords: serendipity, epistemic opacity, ignorance of theoretical structure, (modal) understanding.

References

1. Arfini S (2016): “Thought experiments as model-based abductions". In: Magnani L, Casadio C (eds) Model-based reasoning in science and technology: logical, epistemological, and cognitive issues. Springer, Switzerland.
2. Arfini, S., T. Bertolotti and L. Magnani (2018): “The Antinomies of Serendipity How to Cognitively Frame Serendipity for Scientific Discoveries", Topoi 39: 939-948.
3. Baumeister AA, Hawkins MF, López-Muñoz F (2010): “Toward standardized usage of the word serendipity in the historiography of psychopharmacology" J Hist Neurosci 4(1):91-98.
4. Copeland S (2015) The case of the triggered memory: serendipitous discovery and the ethics of clinical research. PhD Thesis: Dalhousie University, Department of Philosophy.
5. Creel, K.A. (2020): “Transparency in Complex Computational Systems", Philosophy of Science, 87 (4).
6. Floridi, L. (2012): “Big data and their epistemological challenge", Philosophy & Technology 25(4): 435-437.
7. Humphreys, P. (2009). “The philosophical novelty of computer simulation methods.” Synthese, 169(3), 615-626.

8. Leonelli, S. (2014): “What difference does quantity make? On the epistemology of Big data in biology", Big data & Society 1(1): 1-11.
9. Magnani L (2009): Abductive cognition: the epistemological and ecocognitive dimensions of hypothetical reasoning. Springer, Berlin.
10. Martínez-Ordaz, M. del R. (2020): “The ignorance behind inconsistency toleration", S.I. Knowing the Unknown, Synthese.

María del Rosario Martínez-Ordaz is a postdoctoral fellow at the Federal University of Rio de Janeiro, Brazil. I received my PhD in Philosophy of Science from the National Autonomous University of Mexico, UNAM, in Fall 2019. My doctoral supervisor was Luis Estrada-González. I work mainly in epistemology of science and philosophical logic. My interests extend into history of science and (non-classical) logics. My current research focuses on studying the connections between understanding and the rational use of defective information in science and philosophy. Currently, with Moisés Macías-Bustos, I coordinate the Iberoamerican Philosophy of Science seminar.

Big Data and mechanicism in science: New horizons in the understanding of serendipitous discoveries?

Maria Eunice Gonzalez

eunice.gonzalez@unesp.br

Where is the wisdom we have lost in knowledge?
Where is the knowledge we have lost in information?
(Eliot 1934)

With the advent and development of *Big Data* analytics, new methods of scientific research have been proposed that aim at modelling complex sets of regularities found especially in Biology, Chemistry, Communication, Physics, and Social Sciences. In this scenario, the main question discussed here is: What are the new horizons brought about by *Big Data* in the understanding of serendipitous scientific discoveries? We discuss the pros and cons of *Big Data* methods of analysis used to understand the occurrence of one type of Serendipity, understood as a kind of unexpected discovery in science that involves chance and wisdom (Copeland, 2017). Central to this notion of Serendipity is wisdom, understood as the ability to integrate data, information, knowledge, and ethical intelligence, in accordance with contextual criteria of relevance. We argue that even though *Big Data* analytics may offer

powerful resources and methods to assist the scientist in modelling processes of discovery that involve unexpected regularities, emergent from the dynamics of complex systems, it might lack wisdom in relation to the human-machine interaction. As a case study, we discuss the experiment of Kusne et al. (2020) concerning an autonomous machine learning discovery methodology, applied to functional inorganic compounds. According to the authors, this methodology based on active machine learning allows scientists to learn faster, while improving trust in scientific results and machine learning tools. A provisional answer to our initial question is that *Big Data* resources incremented with active machine learning might help with mechanical *simulations* of scenarios that optimize the process leading towards the occurrence of one type of Serendipity. However, as wisdom seems to be lacking in these simulations, no real novelty is brought about in the understanding of Serendipity, despite the fact that they are rapidly transforming the horizons of the traditional scientific paradigm.
Keywords: Active machine learning, Big Data, Knowledge, Scientific discovery, Serendipity, Wisdom.

Maria Eunice Quilici Gonzalez has a BSc in Physics, and an MSc in Logic and Philosophy of Science. Her PhD thesis “A cognitive approach to visual perception” was completed in 1989 at the University of Essex, UK. She is a founder member of the Brazilian Society for Cognitive Science, and the Research Group on Cognitive Studies at UNESP, the CLE research group on self-organization at UNICAMP, and a Latin-American representative at the Council of the Complex Systems Digital Campus UNITWIN/UNESCO. She is the Brazilian head of the Project Trans-Atlantic Platform (T–AP), Digging into Data: *Understanding opinion and language dynamics using massive data* (OpLaDyn) (FAPESP Number 2016/50256-0). She is interested in the interdisciplinary analysis of philosophical, and ethical issues related to the influence of Big Data and ubiquitous computing in autonomous decision-making processes.

Applied serendipity and information technologies: The corrective role of casual discoveries in judicial activities

Mariana C. Broens
mariana.broens@gmail.com
Everaldo T. Q. Gonzalez
equilici51@gmail.com

The main objective of this presentation is to investigate the possible use of casual, unforeseen, unexpected discoveries of legal errors as a systematic method for correcting flaws in judicial activity and practice of law. In recent years, new social phenomena and scientific and technological advances, especially in information and communication technologies, have imposed a new interdisciplinary agenda on judicial studies and the general practice of law. The use and application of information technologies are common, for example, in various legal activities, from the systematic search for precedents, currently fully automatized, to the use of algorithmic systems to carry out tasks previously reserved for judges or court officials, such as reduction of prison sentences. In particular, information technologies have provided greater fluidity and speed in the circulation of information that benefits the exercise of judicial activity in several ways, either by correcting flaws in the judges' deliberations or by bringing important additional information to such deliberations. Many examples of the corrective role of the fortuitous flow of relevant legal information provided by new information technologies can be offered: correcting errors committed by witnesses in the recognition of suspects in criminal trials, pointing out the partiality of judges or public prosecutors who constantly act according to bias, stereotypes, and prejudices, finding recurrent patterns of sentences in favor of large companies, among others. Considering applied serendipity as the systematic use of accidental, unexpected, unforeseen discoveries in a specific area of investigation, we will try to show that accidental discoveries resulting from the flow of information provided by the new information technologies may offer a method of judicial correction by analysing the causes of recurrent errors in the judicial activity to prevent their repetition.

Keywords: Judicial activity; causal discoveries, information technologies, judicial errors.

Mariana Claudia Broens is an associate Professor of Philosophy at UNESP – University of Sao Paulo State – Brazil, CNPq research fellow and member of the Centre for Logic, Epistemology and History of Science – CLE-UNICAMP – State University of Campinas – Brazil. She teaches Philosophy of Mind, Information Ethics and Theory of Knowledge and has a special interest in the following research topics: Self-organization, embodied embedded cognition, ecological information, information Ethics, and Big Data. She is a member of the Research Group on Cognitive Studies at UNESP, the Research Group on Self-Organization at UNICAMP, and the Complex Systems Society.

Everaldo T. Quilici Gonzalez earned his Bachelor of Law from the Faculty of Law of the University of São Paulo – USP (1984), and later completed at the same university his PhD in Philosophy of Law in 1993 with the thesis *The emergence of the capitalist legal system in Brazil.* He was Professor of the Postgraduate Program in Law at the Methodist University of Piracicaba and Professor at the Salesian University Centre (Unisal). He has professional experience in law practice for over twenty years and, in the academic area of Law, he works with emphasis on Philosophy and History of Law in Brazil, working mainly with the following themes: Philosophy of Law, human rights, history of legal ideas in Brazil, and fundamental rights.

An informational approach to Serendipity: epistemological implications of Big Data for scientific discovery

Mariana Vitti Rodrigues

mvittirodrigues@gmail.com

The increasing availability of massive amount of a wide variety of data, integrated in databases, generated and processed at high speed, has attracted researchers from different areas to the opportunity of data re-purposing. The attempt to find valuable insights from existing data invited researchers to the possibility of mechanically induce serendipity, which is, discoveries made by accident and sagacity. This presentation discusses epistemological implications of the extensive use of Big Data analytics to scientific discovery by investigating the following question: to what extent can the semantic and pragmatic aspects of the process of serendipitous discoveries be automated by applying Big Data analytics? To address this

question, we propose a pragmatic/informational account to serendipity in the context of data-driven sciences. Inspired by Peircean pragmatism, information can be characterized as a process which involves the unveiling of the conceivable consequences of a given object by an attentive agent which allows the adjustment of her conduct. From this pragmatic/informational perspective, serendipity can be understood as a dynamic process grounded on information that allows an unexpected phenomenon, brought about by accidental means, to be perceived as relevant to an attentive agent (or many agents). Our hypothesis is that a pragmatic/informational account of serendipity may shed light to the (im)possibility of an automated detection of relevant patterns by clarifying the role of meaning, and creative chance in the collective and dynamic process of scientific discovery. As a case study, we investigate machine learning applications which aim at inducing serendipity by applying natural language processing algorithms. To conclude we will discuss epistemological implications of Big Data analytics for data-driven sciences by exploring the impact of the automation of scientific practice for serendipitous discoveries.

Keywords: Serendipity, Big Data analytics, information, scientific discovery, data-driven sciences.

Mariana Vitti Rodrigues earned her Bachelor's and a Master´s degree in Philosophy at the State University of São Paulo; Ph.D. in Philosophy of Science and Science Studies at the University of Copenhagen. She is currently, postdoc researcher in philosophy of information and philosophy of technology at the State University of São Paulo funded by FAPESP (n. 2020/03134-1).

Zemblanity and Big Data: The ugly truths the algorithms remind us of

Ricardo Peraça Cavassane

ricardo.peraca@gmail.com

Serendipity refers to a discovery that is unexpected (not only in the sense of surprising, but also in the sense of unpredictable), fortunate (both in the sense of being achieved by chance and in the sense of bringing favourable results), emergent (caused by a conjunction of many different factors in an unforeseeable dynamics, in the form of a happy accident), extrinsic (occurring in the context of an unrelated research or investigation), and active (though

researchers or investigators cannot create serendipity, it can only occur when they are attentive to detail and thus able to spot anomalies, open to recognize such anomalies as potentially valuable, and skilled to determine the meaning of those anomalies); zemblanity, in contrast, is its polar opposite: it is a finding that is unfortunate (in the sense of unfavourable, but not in the sense of brought by bad luck), expected (at least by well-informed researchers or investigators) and/or predictable (and thus not a discovery, but rather a finding researchers or investigators will eventually and inevitably come to), mechanical (caused by a chain or set of factors in a foreseeable dynamics and, thus, not as a bug or by accident, but as a feature or by design), intrinsic (closely related or even internal to the matter under research or investigation), and passive (though researchers or investigators should expect such finding, it becomes zemblanitous by their negligence, ignorance and excessive self-confidence). Serendipitous discoveries may create entirely new objects, methods, or fields of investigation; zemblanitous findings, on the other hand, remind us of problems that we already are (or should be) aware of. We will argue that, while statistical models of social phenomena based on Big Data aim at finding fortunate or unexpected correlations and, with some luck, making truly serendipitous discoveries, they frequently end up resurfacing some old and ugly aspects of our society, in a genuinely zemblanitous fashion. The main example of zemblanity in data science we will explore is that of how the algorithms behind models of face recognition, crime detection, and recidivism risk make evident the already known and by no means surprising racial biases present in our society, especially in the contexts of surveillance, policing and incarceration.

Keywords: Serendipity. Zemblanity. Big Data. Data Science. Racism.

Ricardo Peraça Cavassane is a Ph.D. student from the University of Campinas, Brazil. The provisional title of his doctoral thesis is “Are Big Data capable of predicting the behavior of complex systems?”. He has recently published a paper (in English) with the title “Big Data: Truth, quasi-truth, or post-truth?”. Ricardo is a member of the Research Group on Self-Organization of the Centre for Logic, Epistemology, and the History of Science at Unicamp, and of the Complex Systems Society.

Enabling data serendipity: The role of conceptual modelling in collecting and analyzing serendipitous data

Ryan Murphy
ryan@fulcra.design

An opportunity unlocked by the advent of big data is the value of unexpected insights generated by the high volumes, velocities, and varieties of data in big data projects (Stonebraker, 2012). The generation of these unanticipated insights is a serendipitous process (Copeland, 2019). I define "data serendipity" as an important, estimable quality of a given dataset: its potential to contain unexpected insights. With higher levels of data serendipity, collected data may be more valuable for unanticipated use (Lukyanenko et al., 2019) and reuse (Faniel & Zimmerman, 2011). Approaches to data collection and analysis that increase data serendipity therefore increase the likelihood that data will contain valuable, surprising insights. However, the process of generating insights from serendipitous data is fundamentally constrained by a project's conceptual model (Olivé, 2007, pp. 9-22)—or lack thereof (Parsons, 2018). In conceptual modelling, analysts examine the domain of the data, work to understand the requirements of the project and the nature of that domain, and use this analysis to define the nature of the information that the data project will collect and use. There are two conventional approaches to modelling in data projects, both with their own limitations on data serendipity. The first scenario involves defining a conceptual model first, then collecting data (Parsons, 2018). In this scenario, data valuable for serendipity may never be collected or analyzed if it does not fit the predefined model. The second scenario—data first, model never (2018)—is more common in big data projects where data is often generated of its own accord (e.g., through datafication; (Mayer- Schönberger & Cukier, 2013)). "Raw" data is then analyzed via data science techniques (Provost & Fawcett, 2013) to identify features and patterns. In this scenario, analysis lacks an ontological understanding of the domain and, consequentially, may miss out on important findings (Lukyanenko et al., 2019). In each of these scenarios, effective use of conceptual modelling may facilitate the data serendipity of the resulting dataset. Designing for granularity (Murphy & Parsons, 2020), for instance, ensures that potentially significant details are not lost (Parsons & Wand, 2000). For example, data contributors may want to describe an entity that the dataset was not prepared for—e.g., reporting a type of mosquito never seen in an area to an ecology catalogue. If the

conceptual model only included previously seen species, this data would be lost. Similarly, contributors may want to report that an animal they've seen looks unhealthy. If such an attribute is missing from the reporting interface, it will be lost—yet an index of sick-looking animals in an area could be significant to ecologists. In sum, a dataset's potential for serendipity depends on the conceptual model used to generate or analyze it. As a result, designing data projects for serendipity is possible.

Keywords: data, serendipity, unanticipated use, reuse, conceptual modelling

References

Copeland, S. (2019). On serendipity in science: discovery at the intersection of chance and wisdom. Synthese, 196(6), 2385-2406.10.1007/s11229-017-1544-3

Faniel, I. M., & Zimmerman, A. (2011). Beyond the Data Deluge: A Research Agenda for Large-Scale Data Sharing and Reuse. International Journal of Digital Curation, 6(1), 58-69. 10.2218/ijdc.v6i1.172

Lukyanenko, R., Parsons, J., Wiersma, Y., & Maddah, M. (2019). Expecting the Unexpected: Effects of Data Collection Design Choices on the Quality of Crowdsourced User-Generated Content. Management Information Systems Quarterly, 43(2), 634-647. 10.25300/MISQ/2019/14439

Mayer-Schönberger, V., & Cukier, K. (2013). Big data: a revolution that will transform how we live, work, and think. Houghton Mifflin Harcourt.

Murphy, R., & Parsons, J. (2020). Capturing the Forest or the Trees: Designing for Granularity in Data Crowdsourcing. In (pp. 395-404). Wailea, HI, USA. http://hdl.handle.net/10125/63788

Olivé, A. (2007). Conceptual Modeling of Information Systems. Springer Science & Business Media.

Parsons, J. (2018). Is there a Role for Conceptual Modeling in the Age of Big Data? 37th International Conference on Conceptual Modeling (ER 2018). http://www.nwpu-bioinformatics.com/ER2018/Program.php#content.

Parsons, J., & Wand, Y. (2000). Emancipating Instances from the Tyranny of Classes in Information Modeling. ACM Trans. Database Syst., 25(2), 228-268. 10.1145/357775.357778

Provost, F., & Fawcett, T. (2013). Data Science and its Relationship to Big Data and Data-Driven Decision Making. Big Data, 1, 51-59. 10.1089/big.2013.1508

Stonebraker, M. (2012). What Does 'Big Data' Mean. Blog@CACM. https://cacm.acm.org/blogs/blog-cacm/155468-whatdoes- big-data-mean/fulltext

Ryan Murphy uses design and systems thinking to search for strategic opportunities for change. His emerging research practice examines how data and design may help us

understand complexity, how we may use models of complex problems to design effective strategies in wicked problem-solving, and how changemakers develop and use the skills necessary to engage in all the above.

Ryan is currently a PhD candidate in Management (Information Systems) at Memorial University of Newfoundland. His supervisor is Dr. Jeffrey Parsons. He holds a Master of Design in Strategic Foresight & Innovation from OCAD University and a Bachelor of Science (Honours) in Psychology and Computer Science (Software Engineering) from Memorial University