Introduction to this Volume

Don Berkich
President, the International Association for Computing and Philosophy
Department of Humanities
Texas A&M University - Corpus Christi
6300 Ocean Drive
Corpus Christi, TX 78412

August 13, 2018

The International Association for Computing and Philosophy

The intersection between philosophy and computing is curiously expansive, as the articles in this volume amply demonstrate. New vistas of inquiry are being discovered and explored in a way that neither field alone, philosophy nor computer science, would suggest. An exemplar for the fruitfulness of interdisciplinary work can be found in the International Association for Computing and Philosophy (IACAP).

At its inception in the mid-1980's, Computing and Philosophy conferences were almost wholly devoted to discussing the pedagogical uses of the freshly-deployed desktop computer. Much of that work now seems quaint in light of the many ways in which computers and networks have subsequently become integral to the functioning of the modern university's educational mission. Yet it is interesting to note that philosophers were 'early adopters', front and center in discussions of how best to adapt the computer to help in teaching.

This pedagogical focus would persist through the early 1990's. However, the long history tying philosophical, mathematical, and computational investigations together (the work of Hobbes and Leibniz looms particularly large in this regard) would soon draw philosophical and mathematical logicians, computer scientists, neuro-scientists, ethicists, roboticists, psychologists, information theorists, and philosophers of mind into discussions at turns historical, foundational, and applicable. In the subsequent decades, individual threads of inquiry and subsequent discussions have been woven into the fabric of important research agendas well furthered by papers presented at the 2016 meeting of the International Association for Computing and Philosophy at the University of Ferrara, Ferrara, Italy, June 14-17. Hosted by Professors Marcello DAgostino and (my co-editor) Matteo DAlfonso, IACAP 2016 was graciously sponsored by the University of Ferrara, the Department of Economics and Management, and by the Dipartimento di Studi Umanistici.

The 21 contributions to this volume neatly represent a cross section of 40 papers, 4 keynote addresses, and 8 symposia as they cut across fully six distinct research agendas. Now, I take it that an editor's duty is not merely to describe the ways in which the contributions further the research agendas, but to help frame and better set those agendas for readers and researchers alike. Briefly, then, this volume begins with foundational studies in (1) Computation and Information, (2) Logic, and (3) Epistemology and Science. Research into computational aspects of (4) Cognition and Mind leads neatly into (5) Moral Dimensions of Human-Machine Interaction, followed finally by broader social and political investigations into (6) Trust, Privacy, and Justice. Consider each in turn.

1. Computation and Information

In the abstract it is conventional to formally characterize computation in various extensionally equivalent ways–viz., Turing Machine Computability [Turing, 1936], $\lambda-$Calculability [Church, 1932], Primitive Recursion [Gödel, 1931], or even Abacus Computability [Boolos and Jeffrey, 1989]. Such purely formal characterizations themselves do little, however, to answer a host of important, subtly difficult, and deeply related questions: What is computation? Does computation elucidate mechanism, or does mechanism elucidate computation? Do computational processes describe a natural kind, or is virtually any physical process at some level of description computational in nature? Likewise, what is information? What is the relationship between information, on the one hand, and computation, on the other? Are some or all physical processes inherently informational, or is the notion of information simply a conceptual scheme by which physical processes may be interpreted? Notice that all of these questions, and many other related questions besides, are foundational in precisely the sense in which answers to them are presupposed by such questions as, is the brain a kind of information processing organ? Indeed, this last question motivates the investigations taken up in our first two contributions.

Following Piccinini's excellent survey of the problem [Piccinini, 2017], we want to know which complex physical systems implement computations (artifacts like smart-phones, say, or naturally occurring systems like mammalian nervous systems) formally characterized by some conception of algorithm or other and which do not (a freshly painted wall, a stone garden pathway, or a pile of sand). A chunk of carved quartz crystal would not be a smart-phone, no matter how careful the carving and close the resemblance, presumably because the quartz crystal lacks the smart-phone's capacity to, variously, realize, concretize, or implement computations as defined formally and abstractly. Piccinini dubs this 'the problem of concrete computation'.

If concrete computation implements formal computation merely by it happening to be the case that there exists a state-preserving mapping from formal computational states into physical states [Putnam, 1975], then *pancomputationalism* threatens–to borrow Piccinini's terminology. That is, *any* sufficiently complex physical system–among them the molecules of paint drying on a wall, a pile of sand, or our quartz crystal smart-phone facsimile–will implement formal computation. Thus to the question of whether the brain is computational in nature, we answer, only in the same vacuous sense in which any physical system is computational in nature. Put another way, concrete computation is eliminated as a natural kind on such an account.

Paul Schweizer's "Computation in Physical Systems: A Normative Mapping Account" and Vincenzo Fano, et. al.'s "When is a Computation Realized in a Concrete Physical System?", both informed by Piccinini's sketch of the terrain and his terms of the debate, offer competing analyses of concrete computation in attempting to counter pancomputationalism.

Explicitly echoing Dennett's Intentional Stance, whereby intentional states like beliefs and desires are ascribed as such insofar as doing so yields explanation and prediction [Dennett, 1987], Schweizer proposes that, while there is no concrete computational natural kind–and, thus, any physical process can in principle implement formal computations–we avoid the threat pancomputationalism appears to pose to computationalist theories of mind by virtue of the fact that some physical processes are more suited to our pragmatic interests in concrete computation. Taking the *computational stance*, the physical properties of the smart-phone (in terms of high and low voltages and the complex electrical properties of semiconductors) make it vastly more useful to us than any attempt at treating its quartz crystal analog computationally would do, despite the fact that we could in principle take Schweizer's computational stance with respect to it. Likewise the brain: With respect to explanation and prediction, it is more useful to take the computational stance with respect to neural processes, so computational theories of mind are not undermined by the fact that we could also treat piles of sand as concrete computations.

Where Dennett hedges somewhat on outright rejection of the existence of intentional states (he calls himself a 'quasi-realist' with respect to them), Schweizer sees the computationalist stance as justifying, explicitly, anti-realism with respect to concrete computation. It is not altogether clear, though, whether the possibility that concrete computation boils down to an observer-relative computational stance licenses thorough-going anti-realism about concrete computational systems. After all, as Dennett himself points out [Dennett, 1987] in regards to original intentionality, there must be *some* states of the organism which serve to ground the success of taking the intentional stance in explanation and prediction. Similarly, if taking the computational stance is successful in explaining and predicting the behavior of some physical system, surely the most one can assert is a kind of agnosticism with respect to concrete computation.

That taking the computational stance is sometimes useful in explanation and prediction and sometimes not seems itself a curious fact to be explained. An anti-realist would of course point out that this is question-begging: As Schweizer well argues, our pragmatic interests in taking one stance or another just are the whole of the explanation. Nothing more need be added. Nevertheless, the agnostic may suspect that *some* concrete property of the physical system computationally viewed suits it for computational explanation and prediction.

Oberholzer and Gruner seek to resolve a long-standing debate between Floridi and Fetzer concerning the nature of information. For Floridi–and, for the many reasons Floridi has presented [Floridi, 2007]–information must be true to count as such. Although Oberholzer and Gruner are quick to point out that Floridi allows for information to be both factive and instructional, the grist for the Floridi/Fetzer debate is on information in its factive sense. Oberholzer and Gruner take the factive sense to imply that it is more than merely representative, significant, or faithful: It is necessarily *propositional* given that it (when not instructional) must be true. This, they point out, is out-of-step with classical views of information whereby it i) is conceived as data structured in such a way as to be communicable and thus usable, whether true or false, ii) runs afoul of arguments begun by Fetzer [Fetzer, 2014] and vigorously pursued by Scarantino and Piccinini [Scarantino and Piccinini, 2010], and, writing as computer scientists themselves, iii) raises puzzles for how best to conceive of the stimuli-response relationship involved in engineering behavioral robotic systems.

While their sympathies are clearly on the Fetzer, Scarantino, and Piccinini side of the debate, Oberholzer and Gruner seek a resolution by arguing that it rests on an equivocation over 'information'. Drawing on Frege's distinction between sense and reference, they suggest in light of various arguments by Scarantino and Piccinini that 'information' classically construed is to be understood in terms of the thought a proposition expresses, whereas Floridi's more restrictive, factive notion of information is better suited to the true proposition's reference. Of course, as the Frege Argument [Frege, 1980] shows, all true propositions have the same reference, so it is not clear whether the Fregean distinction does much useful work here. Nevertheless, by adapting the *Afrikaans* terms 'inligting' and 'informasie' to 'enlightation' and 'information', respectively, Oberholzer and Gruner seek to expose what they see as the underlying equivocation by applying 'enlightation' to Floridi's more restrictive sense of 'information' so as to preserve the traditional sense and use of 'information'.

While their distinction between enlightation and information may not settle the Floridi/Fetzer debate, Oberholzer and Gruner remind us that the intersection of computing and philosophy imposes a crucial *computational obligato* on philosophical inquiry. That is, computationally informed philosophical inquiry is also constrained by the computational (logical, mathematical) facts and the way those facts bear on engineering questions, modestly echoing the attempt by logical empiricists to impose an *empirical obligato* on philosophical inquiry. Philosophical disputes on matters of computation and information particularly are constrained by the computational facts and fruitful to the extent that they inform simultaneously the technical and the technological. In a theme woven throughout this volume, the computational obligato serves to undermine, if not entirely eliminate, philosophical flights of fancy which may otherwise be thought to impugn philosophical inquiry.

2. Logic

As with philosophical inquiry into the nature of computation and information, so too is there much to consider at the intersection of computation and logic. From automatic theorem provers to the development of non-standard logics and even to the history and theory of computation itself, logicians both mathematical and philosophical have laid the foundation for computation and worked alongside engineers to develop and refine computer technology. Two contributions to this volume, Khudari's "Modal $\Omega-$Logic: Automata, Neo-Logicism, and Set-theoretic Realism" and Mario Piazza and Marco Pedicini's "What Arrow's Information Paradox Says (To Philosophers)" are rooted in this tradition, yet each in their own way seek broader philosophical implications.

In perhaps the most technically demanding paper of this volume, Khudairi sets out to explicate $\Omega-$logic validity in modal $\Omega-$logic for ZFC set theory. With an account in hand, he uses the flexible notions of coalgebraic logics and automata to come at the same concept from these two other directions. Given the modal nature of $\Omega-$logical validity so-defined, Khudairi pauses to describe its epistemic variation and briefly argue that it has application in virtue of its automata definition to computational theories of mind.

Be that as it may, Khudairi's quarry here is to draw two lessons from the coalgebraic definition of $\Omega-$logic validity for the philosophy of mathematics. First, Khudairi suggests that insofar as $\Omega-$logical validity is purely logical, albeit having modal properties, it justifies neo-logicism in that the conceptual truths of mathematics (at least, insofar as ZFC Set Theoretic truths are concerned) are not stronger or more questionable than the underlying $\Omega-$logic expressing those truths. Second, in a series of arguments, Khudairi makes the case that our very grasp of the concept of (the hierarchy of) sets is itself modal in nature to the extent that we understand the meaning qua intension of the concept. Put more simply, modal $\Omega-$logic supplies the resources we require to flesh-out the intuitive notion of a set meant to be captured by the usual extensional characterization of ZFC set theory but which seems in many ways to exceed it. If so, Khudairi reasons, then we *also* have an argument for mathematical platonism. After all, there is some *thing*–the hierarchy of sets, namely–we grasp in our intuitive understanding.

Piazza and Pedicini likewise seek to draw broader philosophical implications–epistemic, in their case–from Arrow's information paradox [Arrow, 1962]. The context here is economic. Shopping for a new bicycle, I need to have as much information about it as possible to determine whether it is worth the cost. Where the product in question is information itself, I need to have as much information about the information to determine whether it is worth the cost. Yet that is just to have the information, without paying any cost whatsoever. The very act of determining the value of information obliterates its value. So either information cannot be construed as a product in the first place–anathema to a free market capitalist system–or there needs to be state intervention in the free market–also anathema to the free market–by establishing and enforcing intellectual property rights. Piazza and Pedicini point out that the economic literature using Arrow's paradox to justify intellectual property rights assumes I'm faithless: Once I have the information to determine its value, I drop the exchange having got what I set out to obtain in the first place without any cost to me.

To take considerations of intellectual property rights off the table, assume I'm honest. Then, Piazza and Pedicini argue, I nevertheless find myself in an epistemic paradox which echoes the *Meno* paradox: I am either blindly pursuing information, not knowing what it is I pursue, or, already knowing it, have no need to pursue it. Modeling my pursuit of information on Shannon's (cryptographic) Information Theory [Shannon, 1949] offers a way to conceptualize certification or verification of the information without its transmission, thus finding a third alternative to blind or unnecessary pursuit of information. That approach aside, Piazza and Pedicini suggest that the larger lesson to be drawn is that curiosity–for the honest agent, at least–comes inevitably at a cost.

3. Epistemology and Science

As computational resources are more cheaply deployed in empirical inquiry than ever before, whether it be for the sake of collecting, storing, and analyzing vast troves of data or constructing and verifying computational models of complex systems, epistemic questions arise which challenge received views about the very nature of scientific inquiry.

Carnap neatly summarizes the crucial transition in science from the teleological (Aristotelian) 'why'-questions characteristic of science prior to Newtonian mechanics to their abandonment in favor of the 'how'-questions characteristic of current science:

> In the nineteenth century, certain Germanic physicists, such as Gustav Kirchhoff and Ernst Mach, said that science should not ask "Why?" but "How?" They meant that science should not look for unknown metaphysical agents that are responsible for certain events, but should only describe such events in terms of laws. This prohibition against asking "Why?" must be understood in its historical setting. The background was the German philosophical atmosphere of the time, which was dominated by idealism in the tradition of Fichte, Schelling, and Hegel. These men felt that a description of how the world behaved was not enough. They wanted a fuller understanding, which they believed could be obtained only by finding metaphysical causes that were behind phenomena and not accessible to scientific method. Physicists reacted to this point of view by saying: "Leave us alone with your why-questions. There is no answer beyond that given by the empirical laws." They objected to why-questions because they were usually metaphysical questions.
>
> Today the philosophical atmosphere has changed. In Germany there are a few philosophers still working in the idealist tradition, but in England and the United States it has practically disappeared. As a result, we are no longer worried by why-questions. We do not have to say, "Don't ask why", because now, when someone asks why, we assume that he means it in a scientific, nonmetaphysical sense. He is simply asking us to explain some-thing by placing it in a framework of empirical laws. [Carnap, 1998, p. 678]

Carnap thus aligns the shift from why to how-questions with Hempel's covering law model of explanation and prediction [Hempel, 1998a], wherein the explanandum is deduced from an explanans containing laws and statements of conditions. The laws in question, whether interpreted as carrying the necessity of causal law or epistemic regularities [Hempel, 1998b], explain and predict by deductively justifying the explanadum. Whether a given such deduction serves predictive or explanatory purposes has nothing to do with the deduction *per se* and everything to do with the scientist's interests.

However crude this gloss surely is, it suffices to highlight the remarkable shift the expanding use of computational methods and so-called Big Data in science have caused. For just as Carnap invites us to drop why-questions and focus exclusively on usefully answerable how-questions, the sophisticated statistical analysis of massive data sets can identify strong correlations without explanatory bearing. Perhaps, then, we should drop how-questions given the size and complexity of the data-sets in favor of *that-questions*: that events are highly correlated regardless of how they are so related confers predictive power without the unnecessary epistemic burden of explanation.

Indeed, our hand may be forced. More sophisticated algorithms for analyzing large data sets for structure beyond mere correlations which might be employed in the service of answering how-questions themselves confront fundamental complexity constraints on what is feasibly computable. The hard limits of those constraints threaten to render large data sets explanatorily impenetrable.

In "Antimodularity: Pragmatic Consequences of Computational Complexity in Scientific Explanation", Luca Rivelli shows how the limits on what is computable in light of complexity constraints for the large input characteristic of scientific inquiry into large systems–e.g., meteorology, ecology, biology, or neurology–raises serious challenges for the epistemic goal of scientific explanation. Specifically, the received view on explanations of such systems is that the system's global behavior can be functionally decomposed into the interactions of sub-systems or *modules* whose (simpler) functional features contribute to, and account for, the super-system's features. Thus the modular specification of a complex system–our ability, that is, to describe it in modular terms–is essential to explanation, or so Rivelli argues.

Yet drawing on the example of network analysis by computational means reveals that the general problem of modular specification is at best a matter of approximation given complexity constraints on such computations, limitations quickly discovered even for not especially large systems. The upshot, Rivelli suggests, is that some systems may be expected to be of such a scale that we can have no confidence whatsoever in any modular specification given by algorithm. Rivelli dubs this *antimodularity*, whereby a system exceeds the limits of even approximate specification. Such a system, Rivelli warns, is functionally impenetrable and inexplicable insofar as explanation presupposes some sort of modular specification. Far from simply aiding in the pursuit of scientific explanation, complexity constraints on computational analysis reveal the limits of explanation and, in the example of antimodularity, the possibility of the inexplicable.

If the modular analysis of complex systems can be foiled by the apparently hard limits of computational complexity, perhaps a creative enterprise like computer science can further illuminate the problems encountered in pursuit of reductive explanations. In "A Software-Inspired Constructive View of Nature", Russ Abbott argues that the practice of computer science–that is, constructing novel functional properties by piecing together simpler functional elements in novel ways–*constructive creativity* in Abbott's terms–provides a metaphor which can usefully be applied to better grasp the limits and nature of scientific explanation.

Although Abbott takes care to point out that it is no more than an analogy, the parallels he draws between explanation in computer science and science generally are striking, particularly regarding complex systems and reductive explanation.

The computer scientist has at her disposal a raft of libraries and low-level function calls suited to creatively constructing, building-block fashion, new, more complex functionality. Although an explanation of the resulting functionality can be given in principle at the level of machine-code and register calls, it would be useless so far as the computer scientist's interests over constructed functional capabilities are concerned. Of course, the computer scientist has the luxury of designing the low-level functionality in such a way that it *permits* creative construction.

Are there parallels to creative construction in nature? That is, are physical scientists in roughly the same position as computer scientists in their prospects for giving reductive explanations? If so, then we ought to find parallels to the functional compatibility we employ by design in computer science to achieve creative construction. Abbott points to three physical analogs which, without the convenience of having been designed, nevertheless underwrite natural creative construction: negative interaction energy, or the attractive sub-atomic forces binding particles together in atomic structures; autopoiesis, the oft-criticized notion of self-sustaining and replicating structures; and, altogether specific to the biological, evolution itself. Just how far the analogy between the physical and the computational can be pressed is open to question, but it does recognize that there is presumably a point of diminishing explanatory relevance the more basic or fundamental the reduction. Just as the computer scientist is properly concerned with features of the available libraries and not the particular states of the microprocessor's registers, the biologist is properly concerned with the organism's capacities and requirements in light of the structure and function of its constitutive organs, say, and not the properties of the sub-atomic particles they contain.

The implications of big data and its computational analysis are no less important for the prospects of explanation in the social and political sciences than the physical sciences, as Teresa Numerico explores in her cautionary "Politics and Epistemology of Big Data: A Critical Assessment". The *exemplum primi* motivating her analysis is Facebook's now infamous 2015 emotional contagion research on nearly a million of its users which, the authors concluded, showed that positive and negative emotions propagate in social networks [Kramer, 2014]. Setting aside the obvious concerns Numerico raises regarding the issue of informed consent which the Facebook researchers utterly ignored, she rightly points out that the wealth of data we create in our online activities should disallow the social science researcher from any pretense that anonymity is a sufficient protection for subjects and, furthermore, the data itself can only be analyzed by algorithms which in turn embed biases which should disabuse researchers of faith in them as objective research tools. Numerico argues in particular that machine learning algorithms deployed for the analysis of big data are *epistemically opaque* in the sense that the methods leading to their results cannot be verified by human researchers. Epistemically, machine learning constitutes a kind of black box in the social scientific endeavor. Thus, what can be quantified about individuals' behavior in online environments leads, by the sheer vastness of the data, to analyses which neither respect the individual nor are answerable to human researchers.

Whether pointing to the hard limits (vis-a-vis the complexity constraints Rivelli marks) or the soft limits (the epistemic opacity of learning algorithms Numerico describes) of computational methods in scientific inquiry–or, indeed, whether computer science can help illuminate natural science, as Abbott argues– the reliance of scientific inquiry on computational methods begs for greater attention by scientists and philosophers of science alike on the ways in which those methods are informing and changing our understanding of scientific explanation and prediction.

4. Cognition and Mind

Big data and its analysis by computational methods are relatively new techniques in biology, physics, and sociology. Nearly from their respective inceptions, however, cognitive science and computation have been pursued so tightly in tandem as to have sprung from the same philosophical roots. Turing's [Turing, 1936] demonstration of the existence of the Universal Turing Machine, a turing machine that can compute any of the denumerable functions computable by some turing machine, in conjunction with the Church-Turing thesis [Boolos and Jeffrey, 1989] that any function computable by some effective procedure is turing machine computable, almost immediately raised the intriguing question of whether cognitive capacities, suitably decomposed in terms of underlying cognitive functions, were not themselves turing machine computable. Put another way, we want to know whether the class of cognitive functions is wholly contained in the class of turing machine computable functions, where the effective procedures in question are neurological in nature.

The audacious hypothesis of cognitive science is that cognition itself is explicable in computational terms. As computer technology advances, the hand-in-glove fit of computation and cognition creates richer opportunities for the study of cognition, perception, action, and their artificial counterparts. The five contributions here capture the breadth and depth of some of the resulting research agendas.

Tjostheim and Leister explore the philosophical foundations bearing on the empirical dimensions of the study of telepresence in their "Telepresence and the Role of the Senses". Consider, for a somewhat concrete example, the operator of a remotely operated submersible such as those deployed by marine scientists and in underwater oil exploration. Using two cameras on the ROS permits depth perception for close work, but it costs the operator the disconcerting feeling of being at once on the ship and submerged 200 meters, simultaneously. Tactile and olfactory senses align with being ship-bound, visual senses with being ROS-bound.

Vaguely understood as the feeling of *being there*, telepresence is something video game and virtual reality designers are eager to exploit for entertainment purposes by creating richly detailed environments. One can, for example, explore a virtual Los Angeles. Pointing out that our capacity to experience telepresence can shed light on the nature of the cognition of sense perception, Tjostheim and Leister are particularly interested in the role of *affordance* in telepresence. Although much work, they note, remains to fully flesh-out the notion that objects present properties suited to their usefulness in agency, what one feels one can 'do' with the virtual objects one finds in a virtual environment surely bears on telepresence. Here, of course, the video-game industry is deeply engaged in developing virtual affordances in the service of telepresence and story-telling. Commercial interests aside, however, the philosophical implications of telepresence range from support for the spinozistic proposition that comprehension entails, at least for an instant, belief, to the nature of subjective experience and methodological puzzles of phenomenological surveys. That said, Tjostheim and Leister's research is both preliminary and promising. As they point out, conceptual analyses of affordance and telepresence are largely unsettled and rich in opportunities for further research.

M. Christina Amoretti et. al. target conceptual analysis itself in their "Ontologies, Mental Disorders and Prototypes". The logical advantages recommending traditional conceptual analysis by giving individually necessary and jointly sufficient conditions contend with the withering criticisms of the later Wittgenstein and the accumulation of empirical evidence that the role of concepts in cognition is better understood in terms of exemplars or prototypes. Using medical practice with respect to psychological diagnosis as a particularly illuminating example, the authors argue that the typicality conditions used in descriptions of mental disorders found in the fifth edition of the Diagnostic and Statistical Manual of Mental Disorders (DSM-V) offer a treasure-trove for formulating new approaches to concept-mapping and ontology development for application in artificial intelligence.

In this contribution, Amoretti et. al. use the Ontology Web Language-Description Logic (OWL-DL)–containing class, role, and individual constructs– to develop a schizophrenia spectrum formal ontology which, better than previous attempts, captures the syndrome, or prototypical description, the DSM-V employs. The point of the exercise is to demonstrate in application the limitations of traditional conceptual analysis in representing knowledge built from the ground up, as it were, not on necessary and sufficient conditions, but on typicality conditions. In representing the DSM-V, however, the authors explain that OWL-DL is a conceptual procrustean bed. Their proposal here, and the direction of their current research, is to adopt a hybrid approach which brings together traditional conceptual analysis (so far as is feasible) with the geometric format of a conceptual spaces analysis. Instances of a concept are modeled as locations in regions (perhaps overlapping) which correspond to concepts. Spatial characteristics like being centered in a convex region can be used to represent the prototypical instance of the corresponding concept, while distances between locations in a region can represent similiarity relations between conceptually related individuals. The promise of such a hybrid approach to knowledge representation would presumably be wider than the diagnosis of mental disorders and apply to conceptual analysis generally.

The application of computational methods to knowledge representation surely has promise in modeling domains of inquiry, yet the more headline-grabbing application is to modeling human neurology, even to large-scale models which seek to explain and predict the behavior of the entire brain. In "Large-scale Computer Models of the Brain: Is There a 'Right' Level of Detail?", Edoardo Datteri takes up the puzzle of just how much detail in brain-modelling is necessary to gain explanatory traction.

A common assumption on the part of the various, flashy whole-brain modeling projects, Datteri points out, is that explanation of human behavior will only be possible with models of exceptionally fine granularity–down to the level of modeling the functional features of individual neurons. Eliasmith and Trujillo [Eliasmith and Trujillo, 2014], however, argue using the analogy to large-scale climate modeling that there is no right answer to just how fine-grained a model must be: The granularity of the model depends on trade-offs between the questions being asked and the computational resources available. The goal of Datteri's novel and carefully argued contribution is to first (and quickly) dismiss the relevance of the abundance or scarcity of computational resources to the epistemic question of what counts as a sufficient neuroscientific explanation of behavior. He then turns to the difficult task of sussing out how the explanandum dictates the appropriate neuroscientific explanans qua computer modeling, specifically as to just how fine-grained the computer model must be to count as a satisfactory explanation. Put another way, and assuming such explanations involve mechanistic decompositions of complex to simpler neurological mechanisms, how simple must the explaining mechanism (computationally modeled) be to count as a satisfactory explanation?

Datteri's answer is nuanced. In some of the cases Datteri describes, what is to be explained neuroscientifically wholly dictates how course-grained or fine-grained the modeling must be, but in many other cases it does not, contrary to Eliasmith and Trujillo's whole-sale assertion. Where it does not simpliciter, further epistemic principles are required guide modeling efforts–to determine, that is to say, the explanatory adequacy of a given model. What is at stake in these philosophical puzzles is nothing less than determining what counts as a good neuroscientific explanation insofar as those explanations rely on computational methods in modeling neurological systems, as Datteri himself points out.

Computational methods and technology surely have application to modeling in cognitive science and extensive epistemic ramifications, at least for that particular science. That said, information and computation theories more broadly may have implications for long-standing problems in philosophy. d'Alfonso takes up one such problem in his "Virtual Information in the Light of Kant's Practical Reason." Consider the fundamental theorem of deduction,

$$\Gamma \vdash \varphi \text{ iff } \vdash \Gamma \to \varphi$$

The fact that any set of postulates Γ entailing some theorem φ is equivalent to a tautology is epistemically problematic: The entailment appears to be informative, yet the tautalogy, being necessarily true, carries no information whatsoever. Thus no information is conveyed by the fact that the euclidean postulates entail the pythagorean theorem. If there is no information gained, then there is no epistemic gain, either. Nothing new is learned in the proof of the theorem, since it is already contained in–as it were–the postulates given the tautological equivalence expressed by the fundamental theorem.

D'Agostino and Floridi [D'Agostino and Luciano, 2009] propose to rescue the presumed epistemic gain of the entailment by appeal to their concept of *virtual information*. That is, in the course of a natural (as opposed to axiomatic) deduction, temporary assumptions are made and later discharged. These temporary assumptions do briefly convey information and thereby signal epistemic gain in the course of the deduction. As d'Alfonso points out, D'Agostino takes this to be a Kantian solution. Deductions which make recourse to the information carried by temporary assumptions are *a priori*, as are all deductions, but synthetic as well. An axiomatic deduction which makes no such dischargeable assumptions is, in D'Agostino's scheme, *a priori* analytic insofar as it conveys no virtual information at all.

d'Alfonso in this contribution seeks to explain, in Kantian terms, the nature of the virtual information in question. In particular, he argues that while the context of the epistemically gainful deduction is *theoretical* in terms of Kant's distinction between theoretical and practical reasoning, our capacity to employ virtual information depends entirely on our practical, or normative, reasoning. Thus the 'should' in the logic professor's exhortation, "You should temporarily assume P so as to infer Q" represents, in Kantian terms, the practical activity essential to the deduction. The epistemic gain of the deduction is in the practical reasoning deployed in its construction, if d'Alfonso is correct.

d'Alfonso draws on the Kantian distinction between practical and theoretical reasoning to develop the D'Agostino and Floridi notion of virtual information, arguing that Kant's distinction neatly explains the epistemic gain from the practical reasoning demanded by mastering an entailment as opposed to its absence on the tautological–and, thus, theoretical–side of the fundamental theorem's equivalence. The long-standing problem in question is an epistemic one raised by the facts of deduction. The solution proposed here hinges on the epistemic relevance of information understood through the lens of Kant's distinction between practical and theoretical reasoning.

Indeed, speculative philosophy has at least since Descartes and the epistemic turn in philosophy focused on the presuppositions the possibility of knowledge (and its character, objects, etc.) place on cognition. What must the mind be like, philosophers have asked, such that knowledge is possible? Competing answers and vigorous debates about the nature of mind have in effect staked out a sort of solution space for cognitive architecture. These proposals, however, have heretofore been speculative–unmoored from any from any sort of verification or testing.

Complicating matters is the metaphysical dimension of the mind-body problem. This may seem an odd claim to make. After all, the mind-body problem, understood as the problem of nature of mind and the nature of body when the properties and relations of mind and body differ so radically as to be utterly distinct, is ordinarily cast first and foremost–if not wholly–as a metaphysical problem. Solutions to the mind-body problem seek to account for the tremendous gap between mental properties and physical properties by working out ways in which the mental and the physical may or may not be distinct substances. For the dualist, the difference in properties signals a difference in substance. The problem then becomes how to account for the apparent ways in which these different substances interact, which generates a plethora of dreaded philosophical 'isms': interactionism, epiphenomenalism, parallelism, etc.

Surely *part* of the problem is metaphysical. Yet the philosophers taking their cue from Descartes, including Leibniz, Berkeley, Hume, Spinoza, Kant, etc., were at least as keen to understand the nature of the properties of mind which constitute the mind-body problem in the first place. Metaphysics aside, here we find extensive investigations which are doxastic, affective, and agential features of mind–cognitive architectures, in short, which, while frequently covered by the mantel of metaphysics, can usefully be divorced from particular metaphysical presuppositions. Neutral monism, for example, does not entail a humean bundle-of-perceptions view of the mind any more than cartesian interactionism excludes it. For the most part, speculations about cognitive architectures can, as psychology has endlessly demonstrated, be conducted while largely ignoring the metaphysics of the mind-body problem.

Yet if investigations of the nature of cognition are to be more than merely speculative, it must be possible to inquire how they would or would not work in practice. Understood in a functionalist sense, the notion of 'work' here opens the door to a sort of computationalist check on what is possible, cognitively speaking. That is, computation provides a sort of proving ground for philosophical speculation about the nature of cognition.

For instance, the widest gap, that between the empiricist and rationalist traditions in philosophy, is reflected–unintentionally, perhaps–in the gap between deep-learning and logic-based inferential approaches to artificial intelligence. In "A Kantian Cognitive Architecture", Richard Evans finds inspiration in the kantian synthesis of empiricism and rationalism to implement a computational model which builds on the strengths of deep-learning and inferential approaches via a computational counterpart to the kantian synthesis. Though admittedly nascent, Evans' project shows promise on the various tests Evans applies to its current implementation. His approach is encouraging inasmuch as it shows how unsupervised learning can be used on a paucity of data points to more efficiently interpret and systematize the data. As Evans points out, the capacity for such efficient deep learning can in general be obtained by building in domain-specific prior constraints. The trouble is how to build in prior constraints for efficient deep learning which are not domain-specific. Enter Kant. Evans takes Kant's Analytic of Principles to provide a set of general prior constraints, shows how they can be rendered in logical terms which can then be translated into computationally tractable terms, and tests the resulting implementation.

There is much for philosophers and computer scientists alike to glean from Evans' project. In the former case, Evans strives to hover as close as possible to Kant's statements of the principles; his logical analysis of those principles is alone a significant contribution. In the later case, his computational implementation of the logical analysis shows how a synthetic approach to developing general prior constraints on deep learning for the sake of demonstrably improved efficiency can be derived in a principled, yet not domain-specific, way. First to last, Evans' ambitious efforts are a step at making good on promises of the philosophical relevance of computation to philosophy. Of course, Evans' project invites a great deal of further discussion on both the philosophical and computational sides. Yet that is rather the point: The specific moves Evans makes on matters of kantian interpretation, logical rendering, and computational implementation each open broad spaces for further debate, discussion, and collaboration.

The need for collaboration between philosophers, neuroscientists, and computer scientists for our grasp of cognition and the development of a cognitive science is where many epistemic questions will, perhaps, find answers. At the same time, the ongoing rapid development of computer technology has raised at least as many moral and legal normative questions which have drawn the attention of a large share of ethicists, roboticists, and researchers in artificial intelligence. Some of the questions, as we shall shortly see, are rather specific, pointing out pitfalls of implementation approaches that ought ethically be avoided, while others are quite general, raising questions about the very nature of a society which is increasingly characterized by the interactions between human beings and the machines they create.

5. Moral Dimensions of Human-Machine Interaction

In "Machine Learning and Irresponsible Inference: Morally Assessing the Training Data for Image Recognition Systems", Owen King identifies a potential moral normative problem arising from many reasonable applications of image recognition software. Applied in particular to human persons and their visibly discernible behaviors, King argues that we should expect the moral problem of *presumption* to arise. If we think of the function of image recognition software as one of classification based on visual evidence and similarity relations, the presumption problem at its most general threatens insofar as *any* classificatory scheme fails to treat individuals as individual persons and thus fails to respect their moral status as such. Note at the outset that the problem of presumption is in no way unique to machine learning contexts. Indeed, King prefaces his discussion with a number of ordinary cases of human-on-human presumption, skillfully using concrete scenarios to guide intuitions about the moral problem of presumption. In the case of this general sort of presumption, consider the predilection we have with stereotyping, for one example, or our tendency towards confirmation bias, for another. More specific instances of presumption involve classificatory schemes grounded in illicit inferences to an individual's intentions. In the ordinary run of things we frequently must infer intention from behavior, including especially verbal behavior. Flubbing the inference, we react to the incorrectly attributed intention with (variously) resentment, dejection, confusion, humiliation, etc. Of course, image recognition software doesn't react, but it does classify and can be expected to be at least as fallible in the inferences drawn as we find ourselves to be. The problem comes in not discovering that presumption qua illicitly inferred intention has occurred and, as a result, the individual's autonomy is unduly restricted, albeit algorithmically.

King distinguishes between a *modular* approach to presumption and an *ingrained* approach. On the modular approach, cases of presumptive inference are (somehow) identified and excluded post classification, whereas the ingrained approach seeks to avoid the presumptuous classification in the first place. Rejecting the modular approach as the obviously question-begging alternative it is, King focuses his efforts in this contribution on how training data can be so restricted as to ensure "responsible judgments"–that is, non-presumptuous or at least minimally presumptuous classifications.

Responsible judgments are one problem, responsible *agents* quite another. As roboticists engineer increasingly sophisticated general applications systems, we confront the thorny problem of whether and how to assess their moral responsibility–viz., moral praiseworthiness or moral blameworthiness. In her engaging and well-argued "Robotic Responsibility", Anna Wilks explores a possible middle ground between two manifestly implausible, yet apparently exhaustive, views of robot moral responsibility. On one hand, we might view robots as either morally neutral, morally innocuous, or perhaps (at most) moral innocents, insofar in each case as they merely express the moral agency of their designers and users, being themselves at most simple tools. Surely, though, there is nothing simple about the contributions a robot makes to its environment, operating as it does quite independently. On the other hand, we might view robots as fully morally responsible agents, which seems absurd both *prima facie* and especially after reflecting on the kantian conception of moral agency qua rational beings capable of authoring and motivating their own agency in light of recognizing the moral duty entailed by moral law. Note that the kantian account of moral agency is notoriously demanding. Less demanding accounts can be given, but in none of them does full robotic responsibility survive the fact of their having been through and through designed, engineered, programmed, and trained by moral agents who seek only to extend their own agency, and the kantian account is anyway Wilks' preferred starting point.

Wilks finds in Margaret Gilbert's work on joint commitment [Gilbert, 2014] the grounds for a middle position between these two positions which holds that there is a sense of collective moral responsibility which is not strictly reducible to the moral responsibility of individuals acting in concert. Thus collective moral responsibility is neither a linear nor a diffuse–that is, some other functional–aggregate of individual assignments of responsibility in a group effort. Wide moral responsibility in this sense presupposes a collective moral agency to which individuals contribute their efforts. As Wilks notes, this necessarily stretches our ordinary conception of moral responsibility inasmuch as irreducibility entails a standalone notion of group moral responsibility. Individual contributions to collective moral agency need not, however, presume full *individual* moral responsibility for all of the members of the collective. At most, Wilks argues, some or even just one member must be fully morally responsible, while the rest require only a degree of intelligence and autonomy for their actions to count towards the collective agency and, thus, moral responsibility. As Wilks puts it, "[i]t is not necessary for the doctor to be also a nurse, and a social worker, and an extremely powerful computational machine. Why then should we require that the machine be a doctor or a social worker, or even a person? Each one contributes something as an individual, but the responsibility for the overall task is ascribed to the whole group–since the utlimate deliberation and actions taken involve the joint commitment of the collective."

Robot colleagues, as it were, cannot be viewed as genuine moral agents if our sense of moral agency is individual, but that does not exclude the necessity of viewing them as potentially important members of a moral community and contributing to communal moral responsibility in their various ways. Of course, much more needs to be said about the degree of intelligence and autonomy required to be so viewed as a member of the community and not merely a tool for its use, yet as Wilks concludes, we at least begin having the altogether necessary conversation of just how we should view the incorporation of sophisticated robotic systems in collective expressions of agency and, ultimately, in assessing group moral responsibility.

A narrow application of robotics which nevertheless carries broad social implications concern Jason Borenstein and Ronald Arkin in their "Robots, Ethics, and Intimacy: The Need for Scientific Research". Sketching the conceptual terrain as best as can probably be done given the nature of the application in question, the authors point to the dearth of answers to important questions regarding the role of social robotics, particularly ones deliberately designed to emulate intimate relationships in such a way as to induce strong feelings in users of attachment and love. Although the prospect of roboticists inquiring seriously about the nature of intimate and loving relationships may strike one as peculiar, science fiction literature and film has long speculated that robots will eventually be so sophisticated as to be capable of perfectly imitating participants in intimate relationships. Still science fiction at this point, the prospect is made more pressing by the propensity humans have to adopt and form relationships–frequently very important social relationships–with non-human animals and, of greater relevance, inanimate objects. Construed as animate objects, robots are readily suited to exploiting this tendency, thereby impacting important aspects of human life, our capacities to value, care, form attachments, and even love.

By carefully articulating a number of important research questions, Borentsein and Arkin lay out an ambitious research agenda for roboticists, philosophers, psychologists, and sociologists to pursue in light of the progress on the engineering front of intimate robotics. The questions tend toward the consequentialist, asking after possible sources of utility and disutility in the application of robotic systems to socially intimate contexts. For example, what psychology (beliefs, desires, and intentions) can the use of intimate robots be expected to engender in the user? What of the user's well-being, psychological and otherwise, particularly in light of the possibility that intimate robots may tend to push out ordinary human relations? Consider in this regard the development of carebots to provide care and companionship for the elderly and infirm, which can only be expected to limit opportunities for human interaction. Perhaps more troubling, how will the prospect of forming intimate attachments with socially sophisticated robots impact our expectations, understanding, and perhaps even capacity for forming ordinary human relationships? The authors remind us that there is a dearth of research on these and many other questions besides, while also pointing out that robotics entrepreneurs will not be reticent to develop and exploit market niches where social robots will be welcomed, for good or ill.

Not all human interactions with robots entail (one way) intimacy or even continuous involvement. Indeed, most of us will only briefly interact with robots as they are deployed by developers, owners, and users on behalf of organizational–including government, corporate, and medical–interests. Frances Grodzinsky et. al.'s "Applying a Social-Relational Model to Explore the Curious Case of hitchBOT" use the example of hitchBOT–the social-media 'hitchhiking' robot star whose summary destruction in Philadelphia seems once and for all to have settled the experimenter's question, "can robots trust humans?"–to argue that robot owners bear responsibility for robot-human interactions even when not present at those interactions.

As the authors explain, what interests them particularly about hitchBOT is that, unlike non-social robots without a shred of 'hooks' to encourage anthropomorphizing, including perhaps hospital delivery robots, vacuuming robots, or even the ubiquitous automatic teller machine, hitchBOT was specifically designed to induce friendly feelings and feelings of trust towards it. That is, if benign, it was nevertheless designed to be deceptive, even though a fair part of that deception included a social media presence. Drawing on research on the moral dimensions of social robotics understood in terms of interactions and social roles, the authors specify the special obligations the designers of unaccompanied robots incur, particularly as the robot interfaces become increasingly sophisticated so as to converge on ordinary human interface–conversationally, say, or visually.

The question of whether and how to consider the moral status of robots need not, however, be solely grounded in terms of social-relational models. Migle Laukyte argues for an altogether different approach to these questions, one derived from considerations in environmental ethics, in his "Against Human Exceptionalism: Environmental Ethics and Machine Question". Specifically, Laukyte starts from the position in environmental ethics known as 'Deep Ecology', which denies any position of special moral privilege–such as being a person, say–in the complex ecological web. Thus Deep Ecology entails a kind of thorough-going ecological egalitarianism, although it is unclear whether the egalitarianism in question extends to geographic features like lakes, mountains, or fjords.

Laukyte makes an important point in noting that our 'environment' has long been, and is being with exponential rapidity, enriched with robots constituting more or less autonomous nodes in what can be viewed as an (albeit artificially constructed) ecological web. This stretches our ordinary understanding of 'ecology', and Laukyte's argument is, in part, to make plausible just such an extension so as to provide Deep Ecology purchase on the problem of the moral status of robots. Setting the stage, his focus is not so much on the obvious example of the autonomous robot reacting to and contributing however it may to the ecological web in question, but on a much wider notion of artificially intelligent agents, regardless of their engineering features or even whether such agents are physically instantiated in some specific robotic form or other.

Attempting to meet the obvious rejoinder, that this application of the central theses of Deep Ecology unacceptably distorts 'ecology' to include both natural and artificial agents, grounded in part perhaps by virtue of the fact that artificial agents are, at least, non-living, is front and center in the challenges Laukyte takes up. His argument here takes place on two fronts: First, the capacities of artificial agents–'mindclones', as Laukyte dubs them–make them difficult to distinguish from natural agents given their success in mimicing behavioral reportoires; second, our ecology in any case has been subject to various substitutions and permutations by selective cultivation and breeding since the development of agriculture. Thus it would be arbitrary to exclude artificial agents from the ecological web, a point well worth considering regardless of any further claims on behalf of, or following from, Deep Ecology.

An issue Laukyte does not directly address is the construction of artificial–that is to say, virtual–environments as a whole, populated with artificial agents (non-player characters, or NPC's) and avatars of human agents. If the notion of an ecology can be stretched to include artificial agents, perhaps it is but one further step to admit an entire virtual environment as a wholly constructed ecosystem. Regardless, the normative features of those environments, particularly for individual human agents represented by avatars in the virtual, is an area of considerable debate. In "The Ethics of Choice in Single-Player Video Games", Erica Neely takes up the puzzle of the moral status of actions in virtual environments, arguing that it is intelligible to speak of harms and benefits caused by the decisions of users (players) and designers alike because of the effects those actions have on them, inside or outside of the virtual environment in which the actions are taken.

Neely draws a distinction between the intravirtual (within game) effects a player's choices in game might have on him or her and the extravirtual effects of those same choices and the potential carryover into extravirtual choices. For one example long discussed in the popular hand-wringing over violent video games and first person shooters, consider that the brutalizing choices the video game player makes while immersed in a given virtual environment may lower social inhibitions to making harmful real-world choices. Neely's argument, however, is far more subtle than this sort of straight-line sorites.

The intravirtual choices a player might make in a virtual environment could encourage the player to entertain or make unethical choices in other contexts, depending in part on how designers of virtual environments encourage or discourage such choices, which in turn depends on the sort of rewards system the designers have built in to the virtual environment. Neely's point, however, is that virtual environment designers seek to make intravirtual choices as close in nature to their extravirtual counterparts as possible–to make them, in terms of the player's experience, 'real' choices with 'real' consequences. Virtual environments thus gain traction with players insofar as they exhibit realism in approximating the gravity of extravirtual choices for players. How well designers themselves grasp the moral import of the degree of such realism they manage to incorporate so as to engage their players raises the moral stakes of the creation and use of virtual environments. The stakes can be for moral ill, as Neely notes. Yet, importantly, she also argues that it can be for moral good, perhaps as players learn in the virtual environment more sophisticated methods of moral deliberation. The onus at least in part is on the sensitivity of designers to such issues, but it also rests with the game player and the lessons they draw from being immersed in the virtual environment.

6. Trust, Privacy, and Justice

Finally, the internet itself and the various social networks it contains are a longstanding source of normative puzzles, particularly as they are the perfect targets for big data collection, its harnassing by algorithmic analysis for purposes of pinpoint profiling, categorizing, and generalizing, and the subsequent exploitation of these analyses by private, corporate, and government interests. Just as we ourselves make use of the networks and services therein provided, those entities and individuals providing them make use of us, often in manipulative, exploitative ways which succeed in part by virtue of their relative invisibility from the network user's perspective. The wealth of scholarship in response has been nothing short of a renaissance in the study and defense of human rights to rival that of the enlightenment, up to having numerous political rammifications. Front and center to these discussions are questions of identity, autonomy, privacy, trust, and justice.

In "Obfuscation and Good Enough Anonymity", Tony Doyle argues in favor of obfuscation–that is, the deliberate muddying of the informational waters, as it were, by the individual's use of misleading or ambiguous data. With characteristic clarity, Doyle draws a straight line from obfuscation to human well-being: Cleverly used, obfuscation can foil big data analytics in such a way as to preserve anonymity and thereby protect privacy as a way to defend, in turn, against manipulation and promote individual autonomy, where individual autonomy tends to promote individual well-being.

There are many caveats and exceptions to be drawn at each stage of Doyle's argument. The use of big data analytics need not be a zero-sum game. Consider their use as simply a matter of efficient and effective discrimination, and note that discrimination *per se* can be just or unjust, depending on the basis for discrimination. Moreover, since the data in big data analytics largely consists of the digital imprint ones online behavior in social and other networks makes, an argument can be made that the resulting discrimination neatly avoids the superficial bases–skin color, say, or height, or attractiveness–we otherwise tend in practice to use, wholly unjustly, to draw distinctions between people.

Yet given the fact of big data analytics and despite its many potential benefits, its service for altogether powerful and particular (though not necessarily malevolent, Doyle is quick to note) interests at the expense of broader social interests and specific personal interests tips the scale of the potential harms of obfuscation–of which there are many Doyle chronicles–in favor of the singular but overridingly important benefit of individual well-being. Or so Doyle argues. Of course, his is a practical as much as it is a philosophical argument. Doyle's prescription of obfuscation requires effort at evasion and something like subterfuge on the part of the network participant to secure even partial anonymity, at which proposal the tendency to throw ones hands up in surrender to the overwhelming force of big data analytics is understandable. After all, we've grown inured to a loss of privacy, just as we grow accustomed, horribly enough, to the potential and sometimes fact of exploitation made possible by our online presence. If 'going off the grid' is not feasible, as for most it is not, then perhaps Doyle has offered at least some line of defense.

Doyle's prescription to manage risk presupposes a broader account of the risk created by complicated online environments. Massimo Durante offers one such account in his "Trust and Security in the Digital Age: Algorithms, Standards, and Risks". Durante draws a crucial distinction between safety and security: Where safety is the immediate defense of life and well-being from threat, security protects ones life projects, including presumably their inception, fostering, and fruition. Frequently–and sometimes, perhaps, deliberately–confused, safety is a necessary condition on security, but not vice versa. The serf obtains safety, for example, but at the expense of security insofar as their life projects are their aristocrat's, not their own.

Security is a uniquely critical feature of well-constructed online environments, since such environments have themselves become decidedly necessary to the projects of today's lives. Yet this puts the individual, unavoidably, in the position of delegating security to corporate entities and government agencies. The complex, altogether distributed nature of the online ecosystem presupposes, for the sake of security, risk-management at the levels of design, implementation, and application, with a particular and altogether necessary emphasis on automated risk-management and the development of trusted systems. Feeble libertarian fantasies aside, no individual has the capacity to ensure their own security in such an environment. Security, in short, presupposes trust, yet trust itself engenders risk. The design decisions made for purposes of risk-management in the development of trusted systems effectively codify and automate social values which, whether by intention or not, may give the appearance of transparency while nevertheless incorporating subtle discriminations and manipulations. The problem is all the more acute because many of the design, development, and implementation decisions are in turn opaque to democratic evaluation.

Durante points to a yet darker possibility: The common confusion of security for safety is ripe opportunity for exploitation by governments. After all, claiming threats to safety as justification for massive surveillance, data-harvesting, and data-analytics as per Doyle's argument dramatically impinges on security as Durante construes the distinction between safety and security. Safety, as a ploy, threatens security and with it the promise of online environments to play an integral role in life's projects. This is an ancient tension, to be sure, yet it is one made all the more pressing by the technology involved.

Ugo Pagallo closes the volume, appropriately enough, examining the legal-philosophical implications of hard legal cases emerging from the use of information technologies. Just as hard cases in ethics are useful to study because of the rift they expose between, say, utilitarian and deontological moral normative analyses, hard legal cases expose the gulf between tolerance-based and justice-based approaches in legal normative analysis–or, as Pagallo dubs them, lockean and platonic approaches, respectively. Complicating matters is the fact that some of the hard legal cases at the leading edge of law and politics regarding information technologies are genuinely novel and surprising, while many others simply continue traditions of posing long-standing legal puzzles and conflicts. That is, some are indeed new wine, while many are old wine in new bottles.

Nevertheless, hard legal cases of information technologies drive dispute among scholars in the first instance on whether a solution exists and, in the second instance, on just how the unique solution, or resolution, if it exists, is to proceed in weighing justice considerations against tolerance, and vice versa. Resolution, if attainable, reveals a legal paradox of sorts, since the cases depend on both the tolerant application of justice and the just use of tolerance, each setting limits on the other. Yet bouncing from tolerance to justice and back from justice to tolerance, otherwise separately at odds with one another in approach and outcome to legal hard cases, leaves their resolution an open question. Focusing on numerous examples from information ethics and jurisprudence, Pagallo argues for a nuanced methodological analysis which shows one way by which a middle ground between justice and tolerance can be found, the one tempering the other in application to the hard cases.

7. Concluding Remarks

Harnessing computation from theory to engineering to application in its many permutations has clearly presented unique scholarly opportunities, all of them so interdisciplinary as to obliterate distinctions of discipline altogether. Are the philosophers roboticists, or are the roboticists philosophers? Are the mathematicians neuroscientists, or are the neuroscientists mathematicians? Are the legal theorists computer scientists, or are the computer scientists legal theorists?

In the end, and in light of the preceding discussions, the only reasonable response is a shrug: It just does not matter. As it has swept through every discipline, the computational turn has succeeded in wiping clean the deploringly artificial distinctions between those disciplines wrought by the balkanization of the resource-deprived modern university. Instead, threads of inquiry are woven throughout and link seemingly disparate research agendas, threads this introduction strives to highlight.

Their multi-disciplinary–better, *a-disciplinary*–investigations reveal the fruitfulness of erasing distinctions among and boundaries between formally established academic disciplines. This should come as no surprise: The computational turn itself is a-disciplinary, and no former discipline, whether scientific, artistic, or humanistic, has been left untouched. Rigorous reflection on the nature of these transformations, as we have seen, opens the door to inquiry into the nature of the world, what constitutes our knowledge of it, and our understanding of our place in it. That these investigations are only just beginning is signaled in part by the many contributions to this volume which close by describing open problems and inviting further research.

References

[Arrow, 1962] Arrow, K. (1962). Economic welfare and the allocation of resources for invention. In *The rate and direction of inventive activity: Economic and social factors*, pages 609–626. Princeton University Press, Princeton, NJ.

[Boolos and Jeffrey, 1989] Boolos, G. and Jeffrey, J. (1989). *Computability and Logic*. Cambridge University Press, Cambridge, UK, 3rd edition.

[Carnap, 1998] Carnap, R. (1998). The value of laws: Explanation and prediction. In Curd, M. and Cover, J., editors, *Philosophy of Science: The Central Issues*, pages 678–684. W.W. Norton, New York, 2nd edition.

[Church, 1932] Church, A. (1932). A set of postulates for the foundation of logic. *Annals of Mathematics (2nd Series)*, 33(2):346–366.

[D'Agostino and Luciano, 2009] D'Agostino, M. and Luciano, F. (2009). The enduring scandal of deduction. is propositional logic really uninformative? *Synthese*, 167:271–315.

[Dennett, 1987] Dennett, D. (1987). *The Intentional Stance*. The MIT Press, Cambridge, MA.

[Eliasmith and Trujillo, 2014] Eliasmith, C. and Trujillo, O. (2014). The use and abuse of large-scale brain models. *Current Opinion in Neurobiology*, 25:1–6.

[Fetzer, 2014] Fetzer, J. (2014). Information: Does it have to be true? *Minds and Machines*, 14(2):223–229.

[Floridi, 2007] Floridi, L. (2007). In defence of the veridical nature of semantic information. *European Journal of Analytic Philosophy*, 3(1):31–42.

[Frege, 1980] Frege, G. (1980). On sense and reference. In Geach, P. and Black, M., editors, *Translations from the Philosophical Writings of Gottlob Frege*, pages 36–56. Blackwell, Oxford, 3rd edition.

[Gilbert, 2014] Gilbert, M. (2014). *Joint Commitment: How We Make the Social World*. Oxford University Press, Oxford.

[Gödel, 1931] Gödel, K. (1931). Über formal unentscheidbare sätze der principia mathematica und verwandter systeme i. *Monatshefte für Mathematik und Physik*, 33:173–198.

[Hempel, 1998a] Hempel, C. (1998a). Two basic types of scientific explanation. In Curd, M. and Cover, J., editors, *Philosophy of Science: The Central Issues*, pages 685–694. W.W. Norton, New York, 2nd edition.

[Hempel, 1998b] Hempel, C. (1998b). Two basic types of scientific explanation. In Curd, M. and Cover, J., editors, *Philosophy of Science: The Central Issues*, pages 808–825. W.W. Norton, New York, 2nd edition.

[Kramer, 2014] Kramer, A.I., e. a. (2014). Experimental evidence of massive-scale emotional contagion through social networks. *Proceedings of the National Academy of Sciences*, 111(24):8788–8790.

[Piccinini, 2017] Piccinini, G. (2017). Computation in physical systems. In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, summer 2017 edition.

[Putnam, 1975] Putnam, H. (1975). *Minds and Machines*, pages 362–386. Cambridge University Press, Cambridge, UK.

[Scarantino and Piccinini, 2010] Scarantino, A. and Piccinini, G. (2010). Information without truth. *Metaphilosophy*, 41(3):314–330.

[Shannon, 1949] Shannon, C. E. (1949). Communication theory of secrecy systems. *Bell system technical journal*, 28(4):656–715.

[Turing, 1936] Turing, A. (1936). On computable numbers, with an application to the entscheidungsproblem. *Proceedings of the London Mathematical Society*, 42:230–265.