The International Association for Computing and Philosophy

2017 Annual Meeting

June 26-28

Courtesy the McCoy Family Center for Ethics in Society and Stanford University

Program (Draft)

	Monday, June 26th	
8:00-10:00	Registration	
	Session A	Session B
8:30-11:30	**Harry Halpin & Alexandre Monnin,** *Philoweb: Symposium on the Web and Philosophy 2017*	TBA
11:30-12:30	**The Simon Award Keynote Address: Andrea Scarantino**	
12:30-2:00	Lunch	
2:00-2:30	**Shannon Vallor,** "Algorithmic Opacity and the Shrinking Space of Moral Reasons in Computing Practices"	**Derek Leben,** "A Rawlsian Algorithm for Autonomous Vehicles"
2:30-3:00	**Anne Gerdes,** "The (Impossible) Art of Balancing National Security and Privacy in a Global Context"	**Arthur Schwaninger,** "Training the moral behaviour of self-driving cars"
3:00-3:30	**Sandra Wachter, Brent Mittelstadt, and Luciano Floridi,** "Explaining Algorithms: On Meaningful Explanations of Automated Decision-making"	**Mikkel Willum Johansen and Henrik Kragh Sørensen,** "Posing and Attacking Mathematical Problems: Human Mathematical Practice, Experimental Mathematics, and Proof Assistants"
3:30-4:00	**Tony Doyle,** "Big Data, Personal Information, and Intellectual Property"	**Robin Hill,** "Deep Dictionary: Analogy, Definition, and Word2Vec"
4:00-4:30	**Karen Frost-Arnold,** "The Epistemic Virtues and Vices of Online Lurking"	**Javier Blanco and Pio Garcia,** "Effectiveness and Programmability"
4:30-5:00	**Jessica Heesen,** "Questions of Technological Determinism in Information Ethics"	**Mate Szabo and Patrick Walsh,** "Gödel's and Post's Proofs of the Incompleteness Theorem"
5:00-7:00	Reception	

	Tuesday, June 27th	
8:00-10:00	Registration	
	Session A	Session B
8:30-11:30	**Gualtiero Piccinini,** *Symposium: Computation, Mind, and Brain* [Neal Anderson (UMass-Amherst) Rosa Cao (Stanford) Corey Maley (KU) Gualtiero Piccinini (UMSL) Michael Rescorla (UCLA) Oron Shagrir (Hebrew U, Jerusalem)]	**Ugo Pagallo, Massimo Durante, and Jacopo Ciani,** *Workshop on the Normative Challenges of Converging Technologies: AI, Big Data, and the Internet of Everything*
11:30-12:30	**The Covey Award Keynote Address: Ray Turner**	
12:30-2:00	Lunch	
2:00-2:30	**John Licato,** "Two Paradoxes and Their Implications for AI-Assisted Analysis"	**Marcello Guarini,** "Carebots and the Ties that Bind"
2:30-3:00	**Ioan Muntean,** "The Small from the Big: Discovering Models and Mechanisms with Machine Learning"	**Steve Mckinlay,** "Swarm Technology and Emergence in Lethal Autonomous Weapon Systems"
3:00-3:30	**Orlin Vakarelov,** "Theory and the Science of the Mind-boggling"	**Zac Cogley,** "Future Autonomous Weapons Will Make Moral Judgments"
3:30-4:00	**Ron Cottam, Willy Ranson, and Roger Vounckx,** "Application of “Logic in Reality” to the Computation of Levels of Consciousness in the Multi-Scaled Brain"	**Erica Neely,** "Augmented Reality, Augmented Ethics: Who Has the Right to Augment a Particular Physical Space?"
4:00-4:30	**Naveen Sundar Govindarajulu, Selmer Bringsjord, Rikhiya Ghosh, Atriya Sen, Kevin O’neill, and James James,** "A Study in Suicide via Defeasible Cognitive Calculi: With Provision for Self-Reasoning"	**Fiona McEvoy,** "Decisions, Decisions: Big Data and the Future of Autonomy"
4:30-5:00	**Paul Schweizer,** "Types, Tokens and Turing Tests"	**Andreas Wolkenstein,** "From roboethics to robopolitics. Why and how knowledge about the Trolley Dilemma should influence the development of self-driving cars"
5:00-7:00	Reception Dinner	

	Wednesday, June 28th	
8:00-10:00	Registration	
	Session A (IT & Democracy)	Session B
8:30-9:00	**John Sullins,** "Politically Motivated Ransomware and the Future of Democracy"	**Björn Lundgren,** "Conceptualising the Values of Anonymity in the 21st Century and Beyond"
9:00-9:30	**Harry Halpin,** "Beyond the State of Exception: Halting Fascism by Inscribing Rights into the Internet"	**Don Fallis,** "Shedding Light on Keeping People in the Dark"
9:30-10:00	**Robin Hill,** "Virtue in the Valley"	**Gary Smith,** "Are Second Order Resemblance Relations that Underpin Representation in Artificial Neural Networks Sufficient for Syntactic Systematicity?"
10:00-10:30	**Otto Kakhidze and Adam Ramey,** "The Obligations of Social Media Companies"	
10:30-11:00	**Leandro De Brasi,** "Personalisation Filters and Deliberative Democracy"	
11:30-12:00	**Claudio Agosti,** "Web Tracking Ecosystem: Mapping the Structure of Corporate Religious Profiling"	
12:00-12:30	**Presidential Address: Don Berkich**	
12:30-1:00	*Open Business and Planning Meeting*	

Keynote Addresses (in order of presentation)

Andrea Scarantino

The 2017 Simon Award
Monday the 26th, 11:30 – 12:30

HOW TO DO THINGS WITH EMOTIONAL EXPRESSIONS

How did the information words carry emerge from natural varieties of information? Charles Darwin was among the first to suggest that emotional expressions may have laid the foundations for the emergence of language. The question is: How? The aim of this talk is to introduce a new framework for the study of emotional expressions I call the Theory of Affective Pragmatics (TAP). As linguistic pragmatics focuses on what utterances mean in a context, affective pragmatics focuses on what emotional expressions mean in a context. TAP develops and connects two principal insights. The first is the insight that emotional expressions do much more than simply expressing emotions. As proponents of the Behavioral Ecology View of facial movements have long emphasized, bodily displays carry natural information about the signaler's intentions and requests. The second insight TAP aims to articulate and apply to emotional expressions is that it is possible to engage in analogs of speech acts without using language at all. I will argue that there are important and so far largely unexplored similarities between what we can "do" with words sensu Austin-Searle and what we can "do" with emotional expressions. In particular, the core tenet of TAP is that emotional expressions, by virtue of the natural information they carry, are a means of not only expressing what's inside, but also of directing other people's behavior, of representing what the world is like and of committing to future courses of action. Since these are some of the main things we can "do" with words, the take home message of my talk is that, from a communicative point of view, most of what we can do with language we can also do with non-verbal emotional expressions. I conclude by exploring some reasons why, despite the analogies I have highlighted, emotional expressions are much less powerful communicative tools than speech acts.

Ray Turner

The 2017 Covey Award
Tuesday the 27th, 11:30 – 12:30

THE ONTOLOGY OF PROGRAMS

The ontological standing of programs has been the subject of some debate in the relatively small amount of philosophical literature that concerns mainstream computer science. In one guise they are symbolic entities with an abstract content imposed by the semantic account of their containing programming language. In another they are concrete devices that causally determine mechanical computations. How do we conceptually package these two facets of programs? In this talk we shall investigate matters from the perspective of technical artefacts, one of the core notions from the philosophy of technology. In doing so we shall encounter some of the central questions in the philosophy of computer science.

Symposia (in order of presentation)

Harry Halpin & Alexandre Monnin, *Philoweb: Symposium on the Web and Philosophy 2017*: 27th, 8:30 – 11:30

> The Web is without a doubt having a profound effect on who we are and how we think. The cognitive - the fundamental aspects of intelligence given by attention, learning, and representation - is being shaped by ubiquitous penetration of the Web into daily life. Likewise, these effects are now leaking into the world of large, provoking new approaches to questions from ethics and epistemology.
> The focus on the foundational seminar will be panel discussion over how the Web transforms concepts from philosophy and artificial intelligence, and how this transformation effects (if at all) problems in philosophy of the computation, information, language, and mind that will be of interest to IACAP members. Particular interest will be taken in the relationship between ethics and Web-scale artificial intelligence (including AI with “humans-in-the-loop”), the extended mind hypothesis, the differences between classical artificial intelligence and more social oriented collective intelligence as guiding frameworks, and how mutually beneficial relationships can be established between 'predictive coding' in neuroscience and the philosophy of the mind and machine-learning AI, and the relationship of the philosophy of the Web to larger philosopical projects such as Floridi's philosophy of information.
> We expect the results of this workshop to influence future generations of philosophers at IACAP to focus on questions around the intersection of cognition and ethics in the space of Internet-enabled technologies, including discussions of both general philosophical questions and low-level technical detail, including the debate between the two.

Gualtiero Piccinini, *Symposium: Computation, Mind, and Brain*: Neal Anderson (UMass-Amherst) Rosa Cao (Stanford) Corey Maley (KU) Gualtiero Piccinini (UMSL) Michael Rescorla (UCLA) Oron Shagrir (Hebrew U, Jerusalem)

> Ever since the cognitive revolution, computation has been a central notion in the study of the mind and brain. The brain was seen as a computer and the mind as its software. Recent developments in both the sciences of computation and the sciences of mind and brain‹such as the rise of cognitive neuroscience, computational neuroscience, Bayesian approaches, and deep learning methods‹are prompting a philosophical reassessment of computation and the role it should play in psychology and neuroscience. New accounts of physical computation as well as new ideas about which notion of computation applies to the brain--and how it applies--are being defended. This symposium presents current work at the forefront of these debates.

Ugo Pagallo, Massimo Durante, and Jacopo Ciani, *Workshop on the Normative Challenges of Converging Technologies: AI, Big Data, and the Internet of Everything*

IoT, Big Data and AI are intertwined, converging and will drastically influence business models in the digital economy. A collection of everyday physical “smart devices” equipped with microchips, sensors, and wireless communications capabilities and connected to the internet and to each other, shall receive, collect and send myriads of user data, track activities and interact with other devices, in order to provide more efficient services tailored to users’ needs and desires. More devices on the web will lead to orders of magnitude more data.

The very near future will bring us more complex and multi-task intelligent devices that will be using AI to make decisions while relying on external distributed data sources. As the scope of intelligent agents’ activities broadens, it is important to ensure that such socio-technical systems will not make irrelevant, counter-productive, harmful or even unlawful decisions. As the intensity and magnitude of this technological revolution still must be fully comprehended, the law may struggle to evolve quickly enough to address the challenges it poses.

To establish a legal framework which ensures an adequate level of protection of personal data and other individual rights involved, while at the same time providing an open and level playing field for businesses to develop innovative data-based services, is a challenging task.

Therefore, the research question arises as to how the needs for protection and business interests can best be accounted for by legal disciplines. At the outset, the particular features of personal data and possible justifications for legal intervention are to be explored from the perspective of ethics, economics, social sciences, engineering and data protection research.

Papers (in alphabetical order by first author)

Claudio Agosti, "Web Tracking Ecosystem: Mapping the Structure of Corporate Religious Profiling" (Brazil, Coding Rights): 28th, 11:30, Session A-IT & Democracy

> US Government resources cannot be used to make religious and ethnic databases however, private advertisers might have such data, US intelligence or other LEA can easily get these data considering their actual powers.
>
> If an user accesses frequently to pages speaking of islamic faith, indeed an algorithm can assume the user is an active Islamic practitioner. The frequency of access and the intensity of the interest can be measured by the website itself but, the many third party trackers doing that as core business are more dangerous.
>
> These trackers are present in web pages for advertising, analytics, social engagement and technical reasons.
>
> The website is the ultimate responsible of the third party trackers servers to their readers, and often they are not aware of who they are including nor of the consequences.
>
> This project goal is to assess third party trackers in the website and advocates to three targets:
>
> 1) The website management: they might be not aware of the third party trackers presence, because the ecosystem is mostly automatic and rarely raises concern. Yet, now that religious discrimination has been institutionalized, they need to be more responsible than ever before.
>
> 2) The citizen who practice Islam, making them know the meaning of living in this corporate surveillance system. Since the implications are clear only when a physical threat like the current one, exists.
>
> 3) The tracking companies, making them aware why their data can be used to harm people involved. Usually, who is doing analytics or advertising do not see (or don’t want to think about) these potential abuses.

Javier Blanco and Pio Garcia, "Effectiveness and Programmability" (Argentina, Universidad Nacional de Córdoba): 26th, 4:00, Session B

> We focus on the distinction between effectiveness and programmability. We propose that effectiveness can be understood in terms of a mechanistic account through an axiomatic approach. Programmability can be, in turn, analyzed as the characteristic feature of a relational account of computation. In this sense, computability is the intersection of effectiveness and programmability, and being “computational” will turn to be a gradual concept that depends on the degree of programmability of a system. We show how many fundamental questions in philosophy of computing can be better addressed from this perspective.

Leandro De Brasi, "Personalisation Filters and Deliberative Democracy" (Chile, Universidad Alberto Hurtado): 28th, 10:30, Session A-IT & Democracy

> Our lives are increasingly mediated by technologies and I argue in this talk that such technological mediation has changed our epistemic circumstances. In particular, assuming here that democracy is intended to be an excellent vehicle for making intelligent decisions and uncovering the truth, that change epistemically challenges deliberative democracy by perpetuating cognitive biases. I also consider if public engagement is required to redesign these technologies.

Zac Cogley, "Future Autonomous Weapons Will Make Moral Judgments" (Northern Michigan University): 27th, 3:00, Session B

> I have two aims in this paper. One direct, the other more oblique. The direct aim is to respond to two putative principled objections to AWS presented by Duncan Purves, Ryan Jenkins, and Bradley J. Strawser. I do this in part by referencing the abilities of AlphaGo, an artificial intelligence that expertly plays the game Go. Purves, Jenkins, and Strawser have recently argued that AWS will never be able to make moral judgments because such systems will lack intuition and make 'decisions' only by following preprogrammed rules. In response, I demonstrate that AlphaGo works via an informational process similar to human intuition and that it does not simply follow preprogrammed rules. Additionally, they claim that AWS will never be able to act for the right reasons because artificial intelligence cannot act for reasons at all. In reply, I argue that AlphaGo's abilities are due to intentional states that represent certain moves as reasonable to consider more fully and also states that represent the best moves to make. In the course of my argument, I also make progress on my indirect aim: to make it reasonable to think that future artificially intelligent systems will be able to instantiate human moral capacities. I do not defend the development and use of AWS. I simply argue that Purves, Jenkins, and Strawser's recent arguments against AWS aren't sound and, as a corollary, give some insight into the ethical capabilities of future AI systems.

Ron Cottam, Willy Ranson, and Roger Vounckx, "Application of "Logic in Reality" to the Computation of Levels of Consciousness in the Multi-Scaled Brain" (Belgium, Vrije Universiteit Brussel and IMEC vzw): 27th, 3:30, Session A

> We set up four basic criteria for the establishment of a hierarchical representation of the brain's information-processing structure. The first of these criteria is the content of Brenner's book "Logic in Reality". The four positions together lead to a description of consciousness which relies on the brain-wide hierarchical nature of information-processing as a quasi-physical grounding. We characterize this information-processing as a generalized form of computation. We address the relationship between different levels of organization in the brain through a non-nested model hierarchy, rather than a conventional nested one. This represents different scales of organization by recognizable levels which each constitute a model of the entire brain, rather than

models of individual elements at each scale. The initial hierarchy decomposes into a binary pair of partials, one modeling directly locally-scaled information itself, the other its locally-scaled internalized ecosystem. We relate this duality to a modification of Brenner's proposition of ontological duality. The quasi-impossibility of directly accessing specific scales of organization 'from outside' leads to the generation of a pair of simplified hyperscalar representations which support approximate access to all of the individual scales. Mutual observation between the two characters of information-processing at each scale generates locally-scaled internal awarenesses built on an assumed pan-psychic primeval awareness. Here again, a modification of Brenner's 'T-state emergence' comes into play. Mutual observation between the two hyperscalar representations generates the partially internal, partially external high-level consciousness which we experience. We conclude with presentation of possible evidence that this binary derivation of consciousness is a valid description.

Tony Doyle, "Big Data, Personal Information, and Intellectual Property" (Hunter College): 26th, 3:30, Session A

Big data has radically altered flows of personal information, with serious implications for privacy and autonomy. Its novel inferences about our preferences, susceptibilities, and vulnerabilities permit targeted and manipulative advertising and can unjustly limit the ability of many to get insurance, a mortgage, or a job. Nearly all of this happens without data subjects' awareness. Unfortunately, the chief responses to this challenge to privacy and autonomy—notice and consent, legislation, and technology that obfuscates our identity or digital activities—have failed to be panaceas. I examine the case for "propertizing" personal information, wherein data collectors compensate data subjects for use of the latter's data. I argue that propertization fails to provide relief from big data's assault on privacy, offering the following criticisms of the proposal: (1) it would lead to greater inequality; (2) the informed consent that propertization requires will generally be impossible to obtain; (3) unlike other types of property, personal information is not freely alienable; (4) propertization laws would be unenforceable; and (5) personal information would be radically unlike other types of property, intellectual or otherwise.

Don Fallis, "Shedding Light on Keeping People in the Dark" (University of Arizona): 28th, 9:00, Session B

We want to keep hackers in the dark about our passwords and our credit card numbers. We want to keep potential eavesdroppers in the dark about our private communications. In order to address the important ethical and epistemological issues raised by this need for secrecy in the digital world, it is helpful to have a good understanding of the concept of keeping someone in the dark. Philosophers (e.g., Bok 1983, Carson 2010) have analyzed this concept in terms of concealing and/or withholding information. However, their analyses incorrectly exclude clear instances of keeping someone in the dark. And more importantly, they incorrectly focus on possible means of keeping someone in the dark rather than on what it is to keep someone in the

dark. In this paper, I argue that you keep X in the dark about P if and only if you intentionally leave X without a true belief about P.

Karen Frost-Arnold, "The Epistemic Virtues and Vices of Online Lurking" (Hobart & William Smith Colleges): 26th, 4:00, Session A

An epistemically virtuous person attempts to unlearn their socially constructed ignorance of their own privileges and prejudices. The epistemologies-of-ignorance literature has revealed various ways of becoming privilege cognizant, including seeking beneficial epistemic friction (Medina 2013) and 'world'-traveling (Lugones 2003), among others. The internet provides opportunities for these practices. The privileged can read blogs, tweets, Facebook posts, and Reddit threads written by members of marginalized groups in order to learn more about the oppression others face. This is one way in which anonymous lurking (i.e., reading online communications without engaging in the conversation oneself) can be epistemically beneficial. But by merely lurking, an agent forgoes the sustained interactions and relationships that many philosophers argue are important to unlearning ignorance. Lurking also protects one from having one's own prejudices openly challenged. Thus, the virtues of lurking need to be carefully balanced with the virtues of online critical engagement across differences. But there are also dangers in these critical dialogues. Privileged speakers can disrupt, derail, and hijack online spaces for marginalized communities. Even well-intentioned privileged people can damage the climate of trust in online spaces for marginalized communities. Therefore, epistemic practices aimed at remedying ignorance need to be developed in the context of an epistemically virtuous character that can discern when to engage, when to lurk, and how to avoid damaging online epistemic communities. This paper introduces a virtue epistemology for lurking to address these challenges.

Anne Gerdes, "The (Impossible) Art of Balancing National Security and Privacy in a Global Context" (Department of Design and Communication, University of Southern Denmark): 26th, 2:30, Session A

This paper highlights the work of collaborating European journalists, who in a series of articles, under the heading "Security for Sale – the Price we pay to protect Europeans," problematize the European Union Funding framework for security technology research, which unfortunately may enhance business opportunities for mass surveillance systems in non-democratic states. Based on a case, involving a research project in which I participated as an ethical adviser, the paper illustrates how a lack of global perspectives constitutes a weakness inherent in methodologies within design ethics, such as Privacy by Design and value sensitive design. Finally, drawing on the notion of professional idealism (Mitcham, 2003), the paper concludes by arguing in favour of moral activism from a global outlook, which goes beyond the walled gardens of the European Union.

Naveen Sundar Govindarajulu, Selmer Bringsjord, Rikhiya Ghosh, Atriya Sen, Kevin O'neill, and James James, "A Study in Suicide via Defeasible Cognitive Calculi: With Provision for Self-Reasoning" (Rensselaer Polytechnic Institute): 27th, 4:00, Session A

> We look at Schopenhauer's arguments for suicide and formalize them in the deontic cognitive event calculus DeCEC∗. This happens in the context of a robot arguing with its friends (other robots) whether it should or should not commit suicide. We show that one of the argument can be easily overcome, but an egoism based argument is quite hard to overcome, formally at least. We use a self-representation scheme that follows Castañeda's analyses [1999] to formalize the egoism argument.

Marcello Guarini, "Carebots and the Ties that Bind" (University of Windsor): 27th, 2:00, Session B

> Robots are already assisting with eldercare, and nannybots are on the way. Caring for the elderly and for children used to be thought of as something, more-or-less, distinctly human. Other species look after their young, but not as long as we do. And many years of caring for those who have aged and are no longer in a position to look after themselves – we've yet to see that in other species. Caring is an important part of who we are. Robotics research is now being done that could have an impact on the two forms of caring just mentioned: that of the elderly and infirmed, and that of the very young. The point of the paper is to raise some concerns about the potential role robots may end up playing in the care of human beings. The concerns are not motivated by a distrust of technology in general or robots in particular. Rather, the concerns are motivated by (a) the importance of other-regarding (caring) behaviour in our species, and (b) the rather limited discussion around the issue of robots becoming involved in that behaviour.

Harry Halpin, "Beyond the State of Exception: Halting Fascism by Inscribing Rights into the Internet" (MIT): 28th, 9:00, Session A-IT & Democracy

> Today, it appears that the once hopeful and millennial promise of the Internet has become a dark nightmare of surveillance and "fake news." However, this is not necessarily the only possible future of the Internet, as we will argue that a new vision of the future of the Internet is possible where fundamental rights are protected not by laws or coercive force, but by inscribing those rights into the technical protocols of the Internet itself.

Jessica Heesen, "Questions of Technological Determinism in Information Ethics" (Internationales Zentrum für Ethik in den Wissenschaften): 26th, 4:30, Session A

> This paper is concerned with attempts to develop an ethically justifiable approach to Information and Communication Technologies (ICT) as forming a comprehensive or ubiquitous system. It discusses positions concerning the problem of inherent necessity or technological determinism produced by technological systems (Jacques Ellul, Helmut Schelsky) and positions related to the totality of media (Jean Baudrillard, Paul Virilio). It argues that it is crucial to keep open a plurality of manners of living and dealing with ICT. This, in turn, is a precondition for being

aware of the normative structures that are produced by comprehensive systems. Against this background, the paper will outline strategies for dealing with media totality for users and developers, and on a societal level. These strategies involve methods of distancing oneself from ICT, user-friendly design and the particular experience of an ICT-free environment.

Robin Hill, "Deep Dictionary: Analogy, Definition, and Word2Vec" (University of Wyoming): 26th, 3:30, Session B

A recent result that abstracts English words into distributed vector representations that yield simple semantic aspects under vector arithmetic has drawn interest from the natural language processing community and philosophers who follow such developments. Yet what is the nature of that breakthrough? A dictionary both prescribes and describes the use of words in public discourse, and a dictionary is built on a structure of associations among words, their definitions. The type of analogy expressed in the vector calculations looks like, could be, and indeed must be, relective of the cross-references made by those definitions.

Robin Hill, "Virtue in the Valley" (University of Wyoming): 28th, 9:30, Session A-IT & Democracy

After centuries of ethics based on duties--deontological--and ethics based on effects--various forms of consequentialism--the mid-20th century saw the return of virtue ethics. Now we are witness to a de facto attempt to institutionalize virtue ethics at a large scale, in high tech, an attempt that seems to have failed. Virtue ethics holds that each of us should strive to be a good person, according to some ideal; doing right will follow when we reach that standard. While deontology and consequentialism can feel oppressive, the virtues are more positive, more attractive. It's easier for us to admire a person than to admire a theory. The trouble is that the identification and acquisition of virtues leads only tenuously to the Right Thing; guidance on particular action is not offered, and outcomes do not garner the scrutiny that we expect. The selection of virtues is controversial, though a given society tends to agree to a comfortable degree. Yet the virtues carry the appeal of vitality rather than the taint of old-fashioned burdens, the drudgery of the construction work of deontology or the calculated approach of consequentialism.

In high tech, we can trace virtue ethics through the development of social networking, as well as other high-tech enterprises. At prominent start-ups, entrepreneurs assumed that the virtues of sharing, open communication, assistance to humans, and connection would ensure the triumph of the Good; that those values, embedded in people and projects, would bring about the Right Thing. This can be seen in social media such as Facebook, in initiatives such as laptops for children in poor countries, in the push for MOOCs in education, and now in Google's self-driving cars. Proponents of all have assured us that their projects will bring a better world. But the results have been weak or counterproductive, the bad aspects manifest in shaming and bullying, and vicious word and deed, even to the extent of violence, brutality, inhumanity. Facebook and its ilk were mistaken in the expectation that solutions to the world's problems

would follow naturally from the global dissemination of social media. Communication has brought a stream of distortion, and connection has fostered groups bent on harm.

Mikkel Willum Johansen and Henrik Kragh Sørensen, "Posing and Attacking Mathematical Problems: Human Mathematical Practice, Experimental Mathematics, and Proof Assistants" (Department of Science Education, University of Copenhagen): 26th, 3:00, Session B

Computers are impacting mathematical research practice in profound ways. Their use for communication and typesetting is long established, and since the 1970s, computers have been important in running extensive searches as parts of mathematical theorems. In the past 20 years, other forms of exploratory uses of computers have emerged in which computers are interactively used to develop hypotheses or aid in the formal proof process.

Yet, mathematical creativity has (certainly not yet) been mechanized, and this raises the philosophical question of what separates human mathematical practice from that which can be (easily) automated and, correspondingly, how human and machine practice can be made to integrate in the field of mathematics.

In this paper we address these two questions by comparing automated and interactive theorem-provers with the practice of (human) research mathematicians. We draw on a qualitative interview study on how research mathematicians choose mathematical problems and how they attack and work with the problems they have chosen. We then discuss how and to what extend the construction of automated and interactive theorem provers can be informed by such insights into the practice of research mathematicians. In particular, we point to important aspects of how the collaboration between human and computer mathematics could be organized such that the strengths and weaknesses of the mechanized system is best matched by the strengths and weaknesses of the human mathematician.

Otto Kakhidze and Adam Ramey, "The Obligations of Social Media Companies" (Centre for Cyber Security, New York University in Abu Dhabi): 28th, 10:00, Session A-IT & Democracy

This paper provides conditions for determining the ethical obligations for social media companies regarding misinformation considering (i) uncertainty of the truth values of news content in the post-truth world, (ii) understanding the function of social media companies as providers of “digital public sphere” and (iii) the right of public to be informed by multiple sources of information. Their responsibility and obligations should be evaluated from a larger perspective as being potential providers of a healthy public discourse instead of being ordinary news media organizations.

Derek Leben, "A Rawlsian Algorithm for Autonomous Vehicles" (University of Pittsburgh, Johnstown): 26th, 2:00, Session B

Autonomous vehicles must be programmed with procedures for dealing with trolley-style dilemmas where actions result in harm to either pedestrians or passengers. This paper outlines a

Rawlsian algorithm as an alternative to the Utilitarian solution. The algorithm will gather the vehicle's estimation of probability of survival for each person in each action, then calculate which action a self-interested person would agree to if he or she were in an original bargaining position of fairness. I will employ Rawls' assumption that the Maximin procedure is what self-interested agents would use from an original position, and then show how the Maximin procedure can be operationalized to produce unique outputs over probabilities of survival.

John Licato, "Two Paradoxes and Their Implications for AI-Assisted Analysis" (Purdue University - Fort Wayne): 27th, 2:00, Session A

Given the centrality of conceptual analysis to modern philosophy, I suggest the possibility of AI-Assisted Analysis (AAA), and argue that it would be extremely helpful in many areas: legal reasoning, ethical reasoning, cognitive science, the social and political sciences, and so on. What would it take to engineer such a system? I argue that AAA should be based on a hybrid approach, which draws from at least two types of analysis: Carnap's explication, and the formal notion of analysis described by Chisholm. Although explication describes a kind of reasoning that can turn informal, vague concepts into more sharply-defined, formal ones, it falls victim to what Dutilh-Novaes calls the 'paradox of adequate formalization', which is closely related to the well-known paradox of analysis. Chisholm's formalism bypasses the second paradox, and suggests a strategy for avoiding the first. However, if a hybrid approach drawing from both explication and Chisholm's formalism is indeed possible, there are some implications, which I discuss.

Björn Lundgren, "Conceptualising the Values of Anonymity in the 21st Century and Beyond" (KTH Royal Institute of Technology): 28th, 8:30, Session B

In this article I argue the in order to analyse the values of anonymity (i.e. the values that anonymity protects) what we need is not, unlike what the current philosophical literature has focused on, a conception of a normative and/or descriptive account of anonymity, but a normative and descriptive account of the ability to be anonymous. I present definitions and a measurement, which, unlike previous conceptual analysis of anonymity, covers all relevant senses of anonymity and is useful in the analysis of (the values of) anonymity given the current threats of deanonymization technology that we are facing. As such, my proposal is conceptually beneficial contrary to previous accounts and also provides an action-guiding account for the protection of the values of anonymity.

Fiona McEvoy, "Decisions, Decisions: Big Data and the Future of Autonomy" (San Francisco State University): 27th, 4:00, Session B

Many of us think that terms like "Big Data" are only of relevance to technology geeks and Silicon Valley executives. The reality is that so-called "datafication" marks the beginning of a new human epoch that will have huge implications for all of us – especially generations being born right now. Understanding the ethics of tech has never been more critical than it is today, and any comprehensive analysis should have one of the most apparent challenges right at its core:

what Big Data means for our autonomy and notions of free will. Some commentators have already expressed nervousness about data-driven technology leading to the erosion of these central human capacities as we relinquish more and more decision-making to computers.

This paper tries to frame this emerging concern, before articulating three ways in which an increasing emphasis on Big Data might threaten our basic liberty. I will conclude that there is now very that little that societies and individuals can do to sidestep the rapid shift in perspective that characterizes this new era of data, intelligence and mass connectivity. Consequently, more work needs to be done to ensure that some of the more damaging effects of Big Data are mitigated.

Steve Mckinlay, "Swarm Technology and Emergence in Lethal Autonomous Weapon Systems" (Wellington Institute of Technology): 27th, 2:30, Session B

Much has already been written regarding the development and possible deployment of lethal autonomous weapon systems (LAWS). While there are no nations that currently use these weapons, they are under development and have been the subject of debate both academically as well as by various NGOs. Fully autonomous weapon systems raise considerable ethical concerns however; the combination of such weaponry with the development of micro-drone and swarm technologies significantly raises moral concern regarding the permissibility of deploying such systems. This article examines the nature of emergence in swarm based LAWS and compares this to other definitions of autonomous weapons systems. I conclude that while there may be a limited argument in support of weapon systems with very low levels of autonomy, systems that have the potential to exhibit emergent behaviour are by their very nature unpredictable and therefore should not be developed or deployed.

Ioan Muntean, "The Small from the Big: Discovering Models and Mechanisms with Machine Learning" (University of North Carolina, Asheville): 27th, 2:30, Session A

This paper proposes a new discussion on “Big Data” as the primary source of model-building in science, when the computational architecture used is machine learning or evolutionary algorithms (and possibly a combination of them). Is Big Data mining a proper method of discovering and building models? The focus is on discovery and building (rather than justification or confirmation) of existing theories or models; and on relevant epistemic and pragmatic aspects of these computational architectures (rather than on the quantitative features of Big Data). Two competing (or, more mildly put, complementing) types of models used in biology and cognitive science are scrutinized: mechanism modeling and computational modeling (mostly network modeling and dynamical modeling).

This paper aims to show that far from becoming irrelevant in the Big Data era, network models and mechanisms can be discovered and built from Big Data.

The argument is based on the concept of patterns in data, discussed in the context of the relation between data and models (Bogen, Woodward, McAllister). It relates it then to the concept of “small patterns” in Big Data (Floridi) as hidden aspects of mechanistic models, hard to fathom by scientists. Machine learning, as a tool to categorize and characterize real patterns is assessed in the context of mechanistic and network models. Evolutionary computation is assessed as a method to optimize the search for mechanisms and complex networks.

To compare and contrast the mechanistic account and its alternatives, this argument builds on two concepts central to all approaches: modularity (as related to decomposability), and organization (Bechtel, Darden, Craver), which both come in degrees and can be discovered through machine learning or evolutionary computation in Big Data (cf. E. Ratti and W. Pietsch). The paper concludes with the claim that Big Data, when it qualifies as scientific evidence, most likely has and will have a fundamental impact on the way we discover and build computational models in science.

Erica Neely, "Augmented Reality, Augmented Ethics: Who Has the Right to Augment a Particular Physical Space?" (Ohio Northern University): 27th, 3:30, Session B

Augmented reality (AR) blends the virtual and physical worlds such that the virtual content experienced by a user of AR technology depends on the user’s geographical location. Games such as Pokémon GO and technologies such as HoloLens are introducing an increasing number of people to augmented reality. AR technologies raise a number of ethical concerns; I focus on ethical rights surrounding the augmentation of a particular physical space. To address this I distinguish public and private spaces; I also separate the case where we access augmentations via many different applications from the case where there is a more unified sphere of augmentation.

Private property under a unified sphere of augmentation is akin to physical property; owners retain the right to augment their property and prevent others from augmenting it. Private property with competing apps is more complex; it is not clear that owners have a general right to prevent augmentations in this case, assuming those augmentations do not interfere with the owner’s use of the property. I raise several difficult cases, such as augmenting a daycare with explicit sexual or violent images.

Public property with competing apps is relatively straightforward. Most augmentation is ethical; those apps simply function like different guidebooks. Under a unified sphere of augmentation it is unclear whether augmentations should be treated more like public speech (which we value) or graffiti (which we do not) or (most likely) some of each. Further consideration is needed to determine which of these augmentations we view as ethical.

Arthur Schwaninger, "Training the moral behaviour of self-driving cars" (Ludwig Maximilian University of Munich): 26th, 2:30, Session B

> When a self-driving car is about to get involved in an accident, it might be confronted with a moral dilemma such as the "trolley problem" and car developers are required to determine the moral basis on which the car is ought to behave. By incorporating machine learning algorithms, it is suggested that a descriptive ethical system is the choice of preference. This paper opens the discussion about the kind of training data one should apply in the process of developing such an ethical system. Contrary to existing software packages, it is argued that the training data should not be based on the evaluation of questionnaires but rather on the measurements of people's emotional states.

Paul Schweizer, "Types, Tokens and Turing Tests" (University of Edinburgh): 27th, 4:30, Session A

> The paper examines the issue of conceptualizing a (hypothetical) test of the capacities of computational artifacts that would achieve true parity with the evidence available regarding the capabilities of the human mind. I argue that the original Turing Test (2T) is fundamentally inadequate, and go on to investigate Harnad's Total Turing Test (3T), which involves successful performance of both linguistic and robotic behaviour, and which is often thought to incorporate the very same range of empirical data that is available in the human case. However, I argue that the 3T, like the 2T, is conceptually flawed because it only considers isolated tokens of the artificial cognitive structure, and it begins by simply presupposing human language and culture as the test platform. This methodology ignores the fact that in critical respects, mentality is not 'individualistic'. Tokens of the human cognitive variety have the mental states and contents they do in virtue of their dependency upon the capabilities of the human cognitive type, where the type is responsible for the sociolinguistic medium in which the tokens are embedded. Hence for true parity, the artificial type must itself have the ability to develop a comparable sociolinguistic medium.

Gary Smith, "Are Second Order Resemblance Relations that Underpin Representation in Artificial Neural Networks Sufficient for Syntactic Systematicity?" (University of Edinburgh): 27th, 9:30, Session B

> Systematicity is a classic problem for artificial neural networks. How can artificial neural networks, with their distributed representations, exhibit the kind of systematic behaviour characteristic of human language and cognition? If they can not, then they are an insufficient model of cognition. O'Brien and Opie (2002) have shown how second order resemblance relations, in which the representing system and the target system need not share physical properties, but instead share some kind of structural organization, can underpin abstract representation. In this paper, I attempt to push this further and show that second order resemblance can potentially allow artificial neural networks to exhibit systematicity. I give a worked example to show that if the activation space of a network can be divided in such a way as

to stand in a second order resemblance relation with a systematic system, such as language, then that network will be able to behave systematically with respect to that system.

John Sullins, "Politically Motivated Ransomware and the Future of Democracy" (Sonoma State University): 28th, 8:30, Session A-IT & Democracy

Ransomware will be used as a tool to suppress political activity. It can be used to extort political dissidents into informing on their compatriots or to cease their political activities. In this way, it can have a chilling effect on free speech, which is essential for democratic governance. At a higher level of operation it can be used to harass or stop large portions of a government or political party.

We will explore this issue form the standpoint of information ethics and see what this philosophical standpoint can add to our understanding of a future where the political processes we have come to rely on can be easily hacked.

Mate Szabo and Patrick Walsh, "Gödel's and Post's Proofs of the Incompleteness Theorem" (Carnegie Mellon University): 26th, 4:30, Session B

Emil Post worked on questions of incompleteness and undecidability already in the 1920s. To some extent he anticipated Gödel's results, but his work only saw publication much later, in 1965. Instead of trying to claim priority to Gödel, Post emphasized that:

> "with the Principia Mathematica as a common starting point, the roads followed towards our common conclusions are so different that much may be gained from a comparison of these parallel evolutions."

We take up this comparison. After we survey Post's approach and Gödel's proof, we distill and emphasize two key dissimilarities based on their different methodologies.

Orlin Vakarelov, "Theory and the Science of the Mind-boggling" (Duke University): 27th, 3:00, Session A

Some have argued that new Big-Data methods in science lead to the end of theory. This is wrong, but there is genuine tension between some conceptions of theory and scientific domains that are too complex for human comprehension – the science of the mind-boggling. The problematic assumption is that scientific knowledge must increase human understanding, and that theories must be comprehensible by humans. We call this the central dogma of scientific knowledge. The dogma is assumed in traditional meta-theory of science. We argue that if theories are to be saved in mind-boggling domains, the dogma must be relaxed, i.e., there may be useful theories that are not fully comprehensible. We suggest that some databases, extended by additional capacities, may be regarded as theories. Such database+ do not exist yet, but their conception helps us understand how to develop theories of complex domains, e.g., a theory of the whole biosphere.

Shannon Vallor, "Algorithmic Opacity and the Shrinking Space of Moral Reasons in Computing Practices" (Santa Clara University): 26th, 2:00, Session A

> Recent advances in machine learning have generated a new set of ethical problems regarding algorithmic opacity. Increasingly sophisticated yet opaque algorithms constrain and shape what we read, watch and hear online, who we are invited to meet or date, what medical treatments we are advised to undergo, who will hire us, how the justice system will treat us, and where we will be allowed to live. The lack of transparency in such processes raises profound ethical questions about justice, power, inequality, bias, freedom and democratic values in modern computing. Here I focus on a less commonly discussed concern: the potential for opaque algorithmic decision systems to lead to a contraction of what moral philosophers have called 'the space of moral reasons,' a concept that underpins personal and public practices of moral reflection, moral responsibility, moral imagination, moral justification and moral appeal. Using examples from algorithmic decision systems used in jurisprudence, human resources, and law enforcement, I show how contractions of the space of moral reasons can result from decision practices mediated by such systems. I conclude with some reflections on how more ethically-informed design and use of algorithmic decision systems might help us to hold open, or even enlarge, the space of moral reasons in personal and public life.

Sandra Wachter, Brent Mittelstadt, and Luciano Floridi, "Explaining Algorithms: On Meaningful Explanations of Automated Decision-making" (University of Oxford): 26th, 3:00, Session A

> Since approval of the EU General Data Protection Regulation (GDPR) in 2016, it has been widely and repeatedly claimed that a ‘right to explanation’ of decisions made by automated or artificially intelligent algorithmic systems will be legally mandated by the GDPR. This right to explanation is viewed as an ideal mechanism to enhance the accountability and transparency of automated decision-making. However, there are several reasons to doubt both the legal existence and the feasibility of such a right. In contrast to the right to explanation of specific automated decisions claimed elsewhere, the GDPR only mandates that data subjects receive limited information (Articles 13-15) about the logic involved, as well as the significance and the envisaged consequences of automated decision-making systems, what we term a ‘right to be informed’. This gap shows that the GDPR lacks explicit and well-defined rights and safeguards against automated decision-making, and therefore runs the risk of being toothless. As the right to explanation is not legally guaranteed, and before meaningful safeguards are possible, the technical constraints and normative requirements for meaningful explanations of automated decision-making must be clarified.

Andreas Wolkenstein, "From roboethics to robopolitics. Why and how knowledge about the Trolley Dilemma should influence the development of self-driving cars" (University of Tübingen): 27th, 4:30, Session B

> In this paper will argue that, contrary to some recent claims, people's reactions to the Trolley Dilemma do play a role for the ethics of self-driving cars, and more generally, that people's reactions - actual and behavioral reactions, not only thought-experiment reactions - to ethical dilemmas are an important input to the applied ethics of technology. However, the role is not (only or primarily) to be understood as a direct input into political decision-making about SDC or automated technology, but as input into a network of small experimental steps that aim at a bottom-up approach to inventing automated technology. The thesis here is that ethics is important, but not as a means to shape public policy, but as a means to shape private endeavors to produce technology for the sake of humanity. In this sense, then, there is a crucial step from roboethics, studying people's views about ethical questions in robotics, to robopolitics, dealing with the question of how to govern robot development. I will thus argue for the claim that robopolitical questions are not answered by roboethical insights, although the latter can inform the activities that the former should address (though in a different way than traditionally held). Along our way to this conclusion we will discuss recent works in roboethics and the question of TD in assessing SDCs, see how they misrepresent the role ethical reactions play in roboethics, and provide arguments for the idea that (1) experimental roboethics is important, and (2) robopolitics needs only indirectly take issue with roboethical insights.