

**CEPE-IACAP 2015
22-25 June 2015
University of Delaware**

PROGRAM

22 June 2015			
8:00 - 8:30	Registration (Coffee/Breakfast)		
8:30 - 9:00	Welcome: Thomas M. Powers & Mariarosaria Taddeo		
9:00 - 10:00	Keynote Address: William J. Rapaport <i>On the Relation of Computing to the World</i>		
10:00 - 10:30	Coffee Break		
10:30 - 11:00	Steve McKinlay, <i>Maker's Information: A Cognitive Approach</i>		
11:00 - 11:30	Paul Schweizer, <i>Cognitive Computation sans Representation</i>		
11:30 - 12:00	Alexander Ovsich, <i>Mathematical Models of Desire, Need and Attention</i>		
12:00 - 12:30	John Symons & Jack Horner, <i>How Does Software Limit Scientific Inquiry?</i>		
12:30 - 13:30	Lunch		
13:30 - 14:00	Selmer Bringsjord, <i>Well, Zombie Autonomy is Fearsome</i>		
14:00 - 14:30	Mario Verdicchio, <i>A Critique of Machine Ethics from the Perspective of Autonomy</i>		
14:30 - 15:00	Bendert Zevenbergen, <i>Philosophy Meets Internet Engineering: Examining and Aligning Ethics Reasoning</i>		
15:00 - 15:15	Coffee Break		
15:15 - 17:00	<table border="1" style="width: 100%; border-collapse: collapse;"> <tr> <td style="width: 50%; padding: 5px;"> Symposium: <i>Computational Ethics: the "case-based, act-centric" AMA, and Its Alternatives</i> Ioan Muntean, Chair </td> <td style="width: 50%; padding: 5px;"> Symposium: <i>Sex, Virtue, and Robots</i> John P. Sullins, Chair </td> </tr> </table>	Symposium: <i>Computational Ethics: the "case-based, act-centric" AMA, and Its Alternatives</i> Ioan Muntean, Chair	Symposium: <i>Sex, Virtue, and Robots</i> John P. Sullins, Chair
Symposium: <i>Computational Ethics: the "case-based, act-centric" AMA, and Its Alternatives</i> Ioan Muntean, Chair	Symposium: <i>Sex, Virtue, and Robots</i> John P. Sullins, Chair		
17:30 - 19:00	Reception: Deer Park Tavern, 108 W. Main St.		

23 June 2015			
8:00 - 8:30	Coffee/Breakfast		
8:30 - 9:00	Brent Mittelstadt, <i>Self-Care, Patient Empowerment and the Virtue of Personal Health Technologies</i>		
9:00 - 9:30	Luke Roelofs, <i>Rational Agency without Self-Awareness: Could 'We' replace 'I'?</i>		
9:30 - 10:00	Ingvar Tjostheim, <i>Telepresence, the Senses and Merleau-Ponty's Theory of Perception</i>		
10:00 - 10:30	Coffee break		
10:30 - 11:00	Soraj Hongladarom, <i>Big Data, Digital Traces and the Metaphysics of the Self</i>		
11:00 - 11:30	Brent Mittelstadt & Luciano Floridi, <i>Contextual Dissonance in Big Data</i>		
11:30 - 12:30	Keynote Address: Deborah Johnson <i>Getting a Handle on Big Data Ethics</i>		
12:30 - 13:30	Lunch		
13:30 - 14:00	Kirk A. Weigand & Ronald L. Hartung, <i>Below Global Coherence: Turtles, No; It's Situations All the Way Down</i>		
14:00 - 14:30	Richard Wilson, <i>Anticipatory Ethics and Brain Computer Interfaces (BCI's)</i>		
14:30 - 15:30	Keynote Address: Shannon Vallor <i>AI and the Automation of Wisdom</i>		
15:30 - 15:45	Coffee break		
15:45 - 17:30	<table border="0"> <tr> <td> Symposium: <i>Philosophy meets Internet Engineering: Ethics in Networked Systems Research</i> Bendert Zevenbergen, Chair </td> <td> Symposium: <i>Dynamics and Computation in Cognition—a Collaborative Approach</i> Orlin Vakarelov, Chair </td> </tr> </table>	Symposium: <i>Philosophy meets Internet Engineering: Ethics in Networked Systems Research</i> Bendert Zevenbergen, Chair	Symposium: <i>Dynamics and Computation in Cognition—a Collaborative Approach</i> Orlin Vakarelov, Chair
Symposium: <i>Philosophy meets Internet Engineering: Ethics in Networked Systems Research</i> Bendert Zevenbergen, Chair	Symposium: <i>Dynamics and Computation in Cognition—a Collaborative Approach</i> Orlin Vakarelov, Chair		

24 June 2015		
8:00 - 8:30	Coffee/Breakfast	
8:30 - 9:30	Keynote Address: Frances Grodzinsky <i>Big Data: The Patterns of Our Lives through the Lens of the Virtues</i>	
9:30 - 10:00	Aimee van Wynsberghe, <i>Ethical Data Provenance</i>	
10:30 - 11:00	Stephen Lilley & Frances Grodzinsky, <i>Should Social Physics Govern the Hyper-Connected, Big Data Age?</i>	
11:00 - 11:30	Coffee break	
11:30 - 12:00	John P. Sullins, <i>Finding the Path to Beneficial AI</i>	
12:00 - 12:30	Michael Devito, <i>The Sentimental Robot: Soldiers, Metal Psychopaths and Artificial Intelligence</i>	
12:30 - 13:00	Wendell Wallach, <i>Deep Learning, AI Safety, Machine Ethics and Superintelligence</i>	
13:00 - 14:00	Lunch	
14:00 - 14:30	Paul B. de Laat, <i>Profiling Contributors to Wikipedia in Order to Combat Vandalism: A Schauerian Approach to Justification</i>	
14:30 - 15:00	Johnny Hartz Søraker, <i>Beyond Right and Wrong - Applying Prudential-Empirical Ethics of Technology (PEET) to Autonomous Cars</i>	
15:00 - 15:30	Hamid Ekbia & Danial Qaurooni, <i>Problematics of Separation: The Case of Drones</i>	
15:30 - 16:00	Keith Miller, Marty J. Wolf, & Frances Grodzinsky, <i>This Ethical Trap is for Roboticians, Not Robots: Why use a robot when a human will do?</i>	
16:00 - 16:15	Coffee break	
16:15 - 17:15	Keynote Address: Michael Rescorla <i>Levels of Computational Explanation</i>	
16:45 - 18:30	Symposium: <i>{SuchThatCast Behind the Philosophy}</i> – Live @ IACAP'15 Johnny Hartz Søraker, Chair	Symposium: <i>Lightning Rounds</i> Don Berkich, Chair
19:30--	Conference Dinner: Caffe Gelato, 90 E. Main St.	

25 June 2015	
8:00 - 8:30	Coffee/Breakfast
8:30 - 9:00	Michele Rapoport, <i>Choosing and Acting Freely in Smart, Persuasive Environments</i>
9:00 - 9:30	James Williams, <i>What Is a 'Persuasive' Technology?</i>
9:30 - 10:00	Mariarosaria Taddeo & Luciano Floridi, <i>The Responsibilities of Online Service Providers in Information Societies</i>
10:00 - 10:30	Coffee break
10:30 - 11:00	Tobias Matzner & Regina Ammicht Quinn, <i>What Does Ethics do in Science and Engineering Research Projects? An Ethical Self-reflection in the Case of Security Research</i>
11:00 - 11:30	Markus Christen, et al., <i>Beyond Informed Consent – Investigating Ethical Justifications for Disclosing, Donating or Sharing Personal Data in Research</i>
11:30 - 12:00	Katleen Gabriels, <i>'I keep a close watch on this child of mine': A Moral Critique of Other-tracking Apps</i>
12:00 - 13:30	Lunch & Concluding Remarks
13:30 - 21:00	Optional trip to Philadelphia, return to the Marriott Hotel

ABSTRACTS

Selmer Bringsjord, *Well, Zombie Autonomy is Fearsome* (22nd 13:30-14:00)

Intuitively, it seems that the following "equation," where 'a' refers to an arbitrary agent, is worrisome:

$$(E) \text{ Autonomy}(a) + \text{Power}(a) = \text{Dangerous}(a)$$

It's of course said that power corrupts, and that absolute power corrupts absolutely. In the above equation is accordingly buried the idea that the more power an agent has, the more dangerous is that agent. In addition, if autonomy too comes in a continuum, the general idea underlying (E) would be that an agent is more and more dangerous as its power and autonomy both grow. Personally, I suspect that something like equation (E) is true, and I suspect that a number of those working in Defense suspect as well that something like it holds. Yet (E) is disturbingly imprecise; one reason is that autonomy is imprecise. The world now speaks routinely of "autonomous" machines. Daimler e.g. has a triumphant commercial in which it declares that one of its Mercedes cars can drive "autonomously." As a formalist, I believe that those publishing this commercial literally have no idea what they're talking about. Ditto for others who now speak this way. Fortunately, the situation can be repaired.

I analyze a prior Kolmogorov-Shannon proposal for formally defining autonomy (versus mere automaticity), and find it severely wanting. (There is *very* little other formal work.) I then offer the first draft of my own rigorous definition of autonomy. When this definition is applied to information-processing machines, the result is that no such machine can be autonomous, and hence by (E) at least, we have little to fear from machines at all, ever. But a variant of the first definition can be used to capture the concept of **zombie autonomy**. If we then interpret 'Autonomy' in (E) to refer to this kind of autonomy, the result is genuinely fearsome.

Markus Christen, et al., *Beyond Informed Consent – Investigating Ethical Justifications for Disclosing, Donating or Sharing Personal Data in Research* (25th 11:00-11:30)

The goal of this contribution is to briefly outline the possibilities and limitations of the classic idea of individual control and consent regarding the use of personal data and to investigate ethical justifications that may support disclosing, donating or sharing personal data, with a focus on using such data in research. This will be done in three steps: First, it is assumed – following several other scholars – that the practice of the ‘art of separation’ or the maintenance of ‘contextual integrity’ is a key moral issue that is at stake due to the recent developments in the field of Big Data. Second, it is argued that the core value of autonomy

(which provides the moral foundation of control and consent) cannot support the defense of privacy by itself, but must be complemented with two other core values – responsibility and fairness – in order to sufficiently describe the moral landscape of the problem under investigation. Third, it is drafted how research relying on (potential) personal data could proceed in order to comply with these values.

Paul B. de Laat, *Profiling Contributors to Wikipedia in Order to Combat Vandalism: A Schauerian Approach to Justification* (24th 14:00-14:30)

Profiling, wherever practiced, is a loaded subject that always brings up heated polemic, in particular as soon as sensitive features like race or religion are involved in the process. In this article, the pros and cons of Wikipedian profiling are explored. To what extent can it pass muster from a moral perspective? Do some elements need to be changed, or possibly the approach as a whole? The question seems especially acute since Wikipedia is to be considered a public undertaking that should be held to public standards.

Michael Devito, *The Sentimental Robot: Soldiers, Metal Psychopaths and Artificial Intelligence* (24th 12:00-12:30)

An exploration of two ways of constructing robotic morality: the procedural approach, based on Rawls/Chomsky's nativist moral grammar work, and the emotional approach, based on empirical Humeanism, in the context of the problem of morally-upstanding, extensible, autonomous robot soldiers. Drawing from moral philosophy, psychology and human neuroscience, the argument is made that autonomous robots must be manufactured with a full-fledged capacity for emotion to ensure moral behavior compatible with our human norms, especially in the case of military robots. Possible methods of constructing an "emotional bootstrap" to allow human-like moral development and the pitfalls of a non-emotional approach, including the possibility of creating dangerous robotic psychopaths, are explored.

Hamid Ekbia & Danial Qaurooni, *Problematics of Separation: The Case of Drones* (24th 15:00-15:30)

Drones, or Unmanned Aerial Vehicles (UAV), are aircraft without a pilot aboard. Taking the pilot out has, at once, multiple implications. The decrease in aircraft weight and size improves efficiency; safety requirements that constrain the design of manned aircraft are rendered irrelevant; applications that would put the pilot in harm's way are made possible. For all these reasons, drones exhibit greater scope, longer range, more endurance — features that have enabled them to occupy several niches that would otherwise be impractical or

impossible to fill. Despite their small, but growing, civil applications however, drones are predominantly employed for military purposes, i.e. targeting and eliminating people and structures from remote distances across the globe.

It is easy to underestimate the cognitive labor that is saved via automation in drones. Nevertheless, compensations for cognitive labor by automatic systems might be misleading in two ways. First, they run the risk of pulling the wool of autonomy over our eyes. In a related and more important sense, the capacity of drones for action at a distance is accompanied by a gap that Suchman (Suchman, 2013) has called the “the problematics of separation.” The most uncomplicated aspect of this separation is the physical distance. Suchman, however, compounds geographical separation with a “reorganization of the human sensorium” that has always accompanied technologies of war.

We flesh out some of the ways in which this reorganization takes place with the intention of bringing out, as it were, the crosscutting dimensions that give Suchman's separation a complex character. Broadly construed, we frame the problematics in terms of spatiotemporal and psychosocial divides between the location of the aircraft (i.e. the locus of sensation/action) on the one hand, and the location of the operating team (i.e. the locus of decision-making) on the other. We draw on insights from ecological psychology and phenomenology, as well as the limited but telling published account of Lt. Colonel Matt Martin, who has provided a close-up view of his thoughts, decisions, and actions as a drone pilot.

Katleen Gabriels, 'I keep a close watch on this child of mine': A Moral Critique of Other-tracking Apps (25th 11:30-12:00)

Smartphones and mobile applications are omnipresent in our present-day lives. At the core of this paper are ‘other-tracking apps’, i.e. mobile applications that make it possible, via location technology, to track others. These apps ensure that we are never unconnected from the network of ubiquitous information and, via that network, from others. Our situation is hence one of ‘continuous connectivity’.

In specific, focus lies on apps designed for parents to remotely track the movement of their child(ren). This particular case can be considered as one example of broader reflection on what continuous technical connectivity means in moral terms. Other-tracking apps give new ground to moral queries related to information technologies. A remarkable field of tension is at play here: on the one hand, the app promises continuous connectivity and ‘togetherness’ with the other; on the other hand, there is the distance of the other in this network of data and information.

Throughout this paper, a number of concerns are expressed and discussed. Ultimately, our concern is that these apps bring us too close to the other, as they engender a situation of ‘overproximity’. An ethical framework is required that emphasizes maintaining the proper, critical distance, to respect the other’s autonomy and privacy.

Keynote Address: Frances Grodzinsky, *Big Data: The Patterns of Our Lives through the Lens of the Virtues* (24th 8:30-9:30)

Big Data is big news. We find references to it everywhere from our local media to scholarly journals. The volume (The quantity of data generated), velocity (speed of generation and analysis) variety (of sources) (traditional 3 v’s) and I would add a 4th, veracity (quality) of the data comprising this socio-technical system raises a number of ethical, empirical, technical issues concerning its generation, collection and use. Patterns are a reliable sample of traits, acts, tendencies, or other observable characteristics of a person, group, or institution (www.merriam-webster.com/dictionary/pattern). Through Big Data analytics, patterns drawn from our behavior emerge that affect us directly giving rise to social issues: the ads we get on our social media sites, and indirectly: the loan we are denied. This talk will analyze Big Data through the lens of the intellectual virtues: what is it; why it is of interest; its methodologies, stakeholders and principles. Using virtue ethics, I will examine what will it take to make this data driven society a “good” society -- i.e., a "flourishing" society in the Aristotelian sense?

Soraj Hongladarom, *Big Data, Digital Traces and the Metaphysics of the Self* (23rd 10:30-11:00)

The main orientation of this paper is not directly an ethical one, though there are obvious ethical implications as we shall see. On the contrary the focus of the paper will be on metaphysics, especially the problem of the constitution of the self in the context of advances in computer networking and ubiquitous computing. Basically the questions to be addressed are: What are the relations between the digital traces one leaves when one appears on line and the conception of one’s own self and one’s identity? In the context of Big Data, where huge information is shared and transmitted over the network, will the self be threatened and reduced to mere numbers or mere set of information? Are we in danger of being reduced to just a set of information? Is “who we are” utterly dependent on our places, or our reputation, in the online world? I attempt to provide some tentative answers to these questions in turn:

- What are the relations between the digital traces and our identity?
- Will the self be threatened and reduced to mere numbers or mere set of information?
- Is “who we are” utterly dependent on our place, or our reputation, in the online world?

Keynote Address: Deborah Johnson, *Getting a Handle on Big Data Ethics* (23rd 11:30-12:30)

A good deal of controversy and public debate was generated last year when the public learned that Facebook had done research on its customers in collaboration with academic researchers at Cornell. The research sought to determine whether the emotional content of users' postings would be affected by manipulating the content of their news feeds. Some customers were exposed to negative emotional content in their news feeds and others were exposed to positive emotional content. Comparing the two groups, the researchers concluded that 'moods are contagious'. Users who saw more positive posts wrote more positive posts and users who saw more negative content wrote more negative posts.

In this paper, I will use this incident and public discussion of it as a starting place for thinking about big data ethics. Public discussion focused largely on the fact that Facebook was doing research on its users. Some emphasized that users had signed a user agreement that, in effect, gave Facebook permission to do such research; others noted that the case was special because it involved not just manipulation but manipulation of emotions. Much, though not all, of big data analytics involves research on human beings and this seems to frame such research as human subjects research. Although most of the literature on this has focused on informed consent, there is much more to be learned from treating big data analytics – using personal data – as human subjects research falling under the umbrella of the Belmont Report.

After exploring the implications of framing big data analytics as human subjects research, the paper discusses other possible frameworks. The best framework is the one that will help to understand both how to think about limits for big data analytics and how to think about the responsibilities of big data scientists.

Stephen Lilley & Frances Grodzinsky, *Should Social Physics Govern the Hyper-Connected, Big Data Age?* (24th 10:30-11:00)

From a historical perspective, social physics is Auguste Comte's unrealized vision. He imagined social physics to be the pinnacle of science in terms of methodology and societal prominence. It would elevate quantitative mathematical analysis of social patterns and become an apex institution directing social reform along secular humanist lines. Comte proposed this in the 1830s in a world lacking supportive administrative and technological infrastructures. Now Alex Pentland, Director of the MIT Media Lab Entrepreneurship Program and prominent network and data scientist, has re-imagined social physics for our hyper-connected, big data age. It too is a bold, compelling model for rational social ordering. It too prioritizes social research and social reform over vested business and political interests.

Pentland and others have developed pilot tests of social physics with “living labs” through which sensors continuously record social and physical data of community members, statisticians run sophisticated analyses of the massive data sets, and social engineers conduct controlled experiments to better “tune” the social networks to yield collective benefits (Pentland, 2014:13-38). He would like to see an exponential increase in sensors and network connections in order to create data-driven societies. This appears to be the way the world is heading with social media, wearable devices, and the Internet of Things, so it is more relevant now than ever to discuss whether social physics should be realized. In this paper we evaluate social physics utilizing James Moor theory of just consequentialism (1999). Although we see clear benefits for public health, we recommend the following checks on social physics to preserve autonomy in an increasingly potent surveillance society: 1) research protections for human subjects, 2) a critical perspective, and 3) attention to the subjective experience of individuals.

Tobias Matzner & Regina Ammicht Quinn, *What Does Ethics do in Science and Engineering Research Projects? An Ethical Self-reflection in the Case of Security Research*
(25th 10:30-11:00)

Many researchers working in ethics of information technology are more or less closely involved with the research and development of technology. Some do research on values in R&D processes, some are counseling, some are working as fixed members of a project team to include ethical issues early on in the design process. Interesting contributions at the CEPE and IACAP conferences over the years have analyzed how ethical issues can be introduced into the design process of products but also into the working environments of R&D. Yet, these structures of involvement also pose ethical questions – for the ethicists themselves.

We have carried out a research project that analyzed the ethical implications of automated video surveillance (often euphemistically called “smart CCTV”), funded by the German Ministry for Education and Research. This project focused particularly on a number of projects, which were developing automated video surveillance technology, also funded by the German Ministry for Education and Research. These technological projects had their own team members from law and the social sciences, which were meant to circumvent societal problems the technology could cause from the very start and ascertain an ethically and legally sound progress. We take this situation and the self-reflection of our research as the basis to assess some fundamental questions for applied ethics. The particular features of automated video surveillance as algorithmic technology in the normatively charged context of security are used as illustration for the ethical issues discussed; at the same time we provide generalizations for other areas of applied ethics.

We assess four core topics:

1. The possible self-concept of ethics and its aims.
2. The legitimacy of ethical influence.
3. The need for collaboration.
4. The role of ethics as a social actor in democratic societies.

Steve McKinlay, *Maker's Information: A Cognitive Approach* (22nd 10:30-11:00)

Recently there has been some discussion regarding the nature of information from the maker's perspective, and its consequent implications for epistemology as well as the philosophy of information. In a 2013 lecture, Luciano Floridi outlined a somewhat austere logical approach to the problem. In this paper I unpack a modest cognitive account of the maker's information, one which may also be applied from a computational perspective, at least theoretically. I outline four components, in terms of knowledge and information, that are required for a maker to hold maker's information. They are as follows, The information yielding event, Prior information, Confirming conditions and Discriminating knowledge. These are discussed in detail in the paper and provide an outline towards a cognitive approach to maker's information.

Keith Miller, Marty J. Wolf, & Frances Grodzinsky, *This Ethical Trap is for Roboticists, Not Robots: Why use a robot when a human will do?* (24th 15:30-16:00)

In a September 2014 article in *New Scientist*, Aviva Rutkin reported on a virtual experiment by roboticist Alan Winfield. In the experiment, Winfield and his team programmed a virtual robot avatar to prevent a virtual human avatar from falling into a hole as the avatars traversed a virtual space. When the scenario had one robot and one human, things went as planned: the robot avatar "pushed" the human avatar out of danger. But when a second human entered the picture, the robot had more trouble. Out of 33 trials, sometimes the robot saved one human, a few times it saved both, but 14 times, the robot managed to save neither of the humans. The author of the article reporting on this experiment wrote that Winfield had "built an ethical trap for a robot."

The work by Winfield, et al. and the subsequent article by Rutkin raise a number of ethical questions that we intend to address in this paper. In the next section we will consider the ethical implications of experiment design in ethical decision making for autonomous agent (AA) research. We will argue that a researcher in this area has to consider two things. The first is the appropriateness of the ethical "task" in the experiment. That is, the experimental situation the AA is tested in should reflect an appropriate use of that kind of technology. The second is that much like a software developer bears responsibility for understanding the socio-technical context the software will be deployed in, the researcher exploring AA ethics

must consider the socio-technical context in which the results of the research will be deployed and understood.

In section 3 we will illuminate some of the issues that are part of the socio-technical context that AA ethics researchers ought to consider. The article by Rutkin demonstrates that even though a researcher might be careful to avoid a particular characterization, it is still possible that someone reporting on the research will not exercise the same level of care. In particular, we will argue that Rutkin's calling the simulation an “ethical trap” for the simulated robot is not an accurate description and will justify our contention in this section.

Even though we object to Rutkin's characterization of the Winfield experiment as an ethical trap, it is important to consider what might motivate such a term, and the role that sort of thinking might lead to in terms of AA ethics research. We take up these issues in section 4. In particular, we argue that this mis-characterization reveals an important mis-understanding of issues crucial to understanding responsibility, accountability, and morality when these terms are applied to artifacts. In addition, we analyze the reasonableness of a variety of conditions that might be part of an ethical trap experiment for a robot. These conditions include aspects of the robot and the situation in which the robot is placed. Although Winfield's experiment does not fulfill these conditions, there is nothing that precludes that an experiment might meet these conditions in the near future.

Brent Mittelstadt, *Self-Care, Patient Empowerment and the Virtue of Personal Health Technologies* (23rd 8:30-9:00)

Recent years have seen an influx of medical technologies capable of remotely monitoring the health and behaviours of individuals to detect, manage and prevent health problems. Known collectively as ‘Personal Health Monitoring’ (PHM), these systems are intended to supplement medical care with health monitoring outside traditional care environments such as hospitals, and range in complexity from single-sensor mobile devices to complex networks of sensors measuring physiological parameters and behaviours. PHM enables new forms of medical care and interactions between patients, clinicians, medical institutions, technology developers, care providers and others, yet the ethical dimensions of these potential futures receive little attention from developers and academics alike. Within this ‘new’ medicine gaps have opened between traditional norms of ‘good’ practice and care mediated by longitudinal monitoring of medical conditions. This paper explores the implications of this gap, suggesting how the moral obligations found in the doctor-patient ‘healing relationship’ are being subtly transferred to the community in the guise of ‘self-care’ and empowerment. A theoretical framework based on Alisdair MacIntyre’s virtue ethics is advanced and applied to the case of PHM deployment in the United Kingdom. Application of the framework demonstrates the broader contribution of virtue-based theories for ethical assessment of

emerging personal health technologies, considered in contrast to dominant utilitarian frameworks which tend to emphasise cost-efficiency.

Brent Mittelstadt & Luciano Floridi, *Contextual Dissonance in Big Data* (23rd 11:00-11:30)

Data can now be generated, aggregated and assessed at an unprecedented rate and scale, heralding what is often referred to as the era of ‘Big Data’ (Bail 2014). Analysis of large-scale, aggregated, continuously growing datasets (IBM 2014) is said to provide unprecedented opportunities for business, policy and research, allowing for products, policies and knowledge which are more efficient, targeted and insightful. One area of particular promise for such data sciences concerns understanding human behaviours and social reality at multiple levels of abstraction, from individual to group (or demographic) to population. Such methods of understanding are undoubtedly of enormous potential value across areas such as public health policy, advertising and resource provisioning, creating opportunities to promote the ‘common good’ through decision-making informed by improved understanding of the behaviours and characteristics of the affected populations.

The appeal of such human-oriented Big Data, or large-scale data sciences which seek to understand and explain human behaviour, is as undeniable as its fundamental assumption is straightforward: decision-making becomes more effective in step with knowledge about the affected phenomena. A future is suggested in which greater collection and assessment of data about the world and the humans occupying inevitably benefit the common good. Questions regarding the ethical acceptability of such a future and the methods of data collection and analysis required to realise it undoubtedly need to be answered as we increasingly move into this new era. Beyond the ethics of the consequences of Big Data, consideration must also be given to the underlying structures of social interactions implied within such a future. In other words, what are the characteristics of social interactions implied by the new modes of understanding made possible through Big Data? Are these characteristics ethically problematic?

This paper explores a gap in knowledge concerning the ethical implications of the epistemic structure of Big Data which may limit opportunities for meaningful participation and protections within public life and decision-making for data subjects, both as individuals and groups.

Alexander Ovsich, *Mathematical Models of Desire, Need and Attention* (22nd 11:30-12:00)

Desire plays an important role in the explanation of behavior in general, for example, in the contemporary Belief-Desire theories. These theories (for example, Bratman’s Belief-Desire-Intention theory) are widely used in the AI applications. However, there is neither much

literature, nor even consensus about the meaning and definition of desire. There is not much clarity about the concepts and mechanisms of need and attention either.

The author presents here simple, closely linked mathematical models of desire, need, and attention. They are based upon the hedonistic principle proclaiming that animals and humans alike are driven by striving to maximize pleasantness of their internal state (Pleasantness of the State of this Subject ("PSS")). What directly follows from this principle is that for such a subject (S), the most important characteristic of any phenomenon (X) should be how much X influences the process of maximization, how much X increases or decreases PSS that is measured by the magnitude and direction of its change (Δ PSS).

Keynote Address: William J. Rapaport, *On the Relation of Computing to the World*
(22nd 9:00-10:00)

I survey a common theme that pervades the philosophy of computer science (and philosophy more generally): the relation of computing to the world. Are algorithms merely certain procedures entirely characterizable in an “indigenous”, “internal”, “intrinsic”, “local”, “narrow”, “syntactic” (more generally: “intra-system”) purely Turing-machine language? Or must they interact with the real world, with a purpose that is expressible only in a language with an “external”, “extrinsic”, “global”, “wide”, “inherited” (more generally: “extra-” or “inter-”system) semantics?

Michele Rapoport, *Choosing and Acting Freely in Smart, Persuasive Environments*
(25th 8:30-9:00)

In addition to fostering an environment that is efficient, comfortable and productive, smart devices can be assigned singular tasks including motivating and supporting healthier lifestyle choices or encouraging the enactment of various social norms and prescriptions. Thus, for example, recent decades have seen mobile applications that assist users in overcoming smoking habits, smart refrigerators that monitor food consumption and promote healthy eating routines, bathroom scales that monitor weight, and devices that restrict the consumption of energy and natural resources. These have been rapidly adopted in the many spheres of human life and have been met with particular interest among health-related service providers.

This presentation will focus on smart, persuasive technologies that are bound to the living of daily life and that encompass the routine and quotidian chores, tasks and assignments that humans engage in regularly, focusing on smart devices that either act or replace human action and that affect the tangible, material world. Smart devices with persuasive capacities have garnered much interest among researchers who see them as “fundamentally about

learning to automate behavior change” and “effectively encode[ing] experiences that change behaviors”. Ultimately, these devices are concerned with behavior and action; their *raison d’être* is to attain desirable results, to evoke certain performances, and to ensure that certain procedures are enacted which will bring about sought-after consequences. They are programmed to instate ‘correct’ and ‘proper’ behavior, behavior which at times may conflict with choices that otherwise might have been made by the user.

Keynote Address: Michael Rescorla, *Levels of Computational Explanation* (24th 16:15-17:15)

Orthodoxy holds that one can fruitfully describe computation at three different levels: the hardware level, the syntactic level, and the representational level. I argue that the orthodox view is true of artificial computing systems but not biological computing systems. Formal syntactic description of biological computation often adds no explanatory value to hardware description and representational description. Overemphasis on formal syntactic description has led many philosophers to underrate the explanatory role that mental representation plays within psychological explanation. I support my position with evidence drawn from computer science, robotics, and cognitive science.

Luke Roelofs, *Rational Agency without Self-Awareness: Could ‘We’ replace ‘I’?* (23rd 9:00-9:30)

It has been claimed that we need singular self-knowledge - knowledge involving the concept ‘I’ - in order to function properly as rational agents. An implication of this is that any artificial intelligence or cyborg which was to be capable of rational agency would need to be built to recognise and identify itself as such - even if it were intended to operate in constant close integration with other agents. I argue that this is not strictly true: agents in certain special relations could dispense with singular self-knowledge and instead rely on plural self-knowledge - knowledge involving the concept ‘we’. I discuss three specific functions for which singular self-knowledge has been regarded as indispensable: connecting non-indexical knowledge with action, reflecting on our own reasoning process, and identifying which ultimate practical reasons we have. I argue in each case that by establishing certain relations between agents - relations I label ‘motor vulnerability’, ‘cognitive vulnerability’, ‘evidential unity’, and ‘moral unity’ - we would allow those agents to do everything a rational agent needs to do while relying only on plural, rather than singular, self-knowledge.

Paul Schweizer, *Cognitive Computation sans Representation* (22nd 11:00-11:30)

According to the traditional conception of the mind, semantical content is perhaps the most important feature distinguishing mental from non-mental systems. And this traditional conception has been incorporated into the foundations of contemporary scientific approaches

to the mind, insofar as the notion of ‘mental representation’ is adopted as a primary theoretical device. Symbolic representations are posited as the internal structures that carry the information utilized by intelligent systems, and they also comprise the formal elements over which cognitive computations are performed. But a fundamental tension is built into the picture - to the extent that symbolic ‘representations’ are formal elements of computation, their alleged content is completely gratuitous. I argue that the computational paradigm is thematically inconsistent with the search for content or its supposed ‘vehicles’. Instead, the concern of computational models of cognition should be with the processing structures that yield the right kinds of input/output profiles, and with how these structures can be implemented in the brain.

Johnny Hartz Søraaker, *Beyond Right and Wrong - Applying Prudential-Empirical Ethics of Technology (PEET) to Autonomous Cars* (24th 14:30-15:00)

Computer ethics has a long tradition of addressing ways in which ICT may diminish core values like privacy, security, equity, trust and accountability. As important as these issues are, there is a need to also go beyond issues of right and wrong – we need to pay more attention to how ICT may serve to enhance quality of life. Similarly, computer ethics scholars have been hard at work providing engineers and designers with tools and frameworks for ‘responsible research and innovation’ (RRI) and technology assessment. There are, however, few if any comparable frameworks to support engineers in understanding, assessing and designing for well-being. Prudential-Empirical Ethics of Technology (PEET) aims to be such a framework. It draws its conception of well-being from empirical research (positive psychology, in particular), provides a methodology for “translating” these empirical findings into concrete technological features, and couples all of this with a robust notion of responsible research and innovation. The purpose of this extended abstract, to be elaborated in the full paper, is to introduce the framework and illustrate its implications by means of an analysis of autonomous cars.

John P. Sullins, *Finding the Path to Beneficial AI* (24th 11:30-12:00)

Some very smart, rich, and powerful people such as Steven Hawking, Elon Musk, and Bill gates are warning us of the potential extinction event that super intelligent Artificial Intelligence represents. This does not seem to be idle talk or speculation on their parts but statements of legitimate concern. In fact Elon Musk has put up ten million dollars towards the funding of researching ways to control any such machine intelligence that might surpass the capabilities of human minds. Part of this grant supports the Future of Life Institute, whose motto is, “Technology has given life the opportunity to flourish like never before... or to self-destruct”. This institute has drafted an open letter on the potential danger of deploying

uncontrolled super intelligent AI applications which has been signed by hundreds of professionals and concerned citizens and produced a report on research priorities needed to develop beneficial AI.

It is argued here that research in the ethics of building super intelligent AI need not ignore short term research such as the recent developments in machine and robot ethics. It is not helpful to build a false dualism between the possibility of the emergence of superintelligence and the more modest gains of every day AI appliances. Given that any advanced form of AI is likely to be a development from earlier designs, if the earlier designs are ethically controlled and verified, then advanced AI applications built on that foundation will have a better chance of being beneficial to humanity. In that case machine ethics is not doomed to eventual irrelevancy; in fact it is of vital importance.

John Symons & Jack Horner, *How Does Software Limit Scientific Inquiry?* (22nd 12:00-12:30)

Computer scientists call a system software intensive if “its software contributes essential influences to the design, construction, deployment, and evolution of the system as a whole” (IEEE 2000). By this definition, much of contemporary science involves software intensive systems. The increasingly central role of computing technologies has changed science in significant ways. While the practical import of this transformation is undeniable, its implications for the philosophical understanding of scientific inquiry are less clear. Broadly speaking, the interests to date of philosophers of science in software-related topics have concerned computational modeling: characterizing how models, in the sense philosophers of science have typically used that term, play in computational regimes. Our objective in this paper, in contrast, is to characterize an important software-based difference between non-software-intensive science and software intensive science and to explain why this difference is relevant to philosophy of science.

Here we argue that, in general, there is no practical way to characterize the error distribution in a software system containing more than a few hundred instructions (roughly, “lines of code”). To the extent a scientific domain epistemically depends on software systems larger than a few hundred lines, therefore, in general the consequences for a scientific domain of our ignorance of that error distribution will remain unknowable.

Mariarosaria Taddeo & Luciano Floridi, *The Responsibilities of Online Service Providers in Information Societies* (25th 9:30-10:00)

In this article we will review the main moral responsibilities ascribed to Online Service Providers (OSPs) in the past two decades. We will focus specifically on two sets of responsibilities. The responsibilities that OSPs bear as agents acting in the web, and those

that OSPs have as agents acting on the web. The former pertain to the regulation of access to information, such as copyrighted material, freedom of speech, censorship, and transparency. The latter are the responsibilities that OSPs bear with respect not so much to the content of the web, but to the access to the metadata concerning the use of the web. The debate on the role of OSPs in collaborating with the US government during the PRISM program concerns OSPs' responsibilities on the web, while the discussion on Yahoo! and Google compliancy with the request of the Chinese government is about the responsibilities of OSPs in the web.

Ingvar Tjostheim, *Telepresence, the Senses and Merleau-Ponty's Theory of Perception* (23rd 9:30-10:00)

The aim of this paper is to discuss the theoretical basis for the telepresence phenomenon with and emphasis on the five senses. We explore theories of relevance for the telepresence phenomenon. In two empirical studies we use a telepresence-measurement that have a sub-construct named perceptual realism. Results from two empirical studies are used to discuss perception and the telepresence experience.

Keynote Address: Shannon Vallor, *AI and the Automation of Wisdom* (23rd 14:30-15:30)

I reflect on the various ways in which automated systems are becoming increasingly involved in the exercise of human judgment, both moral and intellectual, and the implications for human self-cultivation and virtue. After articulating the distinction between machine 'decisions' and machine 'judgment' in the richer sense associated with moral and intellectual wisdom, I will survey a number of scientific, professional and personal domains in which automated systems are increasingly envisioned, and in some cases already positioned, in roles historically reserved for those with wisdom. Drawing upon the philosophical traditions of virtue ethics as well as early modern political theories of agency, I call for a fresh conception of human-machine relations that preserves both the moral and intellectual conditions for living wisely and well.

Aimee van Wynsberghe, *Ethical Data Provenance* (24th 9:30-10:00)

The Carna botnet port scanning scandal of 2012 raised many ethical issues concerning the sharing and use of data when collected in an unethical or highly troublesome manner. The release of data in 2010 obtained from Facebook in the T3 project raised an entirely new set of questions pertaining to the use of data from online social networking sites and risks to anonymity (or lack thereof). The NSA scandal of 2013/4 whereby information of citizens was collected from a variety of sources without consent or knowledge and given to the government for storage and analysis is another example of ethically troubling data sharing.

What all of these cases have in common is that they involve the sharing of data with others. The main problem currently facing computer scientists, computer ethicists and society-at-large is how to ethically evaluate current data sharing practices. Added to this is the question of how ethics can help in data sharing practices.

Many computer scientists and engineers working in either an academic or industrial context are faced with ethical dilemmas pertaining to the sharing of data; should they share data? Where did the data come from? Who has had access to the data? Is this research repeatable? These questions are not new; however, their significance has escalated in the 2010's given: the complexity of data sharing practices (i.e. that a University will have a relationship with a company to share flow or metadata with them), the frequency of these practices, and the lack of understanding of the related practical ethical issues. At the core of questions such as these is the concept of data provenance; the documentation of the origins and ownership of data.

This paper is intended to provide a framework for ethical data provenance to help answer the ethical questions researchers have about the data they work with. The framework is intended to be used by computer scientists and engineers for their own education as well as providing insight for ethicists as to the novelty, complexity and variability of ethical issues related to data sharing and data provenance practices.

Mario Verdicchio, *A Critique of Machine Ethics from the Perspective of Autonomy*
(22nd 14:00-14:30)

In the twenty-first century, video games are ubiquitous. Moving beyond the traditional model of a single player sitting at his computer, games have proliferated wildly. People play casual games on their phones while waiting in line for groceries. They play games online with their friends, whether through social media pages such as Facebook or in Massively Multiplayer Online Role-playing Games (MMORPGs) like World of Warcraft. They can download games easily from Apple's App Store, from third-party platforms such as Steam, or from retailers such as Amazon.com. People can access and play games more easily and in a wider range of environments than ever before. With the increasingly pervasive reach of video games, it is important to consider their moral ramifications; in this paper, I address a variety of questions pertaining to ethics and video games.

Drawing upon discussions of virtual worlds such as LambdaMOO and World of Warcraft, I argue that it is clearly possible to cause harm within a multiplayer game. Any environment in which it is possible to help or hinder another player is one in which it is possible to display morally virtuous or vicious traits; for instance, if it is possible cause frustration or a sense of violation in another player, then it is possible to harm them. Actions which cause that sort of harm have moral status, although we are far from agreeing on how much weight to place on

such harms or how best to redress harms caused in such a way.

A more interesting question is raised when we consider single-player games, as there are no longer any other people who can be harmed or benefited by a player's actions. I raise the question of whether our actions can have moral standing within such a game, distinguishing between two separate sets of questions worth addressing in this context. First, we can consider actions from inside the game world. Hence we may ask how a game handles morality within its own world. How do the player's choices influence the game? Is morality measured with some kind of rigid meter – which generates its own set of problems – or is it more subtly reflected in the changing attitudes of non-player characters within the game? What assumptions are being made by the game's designers in the design of the game world's morality? Does it matter if a player takes an action or only if the character is observed to take the action by another character in the game? All of these questions raise difficult moral issues, and I consider the ramifications of various choices that game designers make.

Wendell Wallach, *Deep Learning, AI Safety, Machine Ethics and Superintelligence*
(24th 12:30-13:00)

Many philosophers, including myself, have expressed skepticism that superintelligent AI systems might be realized in the next 50-100 years. However, recent breakthroughs in computer perception and learning have made formerly skeptical AI researcher consider superintelligence quite likely. Should philosophers also give this prospect more credence, and if so, how might that change philosophical contributions towards AI safety and machine ethics? Deep learning, a relatively recent approach to machine learning, has created renewed excitement among computer scientists for making significant advances toward AI over the coming decade. Should deep learning be seen as truly a breakthrough, or might it be another promising yet limited approach, similar to the former enthusiasm around neural networks and ALife? This presentation will discuss how philosophy might help illuminate the value and possible limits on deep learning as an approach for developing artificial intelligence. It will also outline a deep learning proposal for developing computers and robots capable of making moral decisions.

Kirk A. Weigand & Ronald L. Hartung, *Below Global Coherence: Turtles, No; It's Situations All the Way Down* (23rd 13:30-14:00)

Process philosophy, developmental neuropsychology and cognitive science are applied to the problem of modeling global coherence in narrative such that unconscious experiences may meaningfully be represented as threads that cohere to motifs and overall story purpose. A computer science framework for a computational reasoning architecture is designed and

tested to show that bottom-up meaning vis-à-vis top-down purpose mutually couple relevance and context. This reasoning framework uses compartmentalized knowledge habits. Situations are viewed as composable experiences in an abstraction hierarchy, and situated agents inherit, shape and pass meaning in conceptual spaces across time. This framework proposes a communication protocol for agents representing fundamental, short-lived experiences that relevantly cohere to achieve greater harmony.

James Williams, *What Is a 'Persuasive' Technology?* (25th 9:00-9:30)

While it has become popular to describe some technologies as being ‘persuasive,’ we lack clarity as to what this actually means. In this paper I will propose a conceptual framework for persuasive technology that aims to contribute to the current understanding in three ways. First, I propose a way to more fully consider the question of intention when determining a technology’s ‘persuasiveness.’ Second, I disentangle normative conditions—in particular the requirement that persuasive technologies be noncoercive—from the descriptive conditions of their definition. Finally, I suggest an approach to the language of persuasive technologies that will allow us to handle questions of intention, as well as normative considerations, in a fuller and more nuanced way.

Richard Wilson, *Anticipatory Ethics and Brain Computer Interfaces (BCI's)* (23rd 14:00-14:30)

Brain computer interfaces (BCI’s) have currently developed to a level where engineers and physicians can employ them in experimental and therapeutic settings. The promise of BCI’s for those suffering from a variety of disorders is that BCI technology can move from labs to everyday life, where ordinary users will be able to set them up and use them in everyday circumstances. Ethical and social questions are developing as the technology continues to develop and as researchers, engineers, and users imagine a world where BCI’s have gained widespread use. Ethical questions include such issues as physical risk for users relating to all the stakeholders involved. As BCI’s become less invasive general philosophical issues arise involving agency, individual bases of responsibility, and the changing of the living experience of short term and long term BCI use. All of these issues are interrelated and need to be addressed as a technological paradigm is developed for human machine interactions.

In order to gain a realistic view of the risks, opportunities, and to anticipate future uses of BCI technology, the scientific, medical, anthropological, and societal aspects of effects of BCI’s needs to include a phenomenological description of the living experience of the user’s of BCI’s. Central to the design, experimental use in labs and eventual deployment into everyday life of BCI’s should be a reflection upon the living experience of the users of BCI’s. The descriptive analysis undertaken in this paper employs Husserl’s notion of the

intentionality of consciousness as the bases of this analysis. This descriptive analysis focuses on the 1st person accounts of BCI user's and the 1st person accounts of relatives of BCI users. Central to this analysis is the idea that in order to anticipate ethical issues with BCI's all of the stakeholders involved in the development of BCI technology need to focus on the living experience of BCI user's which should be crucial to current and prospective BCI use, as well as to the future of BCI developments.

Bendert Zevenbergen, *Philosophy Meets Internet Engineering: Examining and Aligning Ethics Reasoning* (22nd 14:30-15:00)

Generally, computer scientists and network engineers tend to apply consequentialist reasoning to research project design - whereby the end justifies the means - to meet research objectives in the most efficient way. However, hardware and software are increasingly 'black boxes' for average people, which makes it difficult for an individual, ethical boards, policy makers, or affected communities to scrutinize the design choices and evaluate the new moral systems created. This can be problematic, because these stakeholders typically reason based on deontology, where actions and design choices are considered more important than the achieved goals.

Primarily, this paper will report on the outcomes of a one-day interdisciplinary workshop held at the University of Oxford on 13 March 2015. Twenty Internet engineers, philosophers, and practical ethicists will discuss several distinct scenarios, where Internet engineering research projects are faced with complex ethical dilemmas. The aim of the workshop is to understand the reasoning of the different groups. The outcomes of the workshop will feed into a larger project that aims to establish guidelines for future Internet engineering research designs. This paper will draw on both computer science and Internet engineering literature, as well as apply ethics methodologies and reasoning to research projects.