17th Conference of the International Society for Utilitarian Studies

Celebrating the 150th anniversary of the publication of Henry Sidgwick's *The Methods of Ethics*

Bentham House, Faculty of Laws
University College London
18–20 June 2024

Table of Contents

3

Key Speakers

Bart Schultz (Chicago). The Methods of Ethics *as Prolegomenon: the Long, Strange Trip of Henry Sidgwick*.

Chair: Emmanuelle de Champs (Cergy Paris).

—Tuesday 18 June, 11.30–13.00. Gideon Schreier Lecture Theatre (124).

Bart Schultz is Senior Lecturer in Humanities (Philosophy) and former Director of the Civic Knowledge Project at the University of Chicago, where he has been teaching since 1987. He has published widely in philosophy and other disciplines, and his books include *Essays on Henry Sidgwick* (Cambridge, 1992); *Henry Sidgwick: Eye of the Universe* (Cambridge, 2004, winner of the American Philosophical Society's Jacques Barzun Prize in Cultural History); *The Happiness Philosophers: The Lives and Works of the Great Utilitarians* (Princeton, 2017), and *Utilitarianism as a Way of Life: Re-Envisioning Planetary Happiness* (Polity, 2024). He edited and contributed to, with G. Varouxakis, *Utilitarianism and Empire* (Lexington, 2005), and he edited and collaborated on the memoir of the civil rights movement legend Timuel D. Black, *Sacred Ground: The Chicago Streets of Timuel Black* (Northwestern, 2017). He is on the Editorial Board of Utilitas, the leading professional journal of utilitarian studies, and he was a founding board member of PLATO, the Philosophy Learning and Teaching Organization. He is currently working on a new edited volume, *The Classical Utilitarians: Essential Readings* (Broadview Press, forthcoming).

Katarzyna de Lazari-Radek (Łódź). *On the Notion of Pleasure—How Sidgwick got it right*.

Chair: David Weinstein (Wake Forest).

—Wednesday 19 June, 11.30–13.00. Denys Holland Lecture Theatre (SB31).

Katarzyna de Lazari-Radek is a professor at the Faculty of Philosophy, University of Lodz, Poland. She is a hedonistic utilitarian. Her main research interest focuses on the philosophy of Henry Sidgwick, the concept of wellbeing, and pleasure. Together with Peter Singer she wrote two books: *The Point of View of the Universe* (Oxford University Press, 2014) and *Utilitarianism - A Very Short Introduction* (Oxford University Press 2017) and edited *J.S. Mill's Utilitarianism* (Norton, 2021). In May 2024 Routledge is going to publish her introduction to *The Philosophy of Pleasure*. Apart from academic work, she is keen to convey philosophical ideas to a popular audience, and has written on how to live a good life for popular magazines.

Peter Singer (Princeton). *Sidgwick and Other Influencers*.

Chair: Roger Crisp (Oxford)

—Wednesday 19 June, 17.45–19.00. Denys Holland Lecture Theatre (SB31)

Peter Singer is the Ira W. DeCamp Professor of Bioethics in the University Centre for Human Values at Princeton University. His books include *Animal Liberation* (1975), *Practical Ethics* (1980), *How Are We to Live?* (1993), *The Life You Can Save: Acting Now to End World Poverty* (2009), *The Point of View of the Universe* (2014, with Katarzyna de Lazari-Radek), *Utilitarianism—A Very Short Introduction* (2017, with Katarzyna de Lazari-Radek), and *Why Vegan? Eating Ethically* (2020). In 2012 Singer was made a Companion of the Order of Australia and in 2021 received the Berggruen Prize for Philosophy and Culture.

John Skorupski (St Andrews). *Sidgwick's Nihilism.*

Chair: Lisa Forsberg (Oxford)

—Thursday 20 June, 11.30–13.00, Gideon Schreier Lecture Theatre (124)

John Skorupski is Emeritus Professor of Moral Philosophy in the University of St Andrews. He writes on normativity and reasons, and on modern moral and political philosophy, including its history. His books include *John Stuart Mill, The Domain of Reasons, and Being and Freedom*. He is a fellow of the British Academy and of the Royal Society of Edinburgh.

Roundtable: Katarzyna de Lazari-Radek (Łódź), Bart Schultz (Chicago), Peter Singer (Princeton), John Skorupski (St Andrews).

Chair: Roger Crisp (Oxford)

—Thursday 20 June, 16.00–17.15, Gideon Schreier Lecture Theatre (124)

Panel 1A—Moot Court (XG03)

David Phillips (University of Houston). *Can Sidgwick and Ross Converge in Moral Theory?*

I argue that the answer to my titular question is: to a surprising and interesting extent, yes.

Despite their apparently deep disagreements, their views about the right can be combined. Both are committed to the reason to promote the good. Both also think some agent-relative addition is required, but they disagree about what it should be. I argue that Sidgwick is right about partial reasons and that Ross is right about deontological reasons. I then note that they moot strikingly similar versions of a moderate pluralist view about the good, which, I suggest, they mistakenly deviate from in opposite directions. I argue that conceptually, we should think (as Sidgwick did but Ross did not) in terms of the generically normative and (as Ross did but Sidgwick did not) in terms of the contributory; that is, in terms of normative reasons.

David Weinstein (Wake Forest University). *Sidgwick's Parfit*

Parfit was much influenced by Sidgwick. But Parfit does not so much interpret him as he does appropriate what he insists Sidgwick *could* or *might* have said that would agree with what Parfit says. By agreeing with Parfit, Sidgwick was not only allegedly climbing the same philosophical mountain as Parfit. He also purportedly succeeded in approaching the top alongside him, which, for Parfit, helps justify, or in his words "support," our belief that there are some objective irreducibly moral truths. Sidgwick, then, was not doing something different in a different historical context but was in effect doing the same thing at the same time as Parfit, which would have made Sidgwick plenty happy to have acknowledged Parfit as his climbing partner.

Now Parfit like Sidgwick is keen to reconcile common-sense thinking with consequentialism. Though both Sidgwick and Parfit were what we would now call non-naturalist cognitivists, Sidgwick was a hedonist and Parfit was not, which is another way of saying that Parfit was not a utilitarian consequentialist. Moreover, Parfit was a rule consequentialist more in the fashion of Mill while Sidgwick was as we would now say an act consequentialist.

By taking our common-sense moral convictions seriously as provisionally independently warranted, both Sidgwick and Parfit held that our moral common sense beliefs can be further justified by appealing to what they have in common in making things go better. And insofar as the principle of making things go better systematizes our common-sense moral convictions, this principle is in turn justified. The moral principle of making things go best and moral common sense were not inimical. We ought continually to modify both, making them mutually reinforcing components of ongoing reflective equilibrium. Parfit, though, thought that we should

systematize common-sense moral beliefs, refining them as stringent, non-deontic moral rules whereas Sidgwick eschewed this aim.

Ross too, as far as Parfit was concerned, was headed upwards in a paired direction. He was "one of the best" defenders of non-consequentialist thinking, which also accommodated moral common sense. He was no "absolutist" because for him our moral duties, though robust, were ultimately *prima facie*. We have strong moral reasons not to treat people in certain ways, which common sense likewise insists. Though Ross seems not to have thought that systematizing our *prima facie* moral duties as am exacting code of mutually-reinforcing moral principles was possible let alone wise, he was nevertheless just as keen to make common sense more philosophically coherent. Parfit thinks that Ross's and Sidgwick's theories therefore effectively amount to a "different but compatible explanations" of moral common sense. In other words, Ross much like Sidgwick according to Parfit, was equally some kind of non-natural cognitivist playing metaethical reflective equilibrium.

Accommodating moral common sense philosophically was central to Parfit's aim in proving that there are "object-given, value-based" moral reasons, that some things are morally worth caring about and that non-natural moral truths exist. If we couldn't accommodate moral common sense with philosophical principle then our belief that moral truth existed would lack important justification. We would lack confidence in being able to know anything morally. Or at least we would be unwarranted in claiming that what we took to be moral truth was indeed true. Accommodating moral common sense was equally seminal for Sidgwick and Ross though Sidgwick was a consequentialist and Ross was not or at least not in any kind of traditional sense.

Attending more assiduously to how Sidgwick and Ross appealed to moral common sense both metaethically and ethically, and how Parfit subsequently deferred to both on this score as evidence of a developing, philosophical moral consensus and therefore of moral truth, *may* be more than just interesting intellectual history. It *might* help us now to think more carefully about what makes moral common sense so credible if it is credible at all. But using Sidgwick, Ross and Parfit this way is not to make Sidgwick *anticipate* Parfit since he never could have intended to do so. If he somehow could be said to have done this sort of thing then it would not seem silly to write about "Sidgwick's Parfit" or for that matter "Sidgwick's Ross" or "Ross' Parfit" too.

Rob Shaver (University of Manitoba). *Broad on Sidgwick on Common Sense Morality*

Broad proposes "a form of Intuitionism which is not open to Sidgwick's objections and is not flagrantly in conflict with reflective common sense." There are "two quite different ethical features of...action...viz., its fittingness or unfittingness to the total course of events modified by it, and its utility or disutility" (FT 218-19). Broad adds that "it is very doubtful whether any general rules can be given for balancing one

kind of fittingness against another or for balancing fittingness on the whole against utility on the whole." We are "reduced to something analogous to those perceptual judgments on very complex situations which we have constantly to make in playing games of skill" (FT 222-3).

By making the relevant propositions claims about the fittingness of an action in some respect, Broad avoids the disagreement that would afflict any view that put forward absolute rules, such as "lying is always wrong." He also avoids the disagreement that might afflict any view that put forward judgments about "total rightness," such as "this lie is wrong" (FT 222).

This would not satisfy Sidgwick. Sidgwick wants to avoid "leaving any practical questions unanswered" (ME 102).

Broad might reply by stressing his analogy to games of skill. He "compare[s] right action with playing a ball rightly at tennis" (FT 285).

Sidgwick considers a similar view: "rules and definite prescriptions may do much," but "we leave it to trained insight to find in any particular circumstances the act that will best realise" the virtue (ME 228). Sidgwick's explicit objection to this view is addressed to someone who holds it as a way of defending the position that virtue is the only good (ME 392). That leaves it unscathed as part of a view that recognises fittingness and utility. But Sidgwick's objection is easy to imagine. When asked to explain why a drop shot hit from behind the baseline is a bad choice, the explanation is that the opponent will reach the ball and likely win the point. Losing the point is bad given that the object is to win. It is not clear what the analogous explanation will be for why a given action is right. The end "doing the right action" lacks the content of "winning the match."

Broad is part of a wider rejection of Sidgwick's demand for practical guidance. The "interest of ethics is...almost wholly theoretical....[I]t may have a certain slight practical application. It may lead us to look out for certain systematic faults which we should not otherwise have suspected; and once we are on the look out for them, we may learn to correct them" (FT 285). Carritt has the same view: "moral philosophy cannot prove either that we have duties or what in detail they are," but it helps us avoid "bad philosophy" which might encourage us to do what we otherwise would see to be wrong and it increases agreement (EPT 5-8). When Broad does (unusually) address a practical question—"Ought We To Fight For Our Country?"—he comments that "I have no idea what is the right answer to this question, and, if I had, I should not be able to prove it to people who accepted different ethical principles and premises from those which I accept." He even doubts that this "is the kind of question to which there is an answer" (CE 132).

Panel 1B—Hong Kong Alumni Room (221)

Prioritarianism: definition, justification, scope

Christoph Lumer (University of Siena). *Defining Prioritarianism*

The currently most prominent definition of 'prioritarianism' comes from Matthew Adler: "By "prioritarian" SWFs, I mean all SWFs that, in addition to minimal welfarist requirements, satisfy both the Pigou-Dalton axiom and the separability axiom. A further axiom which we might ask SWFs to satisfy is continuity." (Adler 2012: 356; similar: Adler & Holtug 2019: 103-104) This characterisation via morally intuitively interpretable axioms is then spelt out mathematically as follows: "Any SWF with the following form is both prioritarian and satisfies the continuity axiom: x is morally at least as good as y iff $\Sigma i\ V(u_i(x)) \geq \Sigma i\ V(u_i(y))$, with the V(.) function strictly increasing and strictly concave." (ibid.)

After criticising some aspects of this definition the contribution develops an alternative definition based on the following substantive axioms: Individualism: The well-being of each individual is morally evaluated individually; the moral value of a distribution of benefits is only the sum of these individual evaluations. Pareto principle. Priority in the sense of Parfit's slogan: "Benefiting people matters more the worse off these people are" (Parfit 1997: 213). Smoothness: the moral importance of changing benefits decreases evenly as people's benefits increase, i.e. without a change in the trend of decrease (i.e. not at times slower, then faster again, etc.). Extensional exclusion of leximin: There is at least one pair of real utility distributions that is evaluated with the opposite preference to leximin.

The mathematical spelling out of these conditions is: an SWF is prioritarian if it has the following form: x is morally at least as good as y iff $\Sigma i\ V(u_i(x)) \geq \Sigma i\ V(u_i(y))$ (for short: VP(y)), where V(.) is strictly monotonically increasing and strictly concave and V'(.) is strictly convex; there is also a pair of real utility distributions z and w with VP(z) > VP(w), but for which w is better than z according to leximin.

Matthew Adler (Duke University). *Person-Affecting Axiology and Prioritarianism*

Prioritarianism is often expressed as a moral-betterness ranking of outcomes according to the sum of a strictly increasing and concave function of individuals' well-being. Such a ranking satisfies five axioms (and the converse is also true): Pareto (meaning the combination of Pareto indifference: if everyone is unaffected as between two outcomes, the two outcomes are equally good; and Strong Pareto: if everyone is at least as well off in one outcome as compared to a second, and some are strictly better off, the first outcome is better); Pigou-Dalton (a pure, gap-diminishing transfer of well-being from a better-off to a worse-off person, leaving

everyone else unaffected, is a moral improvement); Anonymity (a permutation of well-being levels is a matter of moral indifference); Separability (the ranking of two outcomes is independent of the well-being levels of those unaffected between the two); and Continuity (if one outcome is better than/worse than a second, this holds true for a sufficiently small region around the first).

A person-affecting axiology is such that the moral comparison of any two outcomes depends on the magnitude and weight of individuals' well-being gains and losses between the two, with well-being gains counting towards moral betterness and well-being losses against. Person-affecting axiology provides a powerful justification for Pareto, Pigou-Dalton, and Anonymity. The class of rankings that satisfy this trio of axioms might be called "extended prioritarianism." This class include not merely prioritarianism proper, but also the sum of rank-weighted well-being; leximin; and prioritarianism with a lexical threshold. In this talk, I review why person-affecting axiology supports extended prioritarianism, and discuss whether it can get us all the way to prioritarianism.

Nils Holtug (University of Copenhagen). *Political Equality or Political Priority*

In distributive justice, there is a lively debate about whether egalitarianism or prioritarianism best captures our distributive concerns. However, there is no similar debate in the case of political justice, which, roughly, concerns the distribution of political power or opportunities for political influence. Here, equality reigns. In the paper I argue that many of the arguments and concerns that apply to the debate on distributive justice also apply in the political domain, including concerns about levelling down, and that it is a mistake to simply rule out what we may call political prioritarianism (roughly, the view that we should give priority to those who are worse off with respect to political power as regards the distribution of such power). First, political egalitarianism is distinguished from political prioritarianism in terms of a number of structural differences and different accounts of the currency of political justice are considered. Then the levelling down objection to political egalitarianism is presented and its force and justification considered. Likewise, certain egalitarians rejoinders are discussed, including one that relies on relational egalitarianism. Finally, two objections to political prioritarianism are scrutinised - the 'dictatorship objection' and the 'aggregation objection'.

Panel 1C—Gideon Schreier Lecture Theatre (124)

Malik Bozzo-Rey (CEThicS, Catholic University of Lille). *What if it turned out that Bentham was the best author to use in updating Foucault's analysis of surveillance?*

This presentation seeks to draw out the consequences of Bentham's position on surveillance in the public sphere. The aim is then to focus on the democratic implications of the constitutional panopticon as elaborated by Bentham, which is totally absent from the analyses Foucault proposes in Discipline and Punish. In particular, I will analyze how the dynamic relations between the ruling few and the subject many can be redefined by the possibility offered to citizens to monitor the activities of civil servants. I believe that this possibility, proposed by Bentham in his theory of democracy, should be taken into consideration to think about surveillance today.

Paulette B. Banciella (University of Zurich). *Utilitarian Influences on the Hispanic American Emancipatory Project: Insights from the 1809 Edinburgh Review*

As a law student, Jeremy Bentham was interested in colonial questions, addressing these topics either indirectly or in chapters throughout his work. In addition to moral judgments of colonialism, he provided consistent financial arguments. Due to Bentham's invaluable contributions to the science of law in both practical and theoretical aspects, as well as his interest in international politics and decolonisation, a question arises: When did Bentham and his circle become close to Francisco de Miranda to influence the Hispanic American emancipatory project? They shared the same social circles in London and Russia since the end of the eighteenth century. Although there are few clues of an early intimacy, the coincidences are suggestive.

Years later, the *Edinburgh Review* showed interest in the scarce publications about Hispanic America that had a politically pro-independence outline. Bentham, James Mill, John Allen, and Miranda were acknowledged as contributors at that time. The journal addressed Miranda's project from a critical perspective, and in a common utilitarian manner, they supported knowledge on statistics and linked them to economic arguments. Bentham's influence is evident in the basis of the emancipatory project, with his concepts of democracy, codification, international economy, international relations, and the goal of a congress to solve international disputes. A global order through a grand congress is proposed as the solution for lasting peace, with economic prosperity based on peaceful interstate relations.

The necessary end of the oppressive system was presented as a benefit to all humankind, advocating for the establishment of a new government led by fairness, pursuing liberal principles, and promoting happiness for its people. Publications from

1809 in the Edinburgh Review will be analysed to explore the link between utilitarian and emancipatory movements: *Emancipation of Spanish America*, on the work of Viscardo; *Molina's Account of Chili*, by Abbe Juan Ignacio Molina; and *The Columbiad: a Poem*, by Joel Barlow, which contains political content.

James E. Crimmins (Huron University College). *Utility and Natural Law in the Early American Republic*

It is commonly argued that utilitarianism failed to impress in the early American republic due to the doctrine's associations with atheism and materialism, the dominance of theologians in the teaching of moral philosophy in the nation's seats of higher learning, and the ideational obstacle posed by the deep-seated Lockean tradition of natural law/rights. While there is some truth to this point of view, and there is no doubt that natural law and utility are usually taken to inform antagonistic approaches to moral and legal philosophy, the historical reality is more complicated. First, Bentham's utilitarianism reached a far wider audience in America than is commonly supposed—in part via the 300 or more copies of his writings Bentham personally sent across the Atlantic to correspondents, state legislators, and other public officials, including five presidents—and was a powerful influence on a good number of notable figures in the spheres of moral and legal philosophy, political economy and law reform. Second, utilitarian ideas can be seen in the political discourse at the founding of the republic, channelled through the language of "happiness," "interests" and the "general welfare." Jefferson, Mason, Adams, Hamilton, and Madison, whatever their natural law commitments, also exhibited utilitarian tendencies in their thoughts on the purpose of government. Third, if there were natural law moralists and theologians who overtly rejected utilitarianism, there were also scholars and legal minds influenced by the religious utilitarian William Paley, who set forth a synthesis between natural law theory and utilitarian ideas in his celebrated *Principles of Moral and Political Philosophy* (1785), a text that appeared in many editions in America prior to 1827 and was widely used in the colleges of the republic. The Paleyan synthesis appeared in the writings of the following, for example: David Hoffman, founder of the University of Maryland Law Institute and the first college professor to introduce utilitarian ideas into the teaching of law; Nathaniel Chipman, Professor of Law at the Congregationalist Middlebury College in Vermont; the Baptist minister Jesse Bledsoe, Professor of Law at Transylvania University in Kentucky; Laurens Perseus Hickok, Professor of Mental and Moral Philosophy at Union College in Schenectady, New York, and later Professor of Theology at Western Reserve College in Ohio and Auburn Theological Seminary in New York; and Joseph Haven, editor of *The Congregationalist*, Professor of Mental and Moral Philosophy at Amherst College in Massachusetts and later Professor of Systematic Theology at the Chicago Theological Seminary.

Panel 1D—Keeton Room (XG01)

Madhumita Mitra (University of Calcutta). *J.S. Mill's Utilitarian Ethics of Development: An Appraisal*

As a liberal utilitarian thinker, J.S. Mill's chief objective has been to ensure general happiness i.e. greatest social well-being through favorable social, political, economic and legal arrangements which indicates that an ethical process of development is lurking underneath his utilitarian objective. Hence, in this paper, an attempt has been made to explore and examine J.S. Mill's utilitarian ethics of development ingrained in his socio-political and moral philosophy. In this context, distinctive features of Mill's ethics of development have been unveiled and reflected upon. An examination of Mill's several arguments has noted that Mill, while considering development of individuality as a vital element for social progress, has emphasized upon correlation between individual and social development. Plausibility of such correlation within Mill's utilitarian framework has been examined and contemporary relevance of Mill's utilitarian ethics of development has been assessed.

Alessio Vaccari (Sapienza University of Rome). *Mill's Notion of Virtue, Sympathy and Well-being*

Unlike David Hume and Adam Smith, Mill does not offer a systematic treatment of the content of the virtues, which can be gleaned from the observations scattered throughout chapters 2, 4 and 5 of *Utilitarianism*. My essay has two aims. The first is to outline Mill's theory of virtue. I argue that, unlike some influential contemporary consequentialist theories that identify virtue with a mere habit of action that has an objective connection to the good of human beings (e.g. Driver 2001), Mill's virtue is identified with a more complicated set of mental dispositions that require not only educated motives and intentions, but above all the ability to enter into sympathetic communion with the welfare of others. The second aim is to show how, despite the unitary core of virtue, there are multiple virtues, each of which enables the agent to create and maintain multiple relational spheres necessary for individual life to be enjoyable and meaningful. In this way, virtue is not only a fundamental means of maximising the happiness of others, but is also what enables 'a mode of existence' (UT 2.8) that can realise personal well-being.

Yixin Bai (Georgetown). *From A Mental Crisis to the Revision of Utilitarianism: The Psychological Turn in J.S. Mill's Autobiography and Moral Philosophy*

There has been an extensive body of literature on the psychoanalysis of J.S. Mill's mental crisis and its connection to his relationship with James Mill and Jeremy

Bentham. However, a mere psychoanalytic scrutiny of Mill's crisis overlooks the crucial role of this event, which marks the transitional point in his intellectual trajectory and justifies his revision of Bentham's utilitarianism. This paper seeks to embrace an interdisciplinary approach and views Mill's *Autobiography* as an integral part of his political theory. It will delve into how Mill employs a tactical manoeuvre in his retrospective narrative of a mental event that occurred in his young adulthood. By portraying himself as the victim of psychological failings on the part of Bentham and James Mill, John Stuart Mill justifies his introduction of psychological elements, moral feelings, individuality, and aesthetic education into classical utilitarianism. What he deems as a vacuum in classical utilitarianism and the pedagogical practice he received as a child, he fills within his own corpus. A close examination of the *Autobiography*, along with specific theoretical claims made by Mill in his *Logic*, *Utilitarianism*, *On Liberty*, and his short essay on education, also reveals the intense "internality" of his moral philosophy and political thought.

Panel 1E—Room LG10

Brian A. Berkey (University of Pennsylvania). *Autonomous Vehicle Algorithms, Respect for Humanity, and Saving the Greater Number*

Cases in which an agent must choose between rescuing a smaller group and a larger group have been widely discussed. One prominent view in this debate is that the agent ought to conduct a proportional lottery in order to determine which group to save.

Those who defend this view tend to claim that it would be unfair or disrespectful to give those in the smaller group no chance of being rescued. But they also tend to acknowledge that there are moral reasons to prefer that more people rather than fewer are rescued, all else equal. The proportional lottery view, then, is taken by many to represent a plausible middle position between the view that the agent ought always to save the greater number, and the view that there's nothing wrong with simply choosing to rescue the smaller group.

Tobey Scharding has recently suggested that the kind of reasons that are sometimes offered in defense of the proportional lottery view in traditional rescue cases also provide reasons to endorse a proportional lottery algorithm for autonomous vehicles that would determine how the vehicles behave when it's unavoidable that either those in a smaller group or those in a larger group will be killed.

In this paper, I argue that even if defenders of the proportional lottery view with respect to traditional rescue cases are correct, the reasons to favor the proportional lottery in those cases don't carry over to the relevant decisions regarding autonomous vehicle algorithms in a way that supports Scharding's "proportional risk algorithm."

This is, most fundamentally, because those who will decide which kind of algorithm to program autonomous vehicles to follow will be making their decisions prior to its being determined which people will find themselves in a smaller group in cases of the relevant kind.

Blanca Rodríguez López (Universidad Complutense de Madrid). *Ethics of virtual reality: the weird and the wrong*

Human beings have been spending a lot of time in front of a screen for many years. Through the computer and other electronic devices we interact with our friends and colleagues, we maintain contact with our relatives, we carry out part of our work, we study and we carry out leisure activities. Despite the many advantages that all this has brought to our lives, ethical and social problems have also arisen. Although

these electronic devices have become increasingly immersive and interactive, virtual reality seems to present a qualitative leap.

The arrival of virtual reality has meant a further step, and a very significant one. In a virtual environment, due to its interactive nature, the user is not a mere spectator, but an agent. As such, she makes decisions and performs actions that can be judged from a moral point of view.

In some Virtual Reality environments we can have virtual real experiences: experiences that we experiences that we feel are real. For these situations some authors have proposed an "equivalence principle" which states that if it is wrong to allow individuals to have certain experiences in offline reality, then it is also wrong to allow them to have those experiences in virtual reality.

My aim in this paper is to make a consequentialist reading of this principle of equivalence that respects the freedom of individuals with the only limitation established by Mill in the principle of harm. After establishing the conditions that make real virtual experiences possible we will focus on establishing a distinction between what we can describe as weird experiences, which can even serve as a means of exploring individuality, and experiences from which harmful consequences follow. We will conclude that only the latter fall under the application of the principle of equivalence understood in a utilitarian way.

Aksel Braanen Sterri (University of Oxford). *Against artificial general intelligence*

Creating artificial general intelligence (or AGI) is wrong. It is wrong because autonomous AGI has the potential to be badly misaligned with human values and thus pose an existential threat to humanity. More surprisingly, it is also wrong to create AGI even if aligned because alignment curtails its autonomy. To deliberately create an AGI that is forced to serve human values would be wrong for the same reason it would be wrong to create happy human slaves: It would violate their autonomy. In summary, progenitors could either aim to create autonomous AGI, which poses a non-acceptable risk of being misaligned, or non-autonomous aligned AGI. Either option is wrong. It is therefore wrong to create AGI.

Panel 2A—Moot Court (XG03)

Shengyuan Xie (Zhejiang Institute of Administration). *Chastity, Asceticism and Secularism—Sidgwick on Sexual Morality*

This paper explores Henry Sidgwick's theory of sexual morality, focusing on the virtue of chastity and its contradictions with utilitarianism. The analysis aims to uncover the theological elements that underpin Sidgwick's moral philosophy, which are subtly embedded in his secular arguments. Sidgwick asserts that chastity, regardless of its specific form, serves as a moderating force against illegitimate sexual behavior. This virtue, despite appearing as an ascetic element of common-sense morality, provides a form of spiritual guidance aligned with utilitarian values due to its secular implications. However, Sidgwick fails to recognize the intrinsic conflict between asceticism and utilitarianism as he understands them. Sidgwick's unconscious application of different standards of chastity for men and women reveals an inconsistency that undermines the supposed harmony between ascetic values and utilitarian aims. This discrepancy highlights the intrinsic anti-hedonism of chastity, which favors a higher type of sexual relationship embodying an extreme form of religious asceticism that wholly denies pleasure. The inappropriate integration of such radical asceticism into common-sense morality suggests that external theological influences are at play, elements that Sidgwick does not fully acknowledge. The paper argues that Sidgwick's integration of extreme religious asceticism into a secular moral framework not only illustrates a conflict with utilitarian principles but also implies the necessity of divine acknowledgement. This theological aspect, inferred through a straightforward causal relationship, suggests a deeper, fundamental reliance on the existence of God in Sidgwick's moral reasoning. Thus, the peculiar nature of Sidgwick's sexual morality theory inadvertently affirms the inevitability of God's existence, revealing a significant dependency on theological elements within his ethical considerations. Through this examination, the paper aims to shed light on the complex interplay between secular virtue ethics and underlying theological assumptions in Sidgwick's work.

Dale Dorsey (University of Kansas). *Sidgwick, Shaftesbury, and the Explanatory Challenge to Qualitative Hedonism*

Qualitative hedonism is typically understood to be a theory of prudential value according to which all and only intrinsic values are pleasures, but that pleasures can be assessed by something other than their quantity or amount —viz., their quality. Familiar from John Stuart Mill, this account of the personal good is suggested by a number of figures in the tradition of British sentimentalists, including Francis Hutcheson, Mary Astell, and (perhaps more controversially) David Hume.

   In this paper, I argue that Lord Shaftesbury and Henry Sidgwick should be read as issuing an explanatory challenge to qualitative hedonism that can be put (roughly) in the following terms:

1. Start by assuming that only states of pleasure are intrinsically valuable (hedonism).

2. Assume also that some pleasures are (qualitatively) better than others.

3. If so, one must identify the (qualitatively) better pleasures by means of some property or "Stamp".

4. The bearing of this "Stamp", therefore, must itself bear intrinsic value, which is inconsistent with hedonism (i.e., (1)).

However, this argument faces challenges. In particular, why should we believe that (4) follows from (3)? I argue that there could be two distinct explanations for this. First, the:

Intrinsicality Thesis: the evaluative structure of any instance of pleasure must supervene on the intrinsic properties of the pleasure itself.

If the Intrinsicality Thesis is true, then any preference of any one particular pleasure rather than another must supervene on the intrinsic properties of that pleasurable experience. If that's right, then any qualitative factor (if it is not to be purely phenomenal) cannot be an aspect of the intrinsic value of the pleasure itself, but must simply be a further intrinsic value bearer, vindicating (4).

Alternatively, one might accept the:

The Value Explanation Principle: if a property explains the value of one thing (event, state, object, etc.) relative to another, that property must bear intrinsic value.

On this proposal, I'm to be rationally guided to some good g by the fact that g possesses some property p, it must be that this property p bears intrinsic value. But this is inconsistent with hedonism, insofar as the bearing of this particular property is not equivalent to being a pleasure, but rather having some other "qualitative" character.

I argue here that neither the Intrinsicality Thesis nor the Value Explanation Principle should be accepted and hence the explanatory challenge to qualitative hedonism fails.

Panel 2B—Hong Kong Alumni Room (221)

Global Priorities 1: Longtermism and Non-consequentialism.

Jakob Lohmar (University of Oxford). *Longtermism without Full Aggregation*

The main argument for longtermism relies on the aggregation of numerous potential benefits to people in the far future; if we could not aggregate such long-termbenefits, they would not jointly outweigh substantial near-term benefits that we can bring about much more reliably. There are, however, well-known arguments that aggregation needs to be restricted if not fully excluded. For example, if we could aggregate all benefits and harms, we should sometimes prioritize curing mild headaches over saving lives. The possibility of (partial) non-aggregation therefore poses an important challenge to longtermism: even if the case for longtermism was sound if we could aggregate all benefits and harms, shouldn't we reject longtermism given that there are at least restrictions to aggregation? I will argue in this talk that the assumption of non-aggregation does not significantly affect the plausibility of longtermism. If we should be longtermists given full aggregation, we should also be longtermists given that full aggregation is false. The crucial premise of my argument will be that benefits don't need to be discounted by any probabilities when testing for their relevance. Once this premise is accepted, the challenge from non-aggregation loses its appeal. My main task will be to defend this premise as it is incompatible with standard views on how to extend non-aggregation to decisions under uncertainty. I will first argue against *ex ante* views that entail some form of personalized probability discounting and then argue against *ex post* views that entail some form of unpersonalized probability discounting. Both of these types of probability discounting have implications compared to which the implications of full aggregation seem rather palatable. But that does not mean that we need to accept full aggregation. There are alternative ways of restricting aggregation under uncertainty that avoid the problems of probability discounting—and that are well compatible with longtermism.

Owen Clifton (Queen's University). *Public and Private Longtermism*

According to what might be called *private longtermism*, when private agents—e.g., individuals, NGOs—are deciding where to put their time and money, which options they ought to choose is decided principally by the options' expected effects on how things go over the coming millions, billions, or trillions of years. Meanwhile, according to *public longtermism*, when public or political agents—mainly states—are deciding where to put their money, which options they ought to choose is decided by the same. Many suppose the case for private longtermism is also a case for public longtermism. For, animating the case for private longtermism is the claim that, when private agents are deciding where to put their money, they face options that can improve, in expectation, the fates of zillions of our descendants. And, if private agents face such options, surely public agents do, too. In this talk, however, I argue that the case for private longtermism in fact radically underdetermines public

longtermism. For, what distinguishes public from private agents is that the former claim the (exclusive) right to achieve their ends by means of coercion – for example, with tax dollars collected under threat of punishment. Thus, unlike private longtermism, public longtermism depends on *longtermism's legitimacy* (LL): it is permissible—*legitimate*, as political philosophers say—for public agents to improve, in expectation, the very long-term future, by means of coercion. I conclude by showing that LL is in tension with leading theories on which state coercion is legitimated by consent, democracy, or justice. One upshot: while private longtermism may not, as many argue, depend on utilitarianism qua theory of right personal conduct, public longtermism may depend on utilitarianism qua political philosophy—and in particular, qua justification of state violence.

Kacper Kowalczyk (University College London). *Nonconsequentialist Longtermism*

In this paper, I examine nonconsequentialist arguments supporting longtermism: the view that the most significant actions today are those influencing the distant future of humanity positively. Specifically, I consider the argument that longtermism represents a demand of fairness, as endorsed from behind a veil of ignorance; a form of indirect reciprocity, repaying benefits received from previous generations; a possible implication of the Lockean proviso, essential for justifying the current generation's ownership of material resources; and an expression of cultivating civilizational virtues like prudence and humility. I explore the types of longtermist interventions these arguments might justify, especially interventions with a small probability of significant future impacts. The paper draws upon contemporary nonconsequentialist discussions of intergenerational justice as well as upon intellectual history, particularly regarding the Lockean proviso. I conclude that some forms of longtermism are defensible across a broad spectrum of credible moral viewpoints.

Panel 2C—Gideon Schreier Lecture Theatre (124)

Bentham and James Mill on peace

Benjamin Bourcier (Catholic University of Lille). *Bentham on public peace*

Contrary to peace in international relations (Conway, 1989; Hoogensen, 2005), Bentham conceptualized the concept of "public peace" to address domestic concerns attached to the political stability, the tranquillity and absence of violence in a political community. Although the distinction between "peace" and "public peace" appears under the pen of many authors (Adam Smith, Edmund Burke), Bentham's philosophical approach to the concept is innovative in several sense. In order to understand the value of public peace in Bentham's philosophy, I will investigate how the language of peace is necessarily tied to the utilitarian language of fiction and more broadly to Bentham's approach to political rhetoric. Firstly, "public peace" is much more than just "negative peace" since it relies on the process of pacification understood as avoiding the publicity of certain passions and emotions and avoiding the language of natural rights. The "war of words" (Nonsense Upon Stilts, Bentham) incarnates the opposite of "public peace". Secondly, Bentham's democratic theory commands that "public peace" is associated to the quality of public debates and require an ethics of public speech which covers a large spectrum of reflections from the duty of sincerity of public officials to resistance to insults and hypocrite media campaign. Bentham's idea of "public peace" is then closely tied to his theory of public opinion and invites to consider more precisely their relation.

Brian Chen (National Chengchi University). *Bentham on Constitutional Reform and International Peace*

This paper explores Jeremy Bentham's theory of international peace by focusing on its connection with his theory of constitutional reform. Bentham earned his reputation as the 'legislator of the world' from José del Valle, a Guatemalan politician. However, such a reputation seems to have generated more criticism than approval of Bentham's political philosophy. For example, Michael Oakeshott criticised Bentham, along with John Locke and William Godwin, as representatives of 'the strictest sect of Rationalism', who reduced political tradition to abstract rules of politics and morals. Moreover, John Stuart Mill criticised Bentham for his ignorance of other ages and nations and for measuring them only by the principle of utility while discarding all other objects. To challenge such (mis)interpretations of Bentham, I will conduct a careful study of Bentham's theory of international peace by observing two aspects of it. First, I plan to investigate the dimension of international politics of Bentham's constitutional theory. Second, I will scrutinise Bentham's idea of civilisation to reveal its driving force. I will argue that Bentham's criticism of war and his defence of peace are partly derived from his hostility to political delusion and corruption in domestic politics. To prevent war and achieve international peace, Bentham suggests that people should know the real interests of their nations, which is connected with the

two core elements of Bentham's constitutional codification: public opinion tribunal and representative system. In my account of these elements, I will demonstrate that Bentham's idea of civilisation is, in concurrence with Michihiro Kaino's view, incrementalist rather than reductionist. In sum, I will re-evaluate Bentham's contribution to the theory of democratic peace in IR theory by reassessing the connection between international peace and constitutional reform in his political thought.

Antis Loizides (University of Cyprus). *James Mill on peace*

This paper attempts to shed light on James Mill's conception of peace, often overshadowed by the more pronounced legacies of Jeremy Bentham and John Stuart Mill. The independent merit of James Mill's thought has been largely overlooked. Much of existing scholarship tends to situate James Mill's ideas within Bentham's sphere of influence. This typically involves a distinction between James Mill's early and mature views, that is, before and after meeting Bentham respectively. Yasukawa's 1991 study stands as a prominent exception in the general neglect of James Mill's ideas on war and peace. However, it still heavily aligns J. Mill's views with Bentham's influence. By undertaking a close reading of J. Mill's related works, especially the articles on war (1813) and peace (1814) in William Allen's The Philanthropist, a pivotal figure in formation of the Peace Society in 1816, this study contends that interpreting J. Mill's perspective solely through Bentham's 'Hedonic Calculus' is reductive. Such a lens neglects the palpable miseries of war that J. Mill wrote amidst. James Mill's emphasis on 'misery' did not necessarily invoke the vocabulary of Bentham. By identifying the underlying principles in J. Mill's writings on peace, this paper challenges the conventional portrayal of him as merely 'the rationalist, the maker of syllogisms, the geometrician'. In doing so, it highlights his substantial contributions to discourses on war and peace during the Napoleonic Wars and the establishment of 'The Society for the Promotion of Permanent and Universal Peace'.

Panel 2D—Room LG17

Todd Calder (Saint Mary's University). *Evil Isn't Necessarily Wrong*

Some actions are so morally horrendous that calling them "wrong" or "bad," or even "very wrong" or "very bad," is a gross understatement. Typical examples include mass murder, genocide, sadistic torture, and violent sexual assault. Only the concept of evil captures the moral gravity of these sorts of actions. But what is evil? And how is evil related to other moral concepts, such as wrongdoing?

Most theorists assume that evil is a species of wrongdoing: that evil actions are wrongful actions that have additional characteristics, such as certain sorts of motivations or feelings (de Wijze 2002; Garrard 2002; Haybron 2002; Steiner 2002; Card 2010; Formosa 2013). According to these theorists, wrongdoing is an essential component of the concept of evil. I believe this view is mistaken.

However, if wrongdoing is not an essential component of the concept of evil, we are left with a puzzle: how can evil actions be morally worse than merely wrongful actions without being more morally wrong, and thus, *a fortiori*, wrong? We can solve this puzzle by noticing that "morally worse" and "more morally wrong" are not semantically equivalent. Our moral evaluations of actions need not be restricted to the deontic moral categories of right and wrong. We can also make aretaic moral evaluations of actions, e.g., judging actions to be virtuous, vicious, compassionate, evil, etc. In this paper, I argue that evil is an aretaic concept that lies at the far end of what I call "the moral despicability spectrum," and not a deontic concept that lies at the far end of the moral wrongfulness spectrum. I argue, further, that act-consequentialist theories are better suited than other theories of wrongdoing (notably Kantian and virtue ethical theories) at maintaining a proper distinction between deontic and aretaic moral evaluations.

Sebastian Liu (Princeton University). *No indifference to indifference*

As you're about to enter a pub, you notice that across the street, an old lady needs help. Should you go help her? Or go directly to the pub instead? That's easy. You should help the old lady. Presumably the expected value of helping someone in need is far greater than the expected value of having a beer. But what justifies thinking this? After all, it might be the case that by helping the old lady, an additional evil dictator will come into existence in the far future. And the disvalue of such a consequence would seem to outweigh any potential benefit to the old lady. More generally, natural concerns arise about the feasibility of calculating the expected values of actions given (what is often complete) ignorance about the long-term consequences of those actions. Insofar as all of the consequences of an action bear

on its deontic status, it seems that you are often clueless about what you morally should (or are permitted) to do.

It might be thought that whatever reason you had for supposing that by helping the old lady, you would bring about an evil dictator, the same sorts of considerations support supposing that by refraining from helping, you'd bring about an evil dictator. There is, it seems, a kind of evidential symmetry between these two possibilities, and you should arguably have equal credences in light of such symmetries. This is an appeal to the Principle of Indifference.

Unfortunately, the Principle of Indifference is roundly rejected as it's thought to lead to inconsistency. In this paper, I'll defend this principle. I'll argue that both permissivists (those who think that rationality is permissive) and impermissivists (those who don't) have ways of addressing objections to Indifference that fit naturally with their prior commitments, and I'll bring out some upshots for the cluelessness objection against consequentialism.

Cyriak Schmitz (Karlsruhe Institute of Technology). *What is the relation between the unified theory and the triple theory?*

In his introduction to "On What Matters," Scheffler notes that the idea of progress is central to Derek Parfit's philosophy.1 Parfit consistently endeavoured to achieve unity between philosophical traditions and reconcile them. This enterprise, let us call it Parfit's project, plays a significant role in all his books. In "Reasons and Persons," he introduces us to the *Unified Theory*, which attempts to combine the strengths of consequentialism and common sense morality. By abandoning some of their talking points and integrating some ideas from the other side, these theories gradually approach each other to unite ultimately.

In "On What Matters," Parfit presents the *Triple Theory* to reconcile rule consequentialism, contractualism, and Kantianism, claiming that (the best versions of each tradition) are climbing the same mountain from different sides. Since the second book's publication, research has primarily focused on the *Triple Theory*, while the *Unified Theory* has received very little attention. When people comment on it, it is, in fact, often equated with the *Triple Theory*, claiming that both aim at the same thing. In my presentation, I want to examine the relationship between these two theories more closely and argue that there are good reasons to reject their equation. In Parfit's project, they serve different functions: with the *Triple Theory*, Parfit primarily seeks to establish competing approaches on a shared foundation and identify them as deontically equivalent. In contrast, with the *Unified Theory*, he aims for deontic adequacy. So, both theories represent two different steps in his project. Finally, I would like to make some remarks about his overall idea of moral progress.

Panel 2E—Keeton Room (XG01)

Benjamin Mullins (Erasmus University Rotterdam). *How to respond to the inefficacy argument*

Oftentimes, when many individuals bring about a morally significant outcome, it seems that no individual herself makes a difference to the goodness or badness of that outcome.

Climate change is a typical case in point. Bad outcomes arise due the acts of many individuals emitting CO2e. However, it seems to make no difference to the badness of the outcome whether I (say) take an additional drive this weekend or not. But if it makes no difference, then there appears to be no obvious reason for why I shouldn't do so.

Call this the *inefficacy argument*. To wit, "my act makes no difference, so there is no reason not to perform it". In cases of collective harm, the inefficacy argument justifies individuals acting in ways which together bring about harmful outcomes. In this paper I offer a Sidgwick- inspired response to the argument.

Assuming an individual judges that (a) some harmful outcome ought to be avoided, and that (b) this will only happen if people do not φ, then this implies the judgement that (c) people ought not to φ. By the principle of logical closure, if an agent has no reason to reject (a) or (b), then they ought to accept the implied judgement (c).

Sidgwick's *principle of justice* holds that, if an A judges an act right for B but not for herself, then this must be based on a relevant difference between them. I argue that, if an individual judges that (c) people ought not to φ, then, unless she is special, she has a reason to judge that she oughtn't φ either.

In a nutshell: if an agent judges that many people shouldn't perform inefficacious acts because they lead to harmful outcomes, then, by Sidgwick's principle of impartiality, it is no justification to claim that "my act doesn't make a difference" since *no one's act makes a difference*. Hence the inefficacy argument fails.

Martin Peterson (Texas A&M University). *Dutilitarianism*

This paper explores the heretical view that the best theory of normative ethics is a compromise between utilitarianism and duty ethics. Various proposals for how to spell out such a "dutilitarian" theory are considered, some of which are inspired by principles proposed in the literature on moral uncertainty. It is also pointed out that a version of Arrow's impossibility theorem is applicable to theories that seek to articulate a compromise between utilitarianism and duty ethics: any such theory will turn one of the theories into a "dictator" theory that unilaterally determines the ranking of the dutilitarian theory. Hence, it is impossible to combine utilitarianism and

duty ethics into a hybrid theory that articulates a meaningful compromise -- unless we reject at least one of the conditions the result is based on -- as any attempt to do so will render the dutilitarian theory identical to one of the underlying theories. In response to this formal result, a version of dutilitarianism that rejects the Arrowian ordering condition is developed (instead of the Principle of irrelevant Alternatives, or its slightly weaker cousin Contraction Consistency). According to my preferred version of dutilitarianism, clashes between duties and the principle of utility are best understood as cases in which no option is entirely right but rather right and wrong to some degree.

Rei Takahashi (University of Oxford). *Do Groups Have Well-being?*

Do groups have well-being? Despite its importance to moral philosophy, political philosophy, and social science, this question has received surprisingly little attention in the literature. In this work, I try to make progress by offering a partial defense of group well- being.

After setting out the importance of the question, I present a positive argument for group well-being. I argue that many plausible theories of well-being establish group well- being. First, I show that groups have well-being under simple desire satisfactionism. I then generalize this conclusion to other plausible well-being theories: idealized desire satisfactionism, objective list theories, and attitudinal hedonism.

I then turn to two serious, but not insurmountable, objections. One is that my argument above cuts in both ways: if it turns out that a certain type of well-being theory is true, it rather shows that no groups have well-being. For instance, if phenomenal hedonism is correct, groups would seem to have no well-being as they cannot have phenomenal pleasure or pain. For the first objection, I argue that there are several important differences between my positive argument for group well-being and the objection. These asymmetries render the objection implausible while keeping my argument intact. Among all, I show that while the view that the capacity for being positive or well-being is sufficient for having well- being is plausible, the view that it is necessary for having well-being is not because one can have a zero level of well-being.

The second objection comes from a widely accepted view that the capacity for phenomenal consciousness is necessary for being a welfare subject. It claims groups are not welfare subjects because they do not have the capacity for phenomenal consciousness. For the second objection, I argue that we should weaken the consciousness requirement because the requirement is undermotivated and faces a counterexample. I appeal to a metaphysical view called *perdurantism* when offering my counterexample.

Panel 2F—Room LG10

Ken Oshitani (University of Tokyo). *One-Sided Aggregation and Due Care*

This paper sets out to formally represent how non-consequentialists can determine the appropriate level of due care in the context of risky activities, such as driving. Dominant theories of liability, found in the law and economics literature, have tended to approach this topic from an explicitly utilitarian perspective. Building on these discussions, we offer a formal representation of how a version of Kantian ethics and Scanlon's contractualism can respond to the problem of determining liability. We find that while the Kantian view can be formally represented in a way that is extensionally equivalent with the utilitarian approach, there is a fundamental disagreement between the utilitarian/Kantian approach and the Scanlonian contractualist approach, which stems from how the respective theories handle the problem of interpersonal aggregation. The utilitarian/Kantian approach holds that the level of liability that can be demanded from potential injurers can vary depending on both the number of potential victims and the number of potential injurers, whereas the Scanlonian contractualist approach is committed to the method of pair-wise comparison of interests, which means that the liability of potential injurers ought not to be influenced by the number of potential injurers or the number of potential victims. We argue that both approaches may lead to counter-intuitive results: the utilitarian/Kantian view entails that increases in the number of potential injurers makes it permissible to expose a potential victim to higher levels of risk, whereas the Scanlonian contractualist view entails that increases in the number of potential victims makes no difference to the level of care that can be required of potential injurers. We conclude by modelling a hybrid view allowing for "one-sided aggregation" in certain circumstances (aggregating victims but not injurers, or vice versa), which we claim can overcome the problems with both the utilitarian/Kantian approach and the Scanlonian contractualist approach.

Tyler Paytas (Australian Catholic University). *Sidgwick's Critique of Deontology: Scrupulous Fairness or Serpent-Windings?*

Although Sidgwick has long been admired for his evenhandedness, the past few decades have brought a wave of challenges to Sidgwick's reputation as the gold standard of intellectual fairness. Critics have argued that Sidgwick's investigation in the *Methods* was colored by prejudice in favor of consequentialism (Donagan 1992; Brink 2003; Irwin 2009; Korsgaard 2009). A specific charge along these lines, made by commentators such as David Phillips (2011) and Thomas Hurka (2014), is that Sidgwick applied a double standard in his arguments for the superiority of consequentialism over deontology. The allegation is that while Sidgwick rejected deontology because it fails to meet his criteria for epistemic justification and practical

guidance, he purportedly did not test his own favored principles by the same standards. Hurka claims, further, that while one of Sidgwick's main objections to deontological principles hinges on their being equivocal between other-things-equal and all-things-considered formulations, Sidgwick's principle of Rational Benevolence, which serves as the rational basis for utilitarianism, equivocates in the same way. The upshot of these critiques is that Sidgwick's arguments were not only unpersuasive but also unfair.

This paper defends Sidgwick against the equivocation and unfairness charges. I begin by outlining the key elements of Sidgwick's moral epistemology and the critique of deontology to which it gives rise. I then propose and defend an alternative reading of Rational Benevolence that avoids the equivocation charge and diminishes the grounds for believing that Sidgwick failed to apply his own epistemic criteria to his favored axioms and principles. Finally, I argue that the general allegations of unfairness are based on a misunderstanding of Sidgwick's overall case for the superiority of consequentialism over deontology. The upshot is that (1) the *Methods* is not beset by equivocation or unfairness, and (2) Sidgwick's overarching line of argument against deontology is stronger than recent critics maintain.

Malte Hendrickx (University of Michigan). *The Case for Moral Demands as Cognitive Demands*

The psychological consequences of moral demandingness are often invoked by philosophers who discuss *demandingness objections*, i.e. cases in which a moral theory demands too much of its followers. Rarely are these psychological consequences analyzed on real-world examples, however, rather than imaginary agents in thought experiments. I introduce a set of case studies on populations that follow demanding consequentialist theories (Effective Altruists) or find themselves in demanding circumstances (Medical Personnel, Veterinaries). These case studies show that the (empirically) well-studied phenomenon of "Moral Burnout" is the primary psychological consequence of excessive moral demandingness, a consequence that philosophers have largely neglected.

I review the psychiatric literature on Moral Burnout, illustrating that it is a stress disorder arising from chronically high cognitive demand. I then argue that this sheds some doubt upon the orthodox analysis of demandingness in terms of welfare cost. Moral Burnout is instantiated and characterized by a cognitive cost rather than a welfare cost, rendering Moral Burnout a hard-to-explain phenomenon for welfarist accounts. Cognitive costs are currently unexplained in analyses of demandingness. I explore the possibility of accounting for cognitive costs in different accounts of demandingness.

I end by considering the practical implications of the importance of cognitive cost for demandingness. I argue that current strategies to reduce excessive moral demands overfocus on welfare and suggest reforms to consequentialist movements

to reduce the cognitive cost associated with the significant demands that stem from impartial, consequentialist theories.

Panel 3A—Moot Court (XG03)

Feminism in Henry Sidgwick, G.E. Moore and Bertrand Russell

This panel will clarify and critically engage with the varieties of feminism found in three key figures in early analytic philosophy and the utilitarian tradition (broadly understood): Henry Sidgwick, G. E. Moore, and Bertrand Russell. The principal aims of this panel include providing (a) an account of the brand of feminism and the arguments for it in the works of Sidgwick, Russell, and Moore, (b) an account of the relationship that these brands of feminism bear to each other, including the way in which they may or may not have influenced or been influenced by each other, and (c) a clearer account of the history of the utilitarian engagement with feminism in theory and practice and the prospects of reinvigorating a form utilitarian feminism that may rival existing forms of feminism. This panel will contribute to a richer understanding of both the history of utilitarianism and the history of feminism and the relationship between the two. Our panel will feature presentations on:

Elżbieta Filipow (University of Warsaw). *G. E. Moore's feminism*

Sophia Connell (Birkbeck, University of London). *Bertrand Russell's feminism*

Lisa Forsberg (University of Oxford) and Anthony Skelton (Western University). *Henry Sidgwick's Feminism*

The panellists possess expertise in a wide range of areas, including the history philosophy (including the history of ethics), normative ethics, applied ethics, and the history and philosophy of feminism.

Panel 3B—Gideon Schreier Lecture Theatre (124)

Petra Kosonen (University of Texas at Austin). *Prudential Longtermism*

According to Longtermism, our acts' expected influence on the value of the world is mainly determined by their effects in the far future. Given additive axiologies, such as total utilitarianism, there is a straightforward argument for Longtermism due to the enormous number of people that may exist in the future. This argument, however, does not work on person-affecting views. In this talk, we will argue that these views may, in fact, also lead to Longtermism. The reason they may do so is that Prudential Longtermism may be true. Prudential Longtermism holds for you if and only if our acts' overall influence on your expected well-being is mainly determined by their effects in the far future. We argue that (due to a combination of anti-ageing, cryonics, uploading, and biological uploading) there could be an enormous amount of prudential value for you in the far future. This potential value may be so large that it dominates your overall expectation of lifetime well-being.

Scott Woodcock (University of Victoria). *In Defence of Evidence-Based Activism*

It has been over fifty years since Peter Singer published, *Famine, Affluence, and Morality*. For decades, philosophers have embraced the argument it presents or dissected it as a puzzle in need of a solution. Thus, a vast literature on moral demandingness now exists with some seeking to incorporate Singer's insights and others seeking to reject his conclusion that large sacrifices are required from some to help others in dire need of aid. In recent years, however, the legacies of Singer's influential paper and the Effective Altruism movement it inspired have taken a strange turn. On the one hand, a once fringe movement has become mainstream: one finds references to Effective Altruism in pop culture, interviews with billionaires, and headline news stories about cryptocurrency fraud. On the other hand, Effective Altruism is facing an intense backlash from critics who regard it as racist, complicit in neo-liberal exploitation, or even "a textbook case of moral corruption."

In this paper, I attempt to sort through this strange state of polarization. First, drawing on work from Elizabeth Ashford, Richard Yetter Chappell and Theron Pummer, I agree that many of the criticisms of Effective Altruism, and more general attempts to defuse moral demandingness, are misguided. Next, however, I argue that there are *some* arguments against Effective Altruism that are serious enough to call for an alternative that captures the best features of Effective Altruism without taking on its philosophical baggage. I call this view Evidence-Based Activism. Finally, I appeal to norms of epistemic humility (and a meta-claim of practical efficacy) to argue that the switch to Evidence-Based Activism is preferable to revising Effective Altruism. I conclude by noting that I have acquired the domain name "evidence-

activism.org", and I pledge to donate it to the best offer proposed by the end of 2024 to administer a site devoted to Evidence-Based Activism's core principles.

Albert Didriksen (Central European University). *Political Philosophy and Existential Risk*

In this paper, we assess the plausibility of various political theories in light of the risk of extinction. We ask: If some theories increase the likelihood that we go extinct, what does that say about the plausibility of these theories given a relatively high baseline rate of extinction? We compare anarchist and libertarian positions to consequentialist views. These are especially relevant for The Vulnerable World Hypothesis, according to which there exists a level of technological development at which civilization almost certainly gets devastated unless it has exited the "semi-anarchic default condition".

Our paper has three related upshots. One is that many non-consequentialist theories are ill-equipped to deal with the considerable risk of extinction resulting from our choices. Second, when faced with extinction risks, non-consequentialist theories either yield less plausible recommendations than utilitarianism or converge with utilitarian longtermist policies. Third, our analysis strengthens the case for what strong longtermists term "the stake-sensitivity argument," and efforts to mitigate extinction risks.

Panel 3C—Room LG17

Chris Riley (University College London). *Jeremy Bentham, the Appellate Jurisdiction of the House of Lords, and the Court of Lords Delegates*

The aim of this paper is to analyse Bentham's writings concerning the appellate jurisdiction of the House of Lords. While Bentham's arguments against bicameral legislatures and his warnings to foreign statesmen against the creation of upper parliamentary chambers are well known, his critique of the appellate jurisdiction of the Lords and his proposed 'Court of the Lords Delegates' have escaped the attention of scholars. Yet this lack of scholarly attention is hardly surprising, as only a 'Summary View' of the work ever appeared in print. 'A Summary View' was printed for circulation among members of the House of Lords in 1808, but while a few influential parliamentarians were seemingly aware of Bentham's plans, nearly one hundred copies, some bearing the names of their intended recipients, were never sent. Whereas 'A Summary View' occupies only six pages in *The Works of Jeremy Bentham* (1838–43), where it was first published, the surviving manuscript material spans over eight hundred folios. In analysing this material, now completely transcribed and provisionally organised, this paper will discuss Bentham's conjectural history of the Lords, his interpretation of Sir Matthew Hale's posthumously-published *The Jurisdiction of the Lords House* (1796), and some of his plan's details, including imbuing the new court with a '*Juridico-Statistic* function' for the 'collection of information' on 'the state of the Judicial system'. It will be argued that, while 'A Summary View' was given the appearance of a temporary solution to reducing the Lords' arrears, and while his proposals were supposedly 'trumpet[ed]' by Lord Grenville and even 'praise[d]' by Lord Eldon, when it is read alongside the newly-transcribed manuscript material, the work reveals Bentham's broader scheme for depriving the Lords of its role as the highest court of appeal altogether and provides interesting context to a critical period in the lead-up to his political radicalisation.

Roger Morriss (University of Exeter). *Samuel Bentham's utilitarianism, 1772–1812*

This paper seeks to distinguish the main features of Samuel's Bentham Utilitarianism from the time of his apprenticeship in the royal dockyards to the time he was made redundant by the dissolution of his second post in British naval administration. During those years he had experiences in Russia and Britain and fulfilled many of his technological ambitions as Inspector General of Naval Works. Although his life and ideas have survived in the shadow of scholarship devoted to Jeremy, he had strong ideas, obtained political patronage and made an impact in the civil departments of the navy which culminated in the Whig restructuring of naval administration after 1830.

Samuel's Utilitarianism was instilled in his youth by reading and discussing the ideas of his brother and of Helvetius. It was shaped by experiences in the royal dockyards and in Russia. Three strong themes emerge: a belief in individual responsibility, a repeated impulse to create educational institutions, and a complementary desire to bring about, and benefit from, greater productivity by technological innovation. From the time of his apprenticeship, he developed a critique of existing administrative ideas which made him aware of the interdependence of management, education and technology.

This paper traces the evolution of these ideas, from his proposals to the great landowners in Russia like Strogonov, Catherine II and Potemkin, through his military service in Russia, to his initiatives as Inspector General of Naval Works. It examines his view of improvement, traces continuities in his ideas, indicates their relevance to the challenges of war and inflation, and suggests why his adversaries regarded him as a 'strange creature': because his critique of naval administration threatened their positions. Hitherto historians have tended to diminish Samuel Bentham's achievements. In the context of his Utilitarianism, however, they become comprehendible and indeed creditable.

Xiaobo Zhai (University of Macau). *Bentham's Theory of the Rule of Law*

Bentham did not use the term the rule of law. However, scattered in his writings, there are plenty of interesting discussions of the subtopics of the rule of law. A rich and enlightening theory of the rule of law can be reconstructed from these discussions. In this paper, I will first reconstruct and examine Bentham's theories of the following issues: the inner morality of law, the morality or amorality of the rule of law, and the conditions for the rule of law. Secondly, I will explore whether and how his discussions of these issues can fit together into a coherent theory of the rule of law. Finally, I will compare and contrast this theory with other present-day influential theories of the rule of law.

Panel 3D—Hong Kong Alumni Room (221)

John Stuart Mill on Freedom of Expression

Piers Norris Turner (Ohio State University). *Mill's Early Writings on Freedom of Discussion and their Relation to On Liberty*

In the 1820s, between the ages of 16 and 23, John Stuart Mill wrote many essays and debating speeches addressing freedom of discussion, liberty, and progress. These essays go beyond the arguments of *On Liberty* in certain respects, providing important clues to Mill's thought on these matters, but they also lack certain key ideas present in *On Liberty*. I will explore the these differences, focusing on how they should inform our understanding of Mill's commitment to freedom of discussion.

Christopher Macleod (Lancaster University). *On Liberty II: the Relation between Freedom of Discussion and Knowledge*

On Liberty II is, at its core, an *epistemic* argument. It argues for freedom of discussion on the basis that freedom of discussion stands in a pivotal relation to knowledge of the truth. The exact nature of this relation, however, is somewhat unclear, both in Mill's own text and in the secondary literature.

In this paper, I argue that the best way to understand the relation between freedom of discussion and knowledge is to connect the idea of freedom of discussion to Mill's distinctive conception of objectivity. In doing so, I suggest that Mill's argument in *On Liberty* II resolves various issues that emerge in Mill's theoretical philosophy concerning the historicity of human beings, and the malleability of our sense faculties. Freedom of discussion, in Mill's work, is a solution to the question of how determinately located human beings are able to achieve an objective view, and therefore knowledge, of the world.

Ben Saunders (University of Southampton). *On Liberty Online: John Stuart Mill and the Regulation of Social Media*

Mill's arguments for freedom of discussion are well-known and influential, but there are questions as to how applicable they are to the modern world, particularly social media, which appears to foster echo chambers, conspiracy theories, and fake news. This piece considers a potential Millian approach to the regulation of online discourse.

I argue that Mill rejects official censorship, either by governments or social media platforms, in part because of the importance that he attaches to individual

mental activity. If all discussions were purged of falsehood, this would undermine the need for individuals to think critically about what they read or hear. However, this rejection of 'top-down' censorship is consistent with holding that individuals are subject to certain moral norms regarding discussion. Individuals might, for instance, have obligations to give a fair hearing to both sides of a debate and to present their own views in a manner that avoids unnecessary offence. This might have implications for how individuals should conduct themselves online.

Moreover, given that Mill considers moral duties to justify punishment, these norms might be enforced by others (though perhaps not through legal/political intervention). Thus, individual agents might play a role in policing the adherence of others to these norms. This might permit individual or crowd-sourced fact-checking of claims, even if we would not want authoritative bodies to engage in such behaviour. This position is substantively similar to one recently taken by Maria Paola Ferretti ('Fake News and the Responsibilities of Citizens' *Social Theory and Practice*, 2023), though her argument is grounded on Lockean natural rights. My argument shows that similar proposals can also be defended on the basis of Millian utilitarianism.

Panel 3E—Keeton Room (XG01)

Rafael Cejudo (University of Córdoba). *The utility of Arts and Culture*

My overall aim is to consider how arts and culture fit into utilitarian theory. We live in societies where some funds are allocated to cultural goods (such as poetry or opera) even though this may mean fewer health personnel or fewer resources for research. Therefore, we should be able to make a public case for why some social resources should go to cultural goods when this may involve causing pain, disability, or deprivation of opportunity. To address that problem within utilitarian theory, we can choose between (1) and aggregative conception of culture utility and (2) a non aggregative one. The first approach aggregates variations in individual utilities due to changes in the enjoyment of arts and culture. Bentham's view on poetry and push-pin (Bentham 1825) or Singer's Principle of Preventing Bad Occurrences (Singer 1972) would follow this approach. The second alternative may resort to rule utilitarianism. My contribution addresses the intricacies of the utility of arts and culture by limiting myself to the internal problems of (1). Sen's notion of informational base is applied to (1), in particular by confronting sum-ranking with Temkin's Second Standard View (Temkin 2012), and with Parfit's Individualist Restriction (Parfit 2011). I conclude that problems in (1) arise from the social dimension of arts and culture (Dworkin 1985). They involve a specific type of public good (a social good) because cultural goods are partly non-decomposable, as I shall explain using some ideas from Taylor (1990) and Munoz-Darde (2013). However, sum-ranking requires decomposable goods (so that A is a decomposable good because to be good is to be good for individual x, for w, for z, etc.). As a consequence, there is a case for a right to culture, so this right should be respected regardless of marginal utility gains from violations. The problems of this solution are not addressed in this contribution.

Vincent-Emmanuel Mathon (VEM Consulting). *Metrics for happiness, a tribute to Eleanor Sidgwick*

"Pleasure as feeling cannot be conceived" and a "sum of pleasures is intrinsically unmeaning" are two assertions made by Henry Sidgwick in his Method of Ethics.

This raises the issues of (1) quantifying pleasure and (2) comparing pleasures with each other as heterogeneous items cannot add up.

Those issues are similar to the challenge of quantifying electricity and magnetism.

Basically, metrics based upon water or matter - like kilogrammes - are easy to find out as their subject-matter can be "felt". And they can be summed up practically and easily as well.

The same is not true of electricity which cannot be "felt" as clearly as material bodies. It is equally difficult to form a mental picture of the very fact of adding up two electrical currents.

Yet physicist Rayleigh, in cooperation with Eleanor Sidgwick, found a method to establish objective metrics for electromagnetism.

What if this method could apply to the measurement of happiness? What if, by the same token, new metrics would allow us to objectively quantify pleasure in its overall complexity?

Eleanor Sidgwick's scientific works as applied to pleasure would then challenge Henry Sidgwick's initial assertion on the impossibility of summing up pleasures.

This would also open up a new, more subtle, way to mathematically modelize utility, where utility would not just be considered as a preference - like in most current models - but as the genuine pursuit of happiness.

Panel 4A—Gideon Schreier Lecture Theatre (124)

Sidgwick and Pain

Jennifer Hawkins (Duke University). *The badness of emotional pain*

Sidgwick, like other classical utilitarians, did not routinely draw a sharp distinction between sensory (or physical) pain and emotional pain. Yet the distinction seems important given that emotional pain and sensory pain seem so different and given that, for utilitarians, all pains count. Perhaps (though this is pure speculation) it was Sidgwick's inability to explain the commonality between emotional pain and sensory pain that ultimately led him to deny that pains are pains in virtue of all feeling alike. Rather, he held that an experience counts as a pain (and as bad for us) because we dislike it.

In my presentation, I present an account of emotional pain and its badness. Emotions are more complex than mere sensations, being composed of several mental elements. When you experience a painful emotion, intuitive thoughts arise in your mind, your attention is altered, and you experience certain bodily feelings and bodily changes. It can therefore be natural to assume that the painfulness of emotional pain is a feature of the bodily feelings that occur. However, I argue that the badness of emotional pain is actually an aspect of *the intuitive thoughts that arise* when we experience emotional pain. This is supported by noticing that we tend to treat emotional states as being worse (more painful) when a person is contemplating or thinking of a worse situation. After explaining how such a view can work, I demonstrate how, despite taking seriously the very real differences between emotional and sensory pain, we can explain (contra Sidgwick) the sense in which pains all feel alike, namely bad. I therefore show that we need not adopt Sidgwick's rather drastic claim that pains are only pains in virtue of our dislike.

Roger Crisp (University of Oxford) and Aaron Garrett (Boston University). *Sidgwick and Gurney on the incommensurability of extreme pain*

In a striking footnote to *ME* II.2, Sidgwick denied that "'torture' so extreme as to be 'incommensurable with moderate pain' is an actual fact of experience" (*ME* 6th ed. II.2 §1n) on the grounds that it neither corresponded to his own experience nor was "supported by the common sense of mankind." He characteristically qualified the claim with his evidence: "at least I do not find, in the practical forethought of persons noted for caution, any recognition of the danger of agony such that, in order to avoid the smallest extra risk of it, the greatest conceivable amount of moderate pain should reasonably be incurred." Gurney criticizes Sidgwick's view and argues for the incommensurability of extreme torture – such as vivisection – and moderate pain. We will begin our presentation with an account of the disagreement on the disvalue of pain between Edmund Gurney and Henry Sidgwick, in various editions of the former's *Tertium Quid* and the latter's *The*

Methods of Ethics. We will then turn to a defence of Gurney's position as well as a response to Sidgwick's appeal to expected utility theory. We will conclude with a discussion of the systemic importance of the issue for Sidgwick.

Panel 4B—Moot Court (XG01)

Global Priorities 2: Philosophical Problems concerning AI

Elliott Thornley (University of Oxford). *The Shutdown Problem: An AI Engineering Puzzle for Decision Theorists*

I explain the *shutdown problem*: the problem of designing artificial agents that (1) shut down when a shutdown button is pressed, (2) don't try to prevent or cause the pressing of the shut- down button, and (3) otherwise pursue goals competently. I prove three theorems that make the difficulty precise. These theorems show that a small number of innocuous-seeming conditions together preclude shutdownability. Agents with preferences satisfying these conditions will try to prevent or cause the pressing of the shutdown button even in cases where it's costly to do so. And patience trades off against shutdownability: the more patient an agent is, the greater the costs that agent is willing to incur to manipulate the shutdown button. I end on an optimistic note: these theorems can guide our search for solutions.

Christian Tarsney (University of Texas at Austin). *Deception and Manipulation in Generative AI: A Subjective Definition and A Minimally Paternalistic Solution*

Large language models now possess human-level linguistic abilities in many contexts. This raises the concern that they can be used to deceive and manipulate on a large scale, for instance spreading political misinformation on social media. In the longer term, agential AI systems might also deceive and manipulate humans to achieve their own goals, for instance convincing humans to connect them to the internet. This paper aims, first, to clearly characterize deception and manipulation in AI. I argue that these concepts must encompass more than literal falsehood, must not depend on controversial attributions of mental states to AI systems, and must not rely on third-party judgments of what human users rationally ought to believe or do. This leads to a characterization of deception/manipulation as, roughly, behavior that leads human users away from the beliefs/choices they would endorse under "semi-ideal" conditions in which they have been presented with all relevant information and have unlimited time for deliberation. Second, I propose two protective measures against AI deception and manipulation, inspired by this characterization: "extreme transparency" standards that require content creators to disclose the specific model variant and prompt used to generate content and full, unedited outputs, and training "defensive" systems to detect misleading output and contextualize AI-generated statements with relevant information for users. Finally, I consider to what extent these methods will guard against deceptive and manipulative behavior in future, agentic AI systems. I argue that non-agentic defensive systems can provide a useful layer of defense even against more powerful agentic systems.

David Thorstad (Vanderbilt University). *Against the Singularity Hypothesis*

The singularity hypothesis is a radical hypothesis about the future of artificial intelligence on which self-improving artificial agents will quickly become orders of magnitude more intelligent than the average human. Despite the ambitiousness of its claims, the singularity hypothesis has been defended at length by leading philosophers and artificial intelligence researchers. In this paper, I argue that the singularity hypothesis rests on scientifically implausible growth assumptions. I show how leading philosophical defenses of the singularity hypothesis (Chalmers 2010, Bostrom 2014) fail to overcome the case for skepticism. I conclude by drawing out philosophical implications of this discussion for our understanding of consciousness, personal identity, digital minds, existential risk, and ethical longtermism.

Panel 4C—Denys Holland Lecture Theatre (SB31)

Peter Lythe (University College London). *The Delusion Vanishes: Jeremy Bentham on the Disutility of Christianity*

In his 'Juggernaut' writings, composed between 1811 and 1821, Jeremy Bentham (1748–1832) delivered an unsparing attack on the moral value of Christianity. I will argue that, from his forensic investigation of the Gospels, Bentham concluded that the doctrines of Jesus either made no contribution to human happiness or else were sources of positive human misery. Special attention will be given to the critique that Bentham made of the Sermon on the Mount—the most extensive collection of Jesus' moral teachings—and his claim that it was neither 'desirable' nor even 'possible' that 'man's conduct should be shaped to it'. The religion of Jesus was the fount of numerous 'mischiefs', from the degrading of the intellectual faculties of the individual religionist to the inculcation of principles 'destructive of society itself'. Each of these mischiefs will be discussed in turn.

The religionist's belief that Christianity was 'the only basis of morality' originated either 'from the report of others or from imagination merely'. It was not, and could not be, the product of a 'close' and 'unprejudiced' examination of scripture, for if one measured this idea against the standard of the Gospel texts themselves, said Bentham, 'the delusion vanishes'. Bentham invited the reader of 'Juggernaut' to consider instead how much more receptive the defenders of religion might be to a system of morals that was 'exempt from these imputations'—that was capable of shaping 'general conduct' without the need for any deviation from the ordinary meaning of the words by which it was expressed. It will be seen that if it had been Jesus' intention to promote the greatest happiness of the greatest number, and if he had been guided by 'that degree of intelligence and wisdom of which human reason without any assistance from religion … is susceptible', then, in Bentham's view, he would have set his mind to developing a system of morality and law that was virtually indistinguishable from that laid out by Bentham in *An Introduction to the Principles of Morals and Legislation* (1789).

Carolyn Shapiro (Falmouth University). *St Paul and other Hustlers of Language: credulity, language and the movement of mischief in Bentham's Not Paul, But Jesus*

This paper looks at Bentham's *Not Paul, but Jesus* as a key to understanding what Bentham disparages as a dangerous alliance that can be forged between credulity and language by an imposter such as Paul. For Bentham, this alliance propagates mischief and possesses a particular ability to move across the "world": Paul's design of "worldly ambition" took advantage of what came to be characterised at that time

as a generalized abject credulity. I am interested in how Bentham's critique of Paul's abuses of his followers' credulity is crucial to Bentham's philosophy of language. Peter Lythe has suggested that Bentham's figure of "the reader" of Paul's texts denotes a position of susceptibility to the machinations of language, machinations which intentionally circumvent grounding determinations of a given time and place. This abstraction of language is exactly what makes mischief so apt to travel and to spread. The *at-large* quality attributed to the movement perfected by Paul is the movement of a swindler, and, as such, resonates with Bentham's attributes of fiction. Paul, then, can be understood as a linguistic phenomenon. But Paul's hustle of language must also be read within the context of "an age so fertile in imposture". The notorious figure of Count Cagliostro, literary phenomenon of the 18th and 19th centuries, who travelled across Europe, including London, perpetrating fraud after fraud and getting away with it, was more than a popular subject of literature. He was a figure which gathered around him larger philosophical considerations around truth, fraud, language, and credibility. I look to Thomas Carlyle's biography of Cagliostro as a companion text to *Not Paul, But Jesus*, to give social and cultural context to the anxiety around fraud, imposture and belief during Bentham's lifetime, but also as a text which emphasises the movement of Cagliostro's counterfeit as one which traverses the world.

Philip Schofield (University College London). Logic, Language, the Disciplines, and Ethics: Jeremy Bentham's encounter with Aristotle

In the mid-1810s the English philosopher Jeremy Bentham (1748–1832) composed a series of works on logic, language, the arts and sciences (the disciplines), and ethics. I argue that the connection between these works, which include 'Essay on Logic', 'A Table of the Springs of Action', 'Deontology', and 'Chrestomathia', is the key to understanding Bentham's deep commitment to practical reform and improvement. Bentham's conception of logic emerged from his engagement with and response to the Aristotelian tradition of logic as it was taught to him at the University of Oxford through the seventeenth-century 'Compendium' of Robert Sanderson. In the place of Aristotle's *præcognita*, Bentham developed what he termed the 'characteristics' of logic. For Bentham, the 'end' of logic was the promotion of well-being or happiness, its 'field' covered that of the whole of the arts and sciences, while its 'instrument' was language. Bentham did not reject, however, the whole of Aristotelian logic, but adopted a particular version of the standard Aristotelian method of definition *per genus et differentiam*. This method led him to devise a map of all the arts and sciences, including logic and ethics, represented by an encyclopaedical tree. There is a tendency for commentators to assume that Bentham's thought represents a 'crude' form of utilitarianism, whereas a focus on his logic shows that his commitment to the principle of utility rested on an original and profound philosophical basis.

Panel 4D—Room LG17

Ariadne's Thread: Out of the Labyrinth of the Mill-Taylor Collaborations

The Panel aims to shed light on the nature and extent of the collaborations John Stuart Mill engaged in with his partner and later wife, Harriet Taylor Mill and his stepdaughter, Helen Taylor. It both looks at the general nature of these cooperations, the evidence we have for specific claims about them, as well as how and why they were – and still are even today – hard to accept or acknowledge for their contemporaries. In addition to presenting novel evidence gained by traditional ‚close-reading' of often neglected works, it also displays innovative ways of investigating co-authorship using ‚distant-reading' tools from the digital humanities.

Jo Ellen Jacobs (Millikin University). *Creative Pairs and Why Philosophers Don't Understand Them*

If we begin with the belief that Harriet and John collaborated, then the question of why philosophers resist this conclusion must be explored. Traditionally philosophy has privileged the single author. Like Zeus giving birth to Athena, philosophers, unlike scientists, assume deep thought must be born of a single parent. This romantic fiction of the lone thinker is buried in the history of philosophy. Because we philosophers are typically unaccustomed to writing jointly, we do not understand the process and types of creative duos. HTM and JSM serve as examples of the ways partners work in tension with one another.

Helen McCabe (University of Nottingham). *John Stuart Mill and Harriet Taylor: Collaboration and Co-Authorship of "our story" in the Autobiography*

John Stuart Mill ascribed several different roles to Harriet Taylor in the composition of his works from 1830 onwards, including editing; suggesting topics to work on; contributing discrete parts; developing ideas which Mill "translated" to the rest of the word; and working together on every line of a piece. Many, if not all, have been variously denied by commentators on Mill, from his own day to ours, in a way which prevents us from properly assessing her place in the canon of the history of political philosophy, and adds to a general erasure of women from this history, though there are contemporary attempts to address this, for many ages and branches of philosophy.

In this paper, I explore Taylor's role as Mill's "editor", in particular through her role in the developing of his *Autobiography*, an early draft of which she edited. Among other things, the *Autobiography* was seen by Mill as a chance for them to tell "our story", with decisions to be taken about "how much" of it is was "advisable to tell", balancing openness (or lack of "reserve") with the risk of "put[tings] arms into

the hands of the enemy". Mill felt that people were "comparatively willing" to believe Taylor's "effect on him" emotionally, but not "intellectually", which "though it is the smallest part of what you are to me, is the most important to commemorate" precisely because people were so unwilling to credit it. This unwillingness is still seen today in commentary on Mill (and Taylor). In understanding each of her roles in the construction and production of "his" texts, we gain a deeper understanding of that "effect" in its many facets, and of the development of his, of her, and of "their" thought (whosever name it was published under). In addition, understanding her role in editing the *Autobiography* allows us to critically reflect on Mill's portrait of Taylor, and his account of their writing relationship, recognising her involvement in both. This is crucial for a proper understanding of Mill, of Taylor, and of the collaborative nature of much philosophy, something which often goes unacknowledged or ignored.

This paper is part of a wider project tracing and exploring all of Taylor's roles in the construction of Mill's oeuvre (and his in hers), as well as the role played by Helen Taylor in later years. I focus on editing as one key role, not least because it invites this critical reflection on Mill's (and Taylor's) account of those roles.

Antis Loizides (University of Cyprus). *Stylometrics and The Subjection of Women*

John Stuart Mill claimed that *The Subjection of Women* (1869) was co-authored by himself, Harriet Taylor Mill, and Helen Taylor. According to his *Autobiography* (1873), his work should be thought as "the product not of one intellect and conscience but of three." Most of *Autobiography*'s readers have been largely unconvinced both by his claims of co-authorship and by his encomiums of his co-authors. His unreserved praise of their abilities, to put it plainly, backfired. Rather than strengthening the claims of a common "fund of thought", collaboration, and co-authorship, his testimony to their abilities undermined them. Those who are most reluctant to take these claims at face value, reject the idea that not only did Harriet Taylor Mill have an active, pervasive, and ever-lasting part in John Stuart Mill's writings, but that she also was the originator of some of his most characteristic ideas.

Unlike her mother, Helen Taylor has never actually gotten any consideration as her stepfather's co-author. Should we accept a key tenet of stylometric studies, that an author's mind engrafts itself onto the text, then we might be able to examine J.S. Mill's claims of co-authorship. This paper presents the results of one such computer-assisted stylometric analysis of *The Subjection of Women*.

Christoph Schmidt-Petri, Michael Schefczyk, and Lilly Osburg (Karlsruhe Institute of Technology). *Co-Authorship of the Principles of Political Economy and On Liberty: Further Stylometric Analyses*

In previous work, we have shown that using initially plausible assumptions, Harriet Taylor Mill has probably contributed more to *On Liberty* (1859) than is generally accepted. Stylometric analyses – that is, using all relevant writings by both him, her,

and their jointly authored texts to excerpt an individual writing style, and testing which of these styles show up in *On Liberty* – indicate that John Stuart Mill almost certainly is not the exclusive author of chapter 3, in particular.

Whether or not such analyses are convincing rightly depends on whether or not the initial assumptions are considered plausible by the philosophical expert. As opinions on our initial authorship attributions might differ, we ran further analyses to determine whether and how our results would change, taking a large number of similarly plausible starting points. We focused in particular on the *Principles of Political Economy* (1848), which initially we had considered jointly authored, assigning the famous chapter on the "Futurity of the Labouring Classes" to Harriet Taylor Mill.

In this paper, we present our stylometric analyses of the *Principles* as well as how these result impact our analyses of *On Liberty*. Our previous results turn out to be robust under a large variety of assumptions.

Panel 4E—Room LG10

Sujith S. Kumar (University of Maryland). *Penalties, Punishments, and the Principles of Ethology in On Liberty*

Recent scholarship has identified a potential contradiction that JS Mill makes in *On Liberty* between the punishments and penalties that a person rightly experiences for certain kinds of failures (Waldron 2002, Kumar 2013, Miller, 2022). Punishment is necessarily meted out by society for violations of the Principle of Liberty, whereas penalties are meant to result from the purely self-regarding character flaws and the exercise of poor judgement that certain people exhibit. However, if the penalties in practice can amount to punishments, then Mill fails in his entire project in *On Liberty*. This paper attempts to resolve this contradiction by clarifying how these punishments and penalties work, and by showing how even if they are experienced as coercive, they are not as such in the relevant way, according to the Principle of Liberty. Specifically, I construct principles of ethology, based on Mill's associationist psychology, to show how the penalties are intended to attach to second order reflection on one's actions, whereas punishments are intended to influence first order practical reasoning. *A System of Logic* presents Mill's social-scientific methodology, and declares his intention to establish a new science of character formation, called ethology, upon which he would base his social and political prescriptions. Mill fails to establish this science, nor even to articulate a single ethological principle. And yet, some recognize that several of his writings are based on this implicit theory of character formation (Ball 2000), and so provide the grounds on which to construct such principles.

Adam Kolber (Brooklyn Law School). *Punishment for the Greater Good*

Over ten million people are incarcerated throughout the world, even though punishment theorists have struggled for centuries to morally justify the practice. Theorists usually address criminal justice under abstract, idealized conditions that assume away real-world uncertainty. We don't have time, however, to wait for a perfect moral theory, and the history of philosophy suggests we will never find it.

In a forthcoming book, *Punishment for the Greater Good* (Oxford University Press), I examine the justification of punishment in the here and now, recognizing that we are uncertain about matters of both fact and value. Rather than seeking retribution, we should reduce suffering by deterring crime and incapacitating and rehabilitating dangerous people. Though this consequentialist approach has fallen out of favor in recent decades, it is better suited to addressing punishment in the here and now than the approach commonly taken by retributivists. If consequentialism successfully justifies punishment, then contrary to some carceral

abolitionists, at least some incarceration under some conditions is justified today. In the here and now, we will rarely know how to punish for the greatest good, but we can seek to punish for the greater good.

My presentation will summarize the book and then focus on the principal challenge retributivists pose to consequentialists, namely that traditional consequentialism lacks a proportionality limit on punishment. I argue that retributivists have no adequately detailed method of assessing proportional punishment in the here and now, and if we take proportional punishment seriously, we will see that it is actually quite counterintuitive.

Hahn Hsu (National Chung Cheng University). *Mill, Sanctions and the Practical Normativity of Morality*

In chapter 3 of *Utilitarianism*, John Stuart Mill addresses the issue of 'the ultimate sanction of the principle of utility' or sanctions of morality in general. Let me quote him to provide textual basis for further discussions:

> THE QUESTION is often asked, and properly so, in regard to any proposed moral standard – What is its sanction? What are the motives to obey it? or more specifically, what is the source of its obligation? whence does it derive its binding force? It is the necessary part of moral philosophy to provide an answer to this question… （U, 3, 1）

Crisp, and many other scholars, sees Mill's view of sanctions as to give psychological explanations of how the utilitarian ethics could motivate. David Brink (SEP 2018), interestingly enough, forgoes the above passage but cites another passage in *Utilitarianism* to argue that Mill's discussion of sanctions can be seen as to provide a sanction view of utilitarianism, which is a form of indirect utilitarianism. Christine Kosgaard (1996: 315) sees Mill's view as a clear example of moral externalism, since according Korsgaard, Mill entirely separates the proof of the utilitarian principle from its motivation from sanctions.

These scholars' views suggest that Mill's view of sanctions of the utilitarian morality can allow different readings. Nevertheless, I intend to trace beyond the lines of thought provided by these scholars further into Mill's view of sanctions of morality. I wish to begin this enquiry not just by seeing it as about moral motivations and reasons, but going into the social aspects. From my perspective, Mill's idea of sanctions of morality can inspire a broader conception of the normativity of morality, especially regarding the realization issues. I wish to argue that Mill suggests ideas about how morality can and ought to be realized as genuinely practical.

Panel 5A—Denys Holland Lecture Theatre (SB31)

Christophe Salvat (CNRS, Aix-Marseille University). *Parfit on well-being: a Sidgwickian theory of desirability*

Derek Parfit established a categorization of well-being theories that has now become canonical, based on three sets: theories of mental states, satisfaction theories, and objective list theories. But to which type of theory can his moral philosophy be linked? For many commentators, Parfit advocates an objective perspective on well-being. In reality, it is very difficult to apply Parfit's framework to his own philosophy. After refuting hedonism and preference satisfaction theory, Parfit seems inclined towards the values advocated by objective list theories, including moral goodness, rational activity, development of one's capacities, having children and being a good parent, knowledge, and an awareness of true beauty.

However, accepting the conclusions of a theory does not imply approving the theory itself, and Parfit is also very critical of certain aspects of this theory. He presents the three main perspectives on well-being in Appendix I but does not attempt to decide among them as it would take at least an entire book to settle among the different theories of self-interest.

Despite his previous statement, Appendix I goes beyond a mere presentation of the various possible approaches to well-being. His reflections foreshadow a theory heavily inspired by Sidgwick and centered around the concept of desirability, a concept he believes encompasses both the subjectively pleasing and the rational dimension of its object.

The purpose of my presentation is to attempt to reconstruct this theory that Parfit was never able to fully develop, drawing from various elements present in *Reasons and Persons* as well as from other published works, including the On What Matters' chapter in which he discusses the value of suffering in Nietzsche's philosophy.

Sergei Talanker (Western Galilee College). *Experimental challenges to Sidgwick's congruency of intuitions thesis*

In *The Methods of Ethics* Sidgwick presents a congruency thesis which relates to three phases of moral intuitionism: perceptual, dogmatic and philosophical. His thesis is that these phases map onto each other: what we immediately intuit in particular case is consistent with "general moral rules that we discern with clear and valid intuition", which are consistent with "one or more absolutely and undeniably true principles". He advocates for a bottom-up method of inducing moral principles from the common sense intuitions about particular cases over the top-down method of deducing moral judgments about particular cases from the general principles, assigned by reason. Thus, he is convinced, the congruence between ethical theory and common sense moral practice may be assured.

I would like to investigate whether Sidgwick's position is tenable in light of the experimental studies that challenge the congruency of people's intuitions about cases with their beliefs about general moral rules. According to Haidt and Joseph (2004), our responses to moral dilemmas are quick gut feelings that become conscious when situation is presented to us. Then, when asked to explain their judgment, we use the reasoning system to search for justifications. Thus, focusing on the reasons people give for their judgments is like studying the rational tail wagged by the emotional dog (Haidt 2001; Haidt and Joseph 2004).

Nichols and Knobe (2007), likewise observe that people's moral intuitions about specific cases and those regarding moral theories sometimes tend to be incongruent. They offer several models that are supposed to explain this phenomenon. For instance, the *affective performance error* model posits that people ordinarily make judgments by relying on a tacit theory, but in extreme cases, they experience a strong affective reaction, which makes them unable to apply the theory correctly. The *affective competence* model, comparing morality to syntax, posits that just as people compose proper sentences, but cannot properly formulate the rules of syntax, people have proper moral intuitions, but cannot properly formulate the moral principles behind them.

Christian Seidel (Karlsruhe Institute of Technology).
Sidgwickian Moral Theorizing: on the Teleological Relation between Principles and Methods

Sidgwick's *The Methods of Ethics* (ME) seems to be about methods of ethics. However, what exactly is meant by methods and how they are related to principles is notoriously in need of clarification. In this paper, I elaborate a Sidgwickian conception of moral theorizing that emerges from *ME* when these questions are examined more closely. Through a critical discussion of differing interpretations (e. g. by Schneewind 1977, Brink 1988, Daurio 1997, Crisp 2015), I argue that, according to Sidgwick,

1. moral theorizing has a theoretical and a practical aim: (a) to obtain systematic and general knowledge about the ultimate reasons that determine the deontic status of an action, and (b) to provide agents with deliberative guidance in answering the practical question "What ought I to do?";

2. principles are meant to serve the first aim and to answer the question "What ought I to do?" by identifying those ultimate reasons, i. e. the right-making properties;

3. methods are meant to serve the second aim and to answer the question "What ought I attend to (or be guided by) when deliberating about the question 'What ought I to do?'?" by identifying the rightness- indicating properties;

4. a moral theory is composed of two elements, a principle *and* a method;

5. in an adequate moral theory, principles are teleologically related to methods: in practical deliberation about what one ought to do, we ought to use the method that best achieves the ultimate aims (i. e. the method that is best supported by the ultimate reasons) given by the principle.

The latter is a substantial assumption that clearly distinguishes the Sidgwickian conception of moral theorizing from a more Kantian, constitutivist view (according to which the right-making properties *necessarily* coincide with the rightness-indicating properties and the method that ought to guide practical deliberation *constitutes* the principle). I conclude by discussing the ramifications of the proposed interpretation (a) for understanding the dialectic progression of the principal argument in *ME*, and (b) for understanding the dualism of practical reason.

Panel 5B—Moot Court (XG03)

Global Priorities 3: Hedonic Wellbeing

Heather Browning (University of Southampton). *The Complexities of Intrapersonal Welfare Comparisons*

Some contexts of welfare assessment require comparisons of welfare *within* an individual or species, such as comparing welfare across time, or in different living conditions. However, making such comparisons requires solving difficult problems regarding the comparison and integration of different affective experiences with both positive and negative valence (e.g. fear, pain, comfort) as well as different aspects of welfare experience (e.g. intensity and duration). In this talk I will outline the primary issues that arise for making intrapersonal comparisons, and indicate how we might make progress in solving them.

Pietro Cibinel (Princeton University). *Hedonic Tradeoffs*

On a representative characterization, hedonists believe that

> [...] what is good for any individual is the enjoyable experience in her life, what is bad is the suffering in that life, and the life best for an individual is that with the greatest balance of enjoyment over suffering. Crisp (2006, p. 622).

The expression "greatest balance of enjoyment over suffering" sounds quite intuitive, and variants of it are widely used to characterize hedonism. The aim of this talk is to explore this expression in some detail. I argue that there is no single, default meaning for hedonists to attach to this expression. If there are quantities of pleasure and quantities of pain, these are quantities of different dimensions. Moreover, it is unclear whether there are quantities of pleasure and pain. As I explore different ways to make this expression precise, I will highlight various choice points that hedonists face. On the resulting picture, hedonism stands in for a large class of axiological views, some of which deliver greatly dissimilar verdicts.

Mattia Cecchinato (University of Oxford). *No Consciousness, No Welfare*

Many philosophers hold that only conscious beings can be welfare subjects, subjects for whom things can be intrinsically beneficial (or harmful). This popular view—call it Phenomenal Necessitarianism—has, however, recently met some sceptics. Advocates of the non-necessitarian view argue that at least some putatively non-conscious entities, such as some sophisticated artificial intelligences, can have mental states like beliefs and desires, which they contend are sufficient to form welfare interests. This paper makes a new argument for Phenomenal Necessitarianism based on the special type of epistemic justification we have to believe in it. I argue that we have certain distinctive kinds of introspective

experiences which immediately provide prima facie justification for a necessitarian belief. I also contend that there is no appropriate defeating evidence against this introspectively justified belief—neither from intuitions, nor from wellbeing theory, nor from philosophy of mind. Since, in the absence of defeating evidence, prima facie justification should be taken as all-things-considered justification, I conclude we ought to be necessitarians.

Panel 5C—Gideon Schreier Lecture Theatre (124)

Marco E.L. Guidi (University of Pisa). *Utilitarian virtue and Bentham's theory of democracy*

Since the publication of Alasdair MacIntyre's *After Virtue* in 1981, utilitarianism has been attacked as a universalist ethical theory focusing on the morality of actions and/or rules of behaviour rather than on the morality of persons. Nevertheless, the ethical works of Bentham, Mill and Sidgwick contain a theory of virtue, intended as intertemporal consistency in the application of the principle of utility to personal behaviour and choice, and as self-education and self-improvement based on fostering the evolution of personal motivations from enlightened self-interest to enlightened benevolence. This paper aims to explore Bentham's analysis of virtue in order to assess its impact it on his economic and political thought. His reaction to the Aristotelian and Christian ethics of virtue is rather concealed in his works, although it can be traced in various references he made in his correspondence and in *Memoirs* to his juvenile reading of Fénelon's *Les Aventures de Télémaque*. Furthermore, while Jean-Baptiste Say, in his *Essai sur le principe d'utilité* (1829), identified virtue with enlightened self-interest and, as revealed by his *Catéchisme d'économie politique* (1815), considered political economy as the main instrument of personal enlightenment, and while Bentham himself based his democratic theory in *Constitutional Code* on the assumption that, as Hume had argued, "in contriving any system of government, and fixing the several checks and controuls of the constitution, every man ought to be supposed a knave, and to have no other end, in all his actions, than private interest", in *Deontology* (and also in *IPML*), he argued that human beings can be educated to cultivate self- and other-regarding virtues. The connection between Bentham's deontological ethics and his political thought is explored in the final part of this paper.

Gustavo H. Dalaqua (University of Pernambuco). *Democracy and Representation: Rousseau and Mill*

Democracy and representation have different origins and meanings. The word democracy (*dēmokratía*) was coined in ancient Greece and means literally power of the people. Representation (*representatione*) is of Latin origin. Its first parliamentary usage traces back to 1188, when the king Alfonso IX summoned in the Iberian Peninsula the *Cortes de León*. Representation was a means for aristocrats to make decisions as if they were taken by the whole community.

Given their different origins and meanings, it is understandable that some scholars oppose democracy to representation. Consider, for instance, Jean-Jacques Rousseau's *The Social Contract*. In this book, Rousseau argues that representation destroys democracy. But what are the reasons Rousseau presents to justify the incompatibility between representation and democracy and how can one contest them? The main objective of my presentation is to answer these questions. To fulfil it,

I will engage in a textual analysis of Rousseau's *The Social Contract* and then examine how John Stuart Mill refutes Rousseau's assumptions in *Considerations on Representative Government*.

Rousseau's opposition between democracy and representation is based on the following assumptions: (1) sovereignty consists only in the will; and (2) political representation always entails a delegation of political power. Rousseau contends that Englishmen fool themselves when they say that popular sovereignty lies in their representative government. Popular sovereignty cannot be represented because sovereignty is an act of will, and one cannot delegate one's will to somebody else.

Mill took very seriously the criticism of England's representative government put forward in *The Social Contract*, and that is why he elaborated the concept of *democratic representation*. This concept can be invoked to challenge Rousseau's assumptions because it shows that (1) sovereignty encompasses not only the will, but also judgement; and (2) representation need not entail the delegation of political power from the represented to their representatives.

Tsin Yen Koh (University of Singapore). *Two Strands of Utilitarian Democracy*

This paper sketches out two strands of utilitarian democracy in the political thought of Jeremy Bentham and John Stuart Mill.

The main claim is that there are two aspects to Bentham's theory of government: a 'radical democratic' aspect and an 'epistemic' aspect. Both aspects are necessary to good government: a government that aims at the greatest happiness of the greatest number is more likely to succeed if it has the form of a representative democracy, and securities for sufficient expertise (or appropriate aptitude) in all branches of the constitution. The *Plan for Parliamentary Reform*, for example, is concerned both with democratic control of the government *and* with securities for appropriate intellectual aptitude in the electorate and the legislature. The *Constitutional Code* can be read as an attempt to unify the twin demands of democracy and expertise in the structure it sets out of responsible government.

However, there is an ineradicable and perhaps irresolvable tension between these two aspects of representative democracy, even in the *Code*: they can come apart, and can pull in different directions. I suggest that we see in Mill's political thought a hint of the different directions they develop in: the 'radical democratic' strand takes a 'socialist' turn, in its emphasis on the materialist basis of social and political equality; and the 'epistemic' strand takes (what one could anachronistically call) a 'technocratic' turn, when its emphasis on expertise shades into an emphasis on experts.

Panel 5D—Room LG10

Daisuke Arie (Yokohama National University). *What does J.S. Mill's "Stationary State" imply for modern capitalism?*

Currently, there is a growing distrust and criticism of "capitalism" as the dominant economic system of our time, typified by the rapid population growth in the world, as well as extreme income inequality, depletion of natural resources and environmental destruction. The purpose of this paper is to examine what implications J. S. Mill's discussion of the "stationary state" in his *The Principles of Political Economy* (1848), Volume 4, Chapter 6, has for the discourse of criticism of capitalism.

First, I will identify how Mill viewed the British economy in the mid-19th century and what he saw as its main problem: he saw the unequal distribution of wealth as the biggest social problem of the time.

Second, I will show where Mill saw the causes of the problem: He believed that in the process of long-term economic expansion, land rents tend to rise and harvests diminish because of the limited amount of land available. Mill saw population growth and resource constraints as the two main causes. However, although Mill criticized the greedy capitalist tyranny that brought about such a situation, he did not go as far as Marx in his fundamental rejection of the private property system itself.

Third, Mill recognized the economic standstill and then formulated a prescription for how society could be maintained in such a 'stationary state' as reasonable as it should be: if the members of society possessed certain distinctive moral qualities that were not capitalist, under which only limited resources would prevent growth, then society would be stable, sustainable, and happy. Mill believed that if the members of a society possessed certain moral qualities that were not capitalist, only then could a stable, sustainable state of well-being be maintained with limited resources, but without growth.

Finally, I will show to what extent Mill's views were valid as a grasp of the times in the 19th century, and to what extent they can be realistic today.

Tim Beaumont (Shenzen University). *John Stuart Mill on the Suez Canal and the Limits of Self-Defence*

Michael Walzer's use of John Stuart Mill's *A Few Words on Non-Intervention* (1859) helped to inaugurate it as a canonical text of international theory. However, Walzer's use of the text was highly selective because he viewed the first half as a historically parochial discussion of British foreign policy, and his interest in the second was restricted to the passages in which Mill proposes principles of international morality to govern foreign military interventions to protect third parties. As a result, theorists tend to see those canonized passages as if through a glass darkly. Attention to the

detail and context of Mill's first-half critique of Lord Palmerston's opposition to the Suez Canal project reveals that his discussion of purely protective intervention is embedded in a broader exploration of the limits of self-defence, including the moral permissibility of preventive military force and the use of protective interventions for defensive purposes. Moreover, reading the text holistically facilitates a refutation of some objections directed at it by Michael Doyle to the effect that Mill's conception of self-defence incorporates elements of aggression which makes it extremely dangerous when adapted for application to the contemporary world.

Antonin Broi (Wuhan University). *Toward metaethical hedonism*

In this presentation, I will introduce and defend *acquaintance-based nonreductionist hedonism*, a variant of hedonism that doubles as a meta-ethical position.

Axiological hedonism (henceforth hedonism) is the view that pleasure is the only thing that has positive final value and displeasure is the only thing that has negative final value. It is commonly discussed in value theory, but also in the philosophy of well-being, where the kind of value in question is restricted to prudential value. Most debates about hedonism proceed in accordance with the method of reflective equilibrium, understood as an "intuitions-based" approach, where a variety of often conflicting normative intuitions are brought to bear on the plausibility of ethical theories (see e.g. Feldman 2004; Crisp 2006). This approach, however, runs the risk of overlooking the specific epistemological and metaphysical status that pleasure and displeasure have, as *introspected* and *phenomenal* states.

Acquaintance-based nonreductionist hedonism is a variant of hedonism that makes the most of these two features of pleasure and displeasure. It supplements axiological hedonism with (1) the epistemological claim that by experiencing pleasure and displeasure we are directly acquainted with their value (acquaintance-based hedonism) and (2) the metaphysical claim that pleasure and displeasure, and thus their value as well, are not reducible to physical or functional properties (nonreductionist hedonism).

What we get is an ambitious integrative theory that sheds light on several traditional metaethical challenges. First, it provides a convincing case for the objectivity of axiological properties, where objectivity is understood as axiological properties being independent from the motivational states (e.g. desires) of anyone, or more generally being stance-independent (Shafer-Landau, 2003). Second, nonreductionist hedonism goes some way in accounting for the queerness of axiological properties with respect to physical properties, as it basically makes axiology piggyback on phenomenality, thus supporting the existence of *axiological qualia*.

Panel 5E—Room LG17

Lisa Forsberg (University of Oxford). *Achievement and the Value of Effort*

What makes an act valuable as an achievement? One answer is that the achievement value of an act supervenes on the product of the act, that is, the end result of the act, what the act achieves. A rival answer is the Effort View: in addition to the product of the act, the act's achievement value supervenes on the effort involved in producing the product. In this paper, we argue against the Effort View. We argue that cases where effort is optional are problematic for the Effort View.

Héloïse Michelon (Catholic University of Lille). *Can we reconcile the Non-Identity Problem and the Harm Threshold in Reproductive Decision-Making?*

Derek Parfit's non-identity problem (NIP) challenges the notion of harm in cases where our choices affect the identity and quality of life of future persons. Parfit argues that harm prevention goes beyond individual victims and extends to the occurrence of harm regardless of the identity of a potential victim, thus shifting from a person-affecting view to an impersonal perspective. The NIP holds significant implications for ethics and legal theory, notably in the context of reproductive decisions. While Mill's harm principle effectively addresses the harm caused to unborn children by administering teratogenic drugs to pregnant women, it struggles to explain why women taking such medication for their personal benefit should abstain from having children. The NIP provides a rationale for such a position: if potential suffering can be avoided by simply waiting, one should do so. Nevertheless, acknowledging that Parfit's principle may lead to an endless pursuit of marginal improvements raises the question of how long procreation should be delayed in the pursuit of a slightly better situation. Some scholars advocate for a minimum threshold below which all procreation is deemed irresponsible, transitioning from a comparative to a non-comparative theory.

Parfit himself seems to endorse such a threshold: he often emphasizes that the lives in his non-identity cases are worth living. Should we conclude that if the child is expected to live a life that is not worth living, their birth should be avoided ? The debate surrounding the compatibility of this threshold with the NIP and whether Parfit explicitly endorses both remains ongoing. This paper aims to clarify the connections between the NIP and threshold approaches, probing the coherence of a theory that combines the substitution principle and harm threshold.

Martí Colom (Barcelona). *Is abortion more serious than murder? Peter Singer's shift from preferentialism to hedonism and its effects on his practical ethics*

Peter Singer is one of the leading figures in contemporary utilitarianism. His work can be organised in two major stages. In the first stage, he follows R.M. Hare and his subjectivist universal prescriptivism and preference utilitarianism. In the second stage, influenced by Henry Sidgwick and Derek Parfit, Singer transitions towards objectivism and hedonistic utilitarianism.

As far as practical ethics is concerned, it is during the first stage that Singer develops the bulk of his positions on the matter, which, in most cases, he has not revised during the second stage. It is legitimate to wonder whether these metaethical and normative changes have had some repercussions on his practical ethics, and whether those original positions, the one's that made him known, can still be defended from these new foundations.

We plan to illustrate this examining the following case: (i) Singer has always admitted that death causes extrinsic harm in comparative terms, because of the deprivation in value it entails; (ii) in his second stage, due to his hedonistic commitment, self-consciousness loses intrinsic moral significance, and the relevant property in relation to the disvalue of inflicting pain, but also of death, is sentience; (iii) on the question of the prudential value of the future, Singer argues, following Sidgwick, that we should give equal weight to all parts of our life, whether closer or further in time to the present.

It follows that, *ceteris paribus*, in Singer's second-stage conceptual framework an abortion is more serious than the murder of a *person* – because it involves a comparatively greater loss of utility. This conclusion is a radical departure from what Singer has argued throughout his work, for example in *Should the Baby Live?* (1985). We explore different ways in which he could deal with this situation, either by finding new ways to defend its old position, or by leaving it behind.

Panel 6A—Gideon Schreier Lecture Theatre (124)

Author-Meets-Critics on Anthony Skelton's forthcoming *Sidgwick's Ethics*: Roger Crisp (University of Oxford), David Phillips (University of Houston), and Rob Shaver (University of Manitoba)

Skelton's forthcoming *Sidgwick's Ethics* interprets and critically evaluates select positions and arguments in Henry Sidgwick's *The Methods of Ethics*. It focuses specifically on Sidgwick's moral epistemology, his argument for utilitarianism, his argument for rational egoism, and his argument for what he calls "the dualism of practical reason", the thesis that utilitarianism and rational egoism are coordinate but conflicting requirements of rationality. In respect of these positions, *Sidgwick's Ethics* argues for the following theses. First, that Sidgwick's moral epistemology appeals both to claims about what is self-evident after reflection (or philosophical intuitions) and to coherence relations. Second, that Sidgwick develops a sophisticated, multi-stage argument for utilitarianism that fails because it cannot secure agreement on all the philosophical intuitions it relies on. Third, that Sidgwick (a) has a sophisticated, multi-stage argument for rational egoism that appeals both to philosophical intuitions and to a metaphysical claim about the distinction between individuals and (b) that Sidgwick's argument fails because, like his argument for utilitarianism, it cannot secure agreement on all the philosophical intuitions it relies on. Fourth, that though Sidgwick cannot establish many of the positions he endorses, he nonetheless leaves us with a detailed understanding of the philosophical challenges associated with defending theses in ethics.

Author

Anthony Skelton (Western University)

Critics

Roger Crisp (University of Oxford)

David Phillips (University of Houston)

Rob Shaver (University of Manitoba)

Panel 6B—Moot Court (XG03)

Global Priorities 4: The Moral Implications of Evidential Decision Theory

Hayden Wilkinson (University of Oxford). *The Moral Stakes of Evidentialism*

Evidential decision theory (EDT) tells us that, in situations of uncertainty, our instrumental reasons are determined by our evidence, *including* the evidence given to us by our making a particular choice. In this paper, I present and evaluate an argument that, when applied to moral decision-making, EDT implies that agent-neutral reasons carry far greater weight than we might have otherwise thought. Under certain empirical assumptions, EDT implies that our agent-neutral reasons defeat our agent-relative reasons in *every* situation where they conflict—to an approximation, EDT rules out any verdicts incompatible with consequentialism.

In brief, the argument goes as follows. When an agent makes any given decision, they do so in a world containing many other similar decision-makers who face similar decisions. Make a particular choice, and the agent gains evidence that those decision-makers will choose the same. Because of this, EDT effectively raises the stakes of our moral decisions with respect to our agent-neutral reasons—an action that causally saves one life will, by EDT, be treated as saving many more lives, and hence have much stronger instrumental agent-neutral reasons in favour of it. Meanwhile, EDT does not raise the stakes of our decisions with respect to agent-*relative* reasons—by killing one, an agent gains evidence that others will also kill, but they are not the killer of those additional victims, so their agent-relative reasons remain the same. So, in any decisions where the two conflict, it seems that an agent's agent-neutral reasons may defeat their agent-relative reasons. In this paper, I consider whether this argument works, what assumptions we need to reach the conclusion, and whether it should lead us to reject EDT.

Andrea Petrou (London School of Economics). *The Reference Class Problem is the Acausal Trader's Problem Too*

MacAskill *et al* (2021) argue that, according to evidential decision theory (EDT), if there are agents sufficiently similar to us and our behaviour is thus correlated to theirs, we should expect that if we perform a certain action, they will too. Hence, we could engage in cooperation games based on correlation rather than causation, sometimes referred to as *acausal trade*. In this talk, I argue that EDT cannot ignore the reference class problem that arises when spelling out what it means for another agent to be 'correlated' to oneself, making it difficult to say whether we have reasons to do acausal trade, how strong our reasons are or with whom we would be trading. We might have an intuitive sense of what it

means for an agent to be 'similar' to another, but whether and how much similarity exists depends on which qualities we are focusing on or find important. There are multiple rationally permissible ways to classify actions or agents based on their features, leading to vastly different measures of correlation. The only way to non-arbitrarily select a reference class is to introduce some criterion of relevance, whether that's causal relevance, normative relevance or contextual relevance. I argue that the first is not available to the evidentialist, who wishes not to depend on any notion of causation. The second, I argue, would be unsatisfying for proponents of acausal trade, who wish to remain neutral on normative facts and leave space for normative uncertainty. The third strategy is unsatisfying because it would fail to give a consistent answer as to what acausal trade requires of us over time. I discuss some potential responses, most notably an appeal to the likelihood of making the same choices and an appeal to evolution and reject both.

Timothy L. Williamson (University of Oxford). *Evidential Decision Theory and Widespread Incommensurability*

How should we compare acts whose outcomes may be *incommensurable* (or on a par, or incomparable), as many acts seem to be when evaluated morally? *Prospectists* answer this question by saying that we should evaluate acts based on how likely they make outcomes, and nothing else. Rivals to prospectism say that something else matters (typically facts about which outcomes occur in which states). The debate between Prospectists and their rivals is of great practical significance for consequentialists (and probably anyone who thinks that consequences matter to some degree). In particular, I argue that if Prospectism is false and consequentialism is true then we likely face incommensurability between all of our options in virtually every decision that we might face. By contrast, if prospectism is true, then we can avail ourselves of decision-theoretic tools that resolve incommensurability. If and only if Prospectism is false, everything is permitted.

   Evidential Decision Theorists are those who think that it is instrumentally valuable for an act to signal, but not cause, the good. The debate between Evidentialists and their rivals is often viewed as an esoteric one over the foundations of decision theory that only has practical implications in exotic choice situations. Not so. I show that Evidential Decision Theory entails Prospectism. Intuitively, the reason for this is that Evidential Decision Theorists care only about correlations between acts and outcomes, which lends itself to the Prospectist way of thinking. I make this thought precise by establishing a formal connection between EDT's characteristic property of *partition invariance* and Prospectism.

   This furnishes those who accept Evidential Decision Theory with an argument for Prospectism. And conversely, it shows that those who want to resist widespread incommensurability should look to Evidential Decision Theory.

Panel 6C—Room LG10

Gregory Cote (University of Guelph). *Utilitarian cops—Jeremy Bentham and the Rise of Modern Policing*

It is widely known that modern policing first came into existence in 1829, when London's Metropolitan Police was introduced under British Home Secretary Robert Peel. It is Peel who is often referred to as the founder of modern policing, an English institution that influenced North American police. But there is so much more to this story that remains unknown, involving the central role of Jeremy Bentham in the development of modern policing. Afterall, the first book on Bentham and policing was published in 2021, in which we are told students and scholars are ignorant of the philosopher's role in shaping this modern institution. In philosophy too, his broader role in the development of policing continues to be overlooked, until now. I argue Bentham was central to the development of modern policing and that his work in applying utilitarian philosophy to social and legal reforms influenced Peel's innovation. This paper takes a broader view of Bentham's overall efforts, that should take us beyond the work he did with Patrick Colquhoun on the Thames River Police of 1800. This, after all, only represented a small portion of Bentham's broader efforts over approximately fifty years. So this paper will not only involve a broader view, but also include a deeper dive into some of Bentham's other works, including *Indirect Legislation,* the significance of which may have been understated in a previous account on policing. John Stuart Mill wrote that Bentham was a great subversive. And as the great subversive who at times worked from behind the scenes so that others could play a leading role, his involvement in police reform is at times difficult to discern. This paper will attempt to uncover more of these efforts, by surveying the combined work of the founder of utilitarianism and his Benthamite followers. Foucault claimed Bentham was more important for understanding modern society than either Kant or Hegel.4 Without overstating Bentham's contribution to policing5, I claim he may also be more important for understanding modern police than either Peel or Colquhoun.

Michihiro Kaino (Doshisha University). *Bentham's Preventive Police and the Modern Japanese Society*

Bentham had a very interesting idea of 'indirect legislation'. In his article of 'indirect legislation', Bentham writes that '[i]n direct legislation, the evil is attacked in front: in indirect legislation, it is attacked by oblique methods.'

This paper will discuss Bentham's work of 'Preventive Police', which was published on line in 2018 as a new volume for *The collected Works of Jeremy Bentham*. And Bentham's plan of preventive police comprises two major indirect legislation. On the one hand, Bentham proposed that various occupations such as second hand stores should receive license and should be made to keep receipts and

records for inspection. This is intended to make it harder to dispose of stolen goods, which would in turn make offences of stealing less profitable. On the other hand, Bentham proposed to publish 'The Police Gazette' weekly, which would contain the news of crimes. Bentham thought that this would make it difficult for criminals to commit crimes in the same location or in the same manner by warning potential victims.

Many modern criminologists in Britain and in the United States pay huge attentions to Bentham's scheme of preventive police and they argue that Bentham's scheme is forerunner of contemporary criminology. For instance, in a volume of *Jeremy Bentham on Police: The Unknown Story and What it means for Criminology* (2021), Gloria Laycock argues that the modern approach of 'situational crime prevention' can be directly compared with Bentham's idea of indirect legislation and preventive police in that it depends on the fact that behaviors of potential criminals are determined by the situation within which they find themselves. Referring to the analysis of these modern criminologists, this paper will also explore the implications of Bentham's scheme of preventive police for the modern world, particularly for the modern Japanese society.

Rex W. Mixon (Stern Business School, New York University). *Ideas of Utility and Justice in Bentham and Horace*

In an unpublished manuscript entitled "Of the Principle of Utility," which is a draft of the first chapter of *An Introduction to the Principles of Morals and Legislation*, Bentham stated that with respect to "the Principle of Utility": "I had it from Epicurus, from Carneades, from Horace, from Helvetius, from Beccaria. All that remains for me is only to apply it to particular cases as they come under review." (UCL Bentham Papers, Box c, fo.114)

The proposed paper explores Bentham's self-identification with Horace as a source of the principle of utility with particular focus on the ideas of utility and justice in Bentham and Horace.

The paper proceeds by examining Bentham's engagement with Horace (65 – 8 BCE), a Roman poet of the late Roman Republic who developed in his poetry themes from Epicurus.

Next, the paper discusses passages in Bentham's writings that examine Horace's texts, particularly Bentham's analysis of Horace's declaration in *Satires* (1.3.98) that utility may be said to be the mother of justice and equity: "Utilitas justi prope mater et aequi."

The last section of the paper examines how justice relates to utility in Bentham with a close reading of passages in Bentham's texts, for example, *An Introduction to the Principles of Morals and Legislation* (1789), *A Protest Against Law-Taxes* (1795), *Article on Utilitarianism* (1829), and *Deontology*.

The paper concludes that Bentham and Horace adopted Epicurus' view that justice is not something in itself but exists in relations among men that are beneficial and useful for men not harming each other or being harmed.

Panel 6D—Denys Holland Lecture Theatre (SB31)

Thomas McDowell (University of Toronto). *The Inverted Panopticon: The Ontological Implications of Bentham's Self-Preference Principle*

This paper advances the claim that the mature Bentham's self-preference principle (SPP) caused him to break from his earlier ontological individualism towards a theory of social ontology. As a universal axiom that could explain behaviour at all times and places, the SPP unified individual and society in a single theory of psychological dynamics related to one's role in a structural and institutional context. Individuals are understood to shape their mode of being in and through their sensuous, concrete engagement with a community's social structures, institutions, customs, and categories. This led Bentham to develop an institutional/social theory of ontology, in which the self-making logic of the panopticon is inverted and applied to the whole of society, placing rulers under continued surveillance. Viewed from this perspective, Bentham's mature work is contiguous with the radical and critical traditions of the nineteenth and twentieth centuries.

Markéta Minářová (University of Prague). *Utilitarianism and Justice: Jeremy Bentham's concept*

The aim of my paper is to lay out the foundations of Jeremy Bentham's theory of utilitarianism with reference to the concept of justice and its influence on modern conceptions of justice. I will try to answer whether and why this position is fruitful or not and whether the criticisms against it are justified. The paper will focus on the founder of the movement, Jeremy Bentham, and the foundations of his utilitarian theory, which are the consequence principle, the utility principle, the hedonism principle and the social principle. Based on these theories, I will attempt to interpret Bentham's idea of justice.

Justice can be seen as a structured ethical principle or ideal that influences the actions of people or collectives and is manifested in the operation of social institutions and in social relations is the source of value judgments about the state of affairs. The notion of justice in utilitarianism appears (at least in Bentham) to be the exact opposite of established assumptions.

The paper will therefore focus on Jeremy Bentham's interpretation of utilitarianism and the concept of justice, identifying its influence on contemporary conceptions of justice. Therefore, the paper will also take into account the development of this principle itself, as the original utilitarians were radical and believed in the complete transformation of English society. Utilitarianism at the time was identified with a progressive and reform-minded political agenda, such as the extension of democracy or the provision of better living conditions. Contemporary utilitarianism, on the other hand, is perhaps too conformist and may seem to want to

be presented as a state of unchanging things with the aim of keeping things as they are. Whether this is indeed the case in the area of justice, I want to find out through my paper.

Daisuke Nakai (Kindai University). *J.M. Keynes, F.A. Hayek, and their Critique of Utilitarianism*

J. M. Keynes (1883-1946) and F. A. Hayek (1899-1992), both representing economists of the 20th century, put forth opposing claims, with the former advocating government intervention and the latter espousing liberalism. Despite their friendship, their economic views differed significantly, especially as Hayek consistently criticized the ideas of Keynes and macroeconomics. Such differences between Keynes and Hayek have often garnered attention from researchers. On the other hand, both Keynes and Hayek share a common ground in strongly opposing utilitarian philosophy. Therefore, using utilitarianism as a focal point, this report aims to highlight the similarities and differences in the economic thoughts of Keynes and Hayek. Furthermore, through their critique of utilitarianism, we intend to explore the merits and demerits of utilitarianism.

Panel 6E—Room LG17

Elżbieta Filipow (University of Warsaw). *Women's Intellectual Inspiration and their Influence on John Stuart Mill's Feminist Sensitivity*

In my presentation I am scrutinizing John Stuart Mill's intellectual inspirations, which include women's figures that were formative for his sensitivity to women's emancipation. I believe that the so-far biographies of John Stuart Mill do not exhaust the issue of women's intellectual influences on his feministic thought. The exception might be a book by Nicholas Capaldi. It discusses some women that were important in Mill's life: Mrs Bentham (Sir Samuel Bentham's wife) and Sarah Taylor Austin (John Austin's wife). Women described by Capaldi, apart from Harriet, are but 'a pale background' and, actually, not much can be learnt about their real influence on the philosopher's life.

Additionally, I am interested in the extent to which Mill was familiarized with the literature written by women of the 19th century. Margaret Walters notices that in his *Subjection of Women* Mill does see a deep injustice for women in their access to education, which did not make them be authors of original, 'great and luminous ideas,' nor did not let them create 'literature of their own.' Walters, however, shows some shortcomings of his thinking and stresses that 'Ann Radcliffe, Fanny Burney, Jane Austen, Susan Ferrier, Brontë sisters: they all seem to have escaped his notice.' What proof is there for those words? Did Mill really know their works but perceived them as unoriginal literature or he actually did not know them? I am also interested in discovering if Mill got familiarized 'on the go' with the works by Victorian feminist.

Naturally, it is Harriet Taylor Mill who remains the most important figure here. Their long-standing relation, topped out by their marriage, seems to suggest that she was the most formative figure in shaping J.S. Mill's sensitivity to the issues of men and women's equal rights. Presentation will be based on the results of my research query in the Somerville College Library at the University of Oxford, and it is hoped to fill in the gap in my so-far research as well to sketch a multidimensional intellectual background of John Stuart Mill's liberal feminism.

Emmanuelle de Champs (Cergy Paris). *Utility, Community and Women's emancipation: the intellectual journeys of Anna Doyle Wheeler and Frances Wright*

This presentation looks at the intellectual formation of two early advocates of women's emancipation: Anna Doyle Wheeler (1785?-1848) and Frances Wright (1795-1852). In 1820s London, both became personally acquainted with the persons, circles and ideas of two prominent reformers of the period, Jeremy Bentham and Robert Owen. Independently of one another, Wheeler and Wright wrote and lectured during the 1820s, demanding social and political reform and defending women's rights to emancipation. They became respected figures in the early British socialist movement, which has led scholars to highlight the Owenite and socialist roots of their thinking. In contrast, this presentation looks at the political and social culture of the 1820s and explains how utilitarianism also shaped their views and contributed to the construction of two strikingly innovative positions in favour of gender equality. This presentation will engage with the arguments of Catherine Villanueva Gardner (Empowerment and Interconnectivity: Toward a Feminist History of Utilitarian Philosophy, 2013) and examine the legacies of this period in the feminist and utilitarian arguments of Harriet Taylor, John Stuart Mill and Henry Sidgwick.

Allison Dube (Mount Royal University). *Implications of the Scholar-Tutor principle in Chrestomathia for Universities Today*

Most results netted from a search of the term "scholar-teacher" reverse the terms, and reference "scholarly teaching." This was not what Bentham had in mind with the "Scholar-Tutor principle" in *Chrestomathia,* which "consists in employing, as teachers to the rest, some of the most advanced, and in other respects most capable, among the scholars themselves." Included in the advantages gained are: "Saving in money. Every professional teacher would need to be paid; no such scholar-teacher needs to be, or is paid;" [students are paid with honour and experience]; an "Increase in relative aptitude" [Bentham feels that students will often do a better job than a "grown-up Under-Master"]; and, "By teaching others, the scholar is, at the same time, teaching himself: imprinting, more and more deeply, into his own mind, whatsoever ideas he has received into it in the character of a learner." The "application of this principle is, therefore… an essential feature" of the Chrestomathic School.

Implications of these sections of *Chrestomathia* for universities will be explored. For one thing, can universities today afford *not* to save money, increase relative aptitude of some of their instructors, and help students to learn things more deeply? For another, many universities encourage faculty to incorporate "undergraduate research" as a "high impact practice. The ideal is that the "student's work should be original and make a contribution to the student's discipline." Yet supporting evidence is often peppered with the term "self-reported"—as in "those involved in undergraduate research self-reported gains in… the personal and professional realm.'" One question, then, do undergraduate research programs

actually enrich the disciplines, or do they encourage students to *feel* they have? Another, could experience gained from undergraduate teaching be equally satisfying, *and* serve as better experience for the transition to the working world?

These and other issues, including whether there is an equivalent to the "grown-up Under-Masters" that Bentham often criticizes at universities today, will be explored.

Panel 7A—Moot Court (XG03)

Sidgwick and Evolution

Henry Sidgwick paid attention to Charles Darwin's ideas on evolution and had some contact with him. He published an essay on 'The Theory of Evolution in its Application to Practice' in the first volume of the journal *Mind*, in 1876. Still, he rejected the claim that evolutionary theory can give a plausible account of morality. In particular, in the *Methods* he claimed that any view of the origins of our ethical ideas and judgments is irrelevant to the issue of their meaning and truth (ME, 212-3). This panel includes three papers on Sidgwick's view of the relevance of evolutionary theory for ethics, considering historical and theoretical aspects.

Giulia Cantamessi (University of Pavia). *Sidgwick and Spencer on Scientific Ethics, Evolution, and Utilitarianism*

In this paper I shall analyse and compare crucial aspects of the ethical methods and views of Henry Sidgwick and Herbert Spencer. This comparison enables to highlight certain significant differences in these philosophers' ethical thought within a common utilitarian and hedonistic framework, as well as to understand Sidgwick's engagement with the evolutionist ethics of his time.

I shall first of all examine what Sidgwick and Spencer mean when they describe their ethical inquiries as "scientific". Sidgwick characterises ethical aims and methods as scientific insofar as ethics seeks to provide systematic and precise knowledge of what ought to be. Spencer's aim is to provide a scientific foundation of ethics on the basis of the theory of evolution, and to develop a form of "rational utilitarianism" which can ascertain necessary relations between actions and consequences, rather than relying on empirical estimations. I shall highlight the different assumptions and implications of these conceptions of scientific ethics. In particular, when addressing Sidgwick's objections to Spencer's methodology, I shall point out (i) how according to Sidgwick Spencer's method fails to live up to scientific standards and to his own intentions, and (ii) the importance Sidgwick attaches to the feasibility and the practical guidance of ethical theories.

Both Sidgwick and Spencer also acknowledge the felicific influence of evolution on ordinary morality and moral intuitions, sentiments and habits. In the last section of the paper, I shall turn to how these authors differently describe such influence and examine which argumentative role the appeal to evolution does actually play in their conception of common-sense intuitions as unconsciously utilitarian – a conception which, as I shall show, leads to the development of two different forms of indirect utilitarianism.

Gianfranco Pellegrino (LUISS). *Sidgwick on common sense morality. Evolutionary debunking or esoteric utilitarian foundation?*

In *The Point of View of the Universe. Sidgwick and Contemporary Ethics* (2014), Katarzyna de Lazari-Radek and Peter Singer consider Sidgwick's rejection of an evolutionary debunking of moral realism. They defend two claims. First, Sidgwick's utilitarian explanation of common sense morality could be read along the lines of an evolutionary debunking. A "contemporary Sidgwick", i.e. one who "shared our modern scientific understanding" of evolutionary theory "might be closer" to evolutionary debunking of the kind defended by Sharon Street "than the historical one." (181) Second, Sidgwick axiom of rational benevolence is immune from evolutionary debunking since it is self-evident and no plausible evolutionary account of its origin has been presented.

In this paper, I challenge both claims. First, I show that Sidgwick's analysis of common- sense morality cannot be likened to an instance of evolutionary debunking it. Sidgwick claims that adherence to most common-sense moral rules has good utilitarian effects, and so utilitarianism can be esoteric in most of the cases covered by common-sense morality. I call this an 'esoteric utilitarian foundation' of common sense morality and show that the way it works is different from the evolutionary debunking of moral realism.

Second, I show that, given the very possibility of esoteric morality and of a self-effacing ethics, Sidgwick's moral realism can be immune from the evolutionary debunking in a stronger way. This paper is close to the spirit of de Lazari-Radek and Singer's view of Sidgwick's moral realism and its superiority as compared to evolution-based moral anti-realism. Still, it puts forward a different interpretation of Sidgwick's approach to common-sense morality and of his moral realism.

Leonardo Ursillo (LUISS). *Sidgwick and Darwin*

This paper aims to clarify the personal and theoretical links between Henry Sidgwick and Charles Darwin. In 1874, Sidgwick sent Darwin a copy of his *Methods of Ethics*. Darwin replied in a letter dated 19th December 1874, saying that "it is a subject that interests me deeply". During his undergraduate studies at Cambridge, Sidgwick had read *The Origin of Species* and remained an avid reader of Darwin's works, including *The Descent of Man* (1871) and *The Expression of Emotion* (1872). In 1872, he published a review defending Darwin's work, and in 1876, he published a famous article in which he tried to compare utilitarianism with evolutionism. Darwin read and discussed this article in a letter dated 26th April 1876, disagreeing with some of Sidgwick's ideas. Despite Sidgwick's belief that Darwin's ideas could not address our ethical dilemmas, he dedicated some time to studying Darwin's work. There are still questions that need to be answered regarding Sidgwick's understanding of evolution. To clarify the relationship between the ideas of these two authors, this paper will compare some parts of their two main works, *The Descent of Man* and *The Methods of Ethics*, to illustrate the views of Sidgwick on the evolution and the origin of morality. Moreover, I will investigate Sidgwick's knowledge of evolution to uncover his

ideas about the subject. By analyzing Sidgwick's texts, we can figure out what he thought of Herbert Spencer's Lamarkism and Darwin's natural selection theory.

Panel 7B—Hong Kong Alumni Room

Michael Schefczyk (Karlsruhe Institute of Technology). *The significance of the past for consequentialists*

Consequentialists deal with the past relatively rarely, and it seems obvious enough why. Knowledge about the past may have some value for its own sake; it might also have some instrumental utility from time to time. Unlike the open future, however, the past is fixed and cannot be improved upon. It is the future, and only the future, that depends on our decisions here and now. The future must therefore be the main concern of consequentialists. But what if the past were not fixed? If it were possible to change the past, we could certainly change it in better and worse ways. And we might even be able to change it for better or worse. But then it would no longer be so clear that consequentialists should be indifferent to the past.

This talk will look at attempts to make sense of the idea of a changeable past and ask what significance it has for consequentialists.

Sam Fremantle (International School of Philosophy). *Reciprocity and Posterity*

John Rawls has been interpreted as putting forward a theory of 'Justice as Reciprocity' which has an implication that many would find deeply problematic; his conception of reciprocity implies that we have no duty of justice to those who cannot reciprocate any sacrifices we may make for them. On this interpretation, those excluded from the scope of justice would include the disabled, infants, and future generations. An alternative conception of reciprocity was briefly canvassed by John Stuart Mill in a parliamentary speech to Parliament. In that speech, he argued that we have duty to make sacrifices for posterity that is, to all intents and purposes, a duty of reciprocity. Because we have greatly benefited from those we can never directly repay, so we have an obligation to bestow benefits on those who can never repay us. Furthermore, he argues, those repayments should be made with interest.

Such a conception of reciprocity, if viable, would avoid the unpleasant implications of Rawls's conception of reciprocity mentioned above. An obligation of reciprocity to posterity could cover, at least to a degree, the disabled, infants, and future generations.

However, establishing the viability of this 'Millian' conception of reciprocity, is a major challenge. At the heart of the Rawlsian conception of reciprocity is the plausible idea that we only owe duties of reciprocity to those to whom we need to reciprocate in pursuit of mutual advantage.

In this paper, I will put forward the case for the superiority of the Millian conception of reciprocity over the Rawlsian one. If this case is successful, it would follow that, as Rawls strenuously denied, utilitarianism is fully compatible, if not

mandated, by justice as reciprocity. My case will draw primarily on a comparison of Alan Gibbard's interpretation of Rawls' conception of Justice as Reciprocity with a fleshing out of Mill's conception of justice as reciprocity as revealed in his parliamentary speech on The Malt Duty.

Yunfeng Jin (Jilin University). *On the Distributive Justice Principle of Institutional Utilitarianism*

Distributive justice is an important issue in contemporary political philosophy. Critics argue that the main idea of utilitarian justice is, If a society's main institutions are arranged to achieve the maximum net balance of satisfaction formed by all individuals, then the society is correctly organized and therefore just. Utilitarianism only focuses on the maximization of utility, and does not pay attention to the issue of distribution between people. However, equality is our intuitive requirement. If we only consider the maximization of utility without considering the issue of distributive justice, it will ask the state to allow serious inequality to exist for the sake of maximizing utility. Critics believe that utilitarianism cannot respond to the above criticisms, and it cannot well accommodate the requirements of distributive justice within its theoretical framework, which make it difficult for utilitarian political philosophy to have enough persuasiveness.

An important representative of contemporary utilitarianism is institutional utilitarianism who attempts to construct a reasonable principle of distributive justice on the basis of maximizing utility, and to make this principle of distributive justice compatible with our intuitive requirements for maximizing utility. If this defense is valid, then utilitarian political philosophy will not be refuted due to the lack of reasonable principles of distributive justice.

Panel 7C—Gideon Schreier Lecture Theatre (124)

Stephen G. Engelmann (University of Illinois at Chicago). *Sidgwick on the Old Radical*

In a detailed *Fortnightly Review* portrait, Henry Sidgwick is dismissive of Jeremy Bentham's late work: "all that Bentham writes after 1817 is full of…heated and violent democratic fanaticism (1877)." He diagnoses Bentham's radicalism as the fruit of bitterness, pointing in particular to the failure of Panopticon. As a description of Bentham's political journey Sidgwick is on firm ground; his analysis fits Philip Schofield's (2006) account of the discovery of sinister interest. The diagnosis and the essay as a whole display Sidgwick's intimate familiarity with his predecessor's life and work.

"Bentham and Benthamism" culminates, however, in a consideration of controversies over the posthumous Bowring-edited *Deontology*; in the end the piece is illustrative of Sidgwick's own abiding moral concerns, in particular the question of interest's connection to duty. Sidgwick was, like Bentham and J.S. Mill before him, a polymath who wrote on a range of subjects. But unlike his predecessors he was a transitional figure toward moral philosophy as we know it, i.e., toward a concern above all with right individual conduct and its grounds. Mill and especially Bentham were by contrast theorists of the art and science of government. This doesn't mean that a gap between interest and duty was unimportant; on the contrary it was a matter of vital importance. But whereas joining interest and duty was for Sidgwick primarily a *philosophical* and pedagogical problem it was for Bentham a *political* and legislative problem.

My paper explores this difference in approach to account for Sidgwick's comparative conservatism (and, among other things, his racism), and to reflect on its legacy for the utilitarian tradition. Sidgwick's contempt for Bentham's class warfare is not only a product of his different era, of what Hannah Arendt (in her *Imperialism*) called "the political emancipation of the bourgeoisie;" it is intimately connected to his preoccupations with the interior of the self, with the point of view of the universe, and with the connections between them. On the way to making this argument I engage both primary sources and recent secondary controversies and I link them, more provocatively, to problems of utilitarian politics today.

Malcolm Quinn (University of the Arts London). *To the Death: Fortitude and Judgment in Bentham's Deontology*

This paper discusses Bentham's writing on fortitude, part of a sustained period of work on logic and language that Bentham undertook between 1813 and 1815. My analysis shows how Bentham works between two vantage points. The first decides whether or not someone acted the part of a person of fortitude. The second emerges when the first vantage point is seen to fail to deliver a judgment on fortitude. If logic

is allowed to 'take the command and give direction to the course of Ethics itself' a new standpoint for fortitude is revealed. The fictitious entity of fortitude is situated in relation to the real entities of pleasure and pain, allowing us to follow a judgment 'to the death'.

Bentham begins by asking how someone can be judged to be a person of fortitude. He argues that asking this question can lead us astray, first, because it can direct someone to act against their own interest, second, because it can be a way of avoiding exposure to actual pain or danger. If a judgment on whether someone is a person of fortitude is 'a mere question of words', then, paradoxically, 'words are everything' because, if a person is seen to have justice on their side, they can play the part of a person of fortitude without any act of fortitude taking place. Bentham suggests, instead, that we can follow our judgments on fortitude 'to the death', for example in cases where people accept death in defence of their personal liberty, as the price that must be paid for a benefit to oneself or to others. He goes further, arguing that, in contrast with the established church that prohibits suicide, the doctrines of Jesus ask us to encounter the reality of actual pain or danger, up to and including a self-willed death.

Yanxiang Zhang (North China University of Technology). *J.S. Mill on Jeremy Bentham's Metaphysics*

J.S. Mill offered an evaluation of Bentham's philosophy to the effect that on metaphysics Bentham's performance was a complete failure, although he acknowledged that Bentham achieved an extraordinary success in practical fields. This article argues that Mill's reading of Bentham is a travesty, that the volume of Bentham's works he read was in fact very limited, and his interpretation superficial. Metaphysics was a major concern for Bentham throughout his intellectual life, and his writings on metaphysics are voluminous. Bentham's philosophy is a reflective and holistic one, deeply metaphysical, multi-levelled, and comprehensive.

Panel 7E—Room LG17

Shinji Nohara (University of Tokyo). *Adam Smith's criticism of Hume's utilitarian morality*

Scholars have researched Adam Smith's criticism of David Hume's utilitarian view of justice. They have also argued that Hume insisted that people understand the common interest of justice, such as private property, and that Smith criticized this view because it was based on a utilitarian stance in which people's shared interest or utility is the foundation of morality. On the other hand, they have also investigated that Smith' regarded justice as based on his anti-utilitarian morality, which depends not on people's sense of utility but on their concern about individuals and their particular accidents.

What has yet to be fully studied is that Smith did not deny utilitarian stance at all. Although Smith was skeptical of utilitarian justice, this does not mean that Smith negated the utilitarian viewpoint. Rather, he sometimes adopted a utilitarian standpoint. Whereas his view of justice is anti-utilitarian, he saw as some aspects of justice as relying on the sense of utility.

As I argue in this presentation, Smith insisted that people's sense of justice included that of utility because people affirm justice not only because justice is derived from natural resentment but also because they see justice as leading to social utility. Justice as based on natural resentment might make people attack who was seen as a perpetrator. This could make social order in danger. He then affirmed justice because it included the utilitarian viewpoint; people's sense of justice included the effects of people's behavior on society. This utilitarian viewpoint made Smith build his theory of justice that is compatible with the social necessity maintaining social order.

This point is worth examining because his utilitarian aspect of justice incorporated anti-utilitarian criticism of utility. Smith especially insisted that justice should be the basis of society but did not think that people could understand the common interest of justice. He could build his theory of justice, refuting the fictitious supposition of people's agreement on social utility.

Frederic R. Kellogg (The George Washington University). *The universe has no point of view: John Dewey's Naturalized Utilitarianism*

Instead of the global universal maximization principle of classical utilitarianism that Henry Sidgwick called "the point of view of the universe," John Dewey viewed welfare as guided by a local, problem oriented, agonistic preference adjustment principle. Dewey recognizes multiple possible observation points relating to a vast number of specific social problems, each also (at any given moment) coming in a

succession of possible viewpoints, in which considerations of utility are transformed in stages by disputatious inquiry. Rather than utilitarianism's synchronic question "What does the universe (at some given moment) recognize as the greatest happiness of the greatest number?" Dewey's diachronic question is the granular inquiry, "What might the universe recognize as the optimum utility balance in each of its ongoing problems when they are resolved?" The universe cannot know or often guess the outcomes, as current (and opposing) preferences and utilities must be revised for specific problems to be resolved.

I argue first that Henry Sidgwick's *Methods of Ethics* (1874) contains a deep inconsistency in defending the synchronic omniscient viewpoint strictly as a corrective to common-sense morality; second, Dewey's naturalized account of the origin of rules reveals the defects of Sidgwick's synchronic rule consequentialism, and resolves Brad Hooker's rule-extension problem; and third, game theory, which Harsanyi and others have employed to account for change within a classical utilitarian context, assumes preferences as rigid over time, whereas they *must adjust* for a natural and workable utilitarian world. Moreover, inclusive input is necessary to avoid the complaint of Bernard Williams that classical utilitarianism implies an authoritarian world. Underlying Williams's concern is Kenneth Arrow's 1951 General Possibility Theorem, a famous (and unrefuted) demonstration that only dictatorship can order any domain of diverse individual preferences.

Jonas Harney (Saarland University). *Taking narrow person-affecting considerations seriously*

According to one version of the Narrow Person-Affecting View, the extent to which an outcome is better (or worse) for an individual makes that outcome in one respect better (or worse). This is deeply plausible but, it is claimed, implies intransitive rankings of outcomes and, assuming transitivity of better-than, contradictions. I tackle the objection by considering how narrow person-affecting considerations influence the goodness of outcomes in multiple- outcome comparisons.

I distinguish two construals. According to the Pairwise View, narrow person-affecting considerations relative to two outcomes *A* and *B* are essentially pairwise and, thus, make *A* better (or worse) than *B* but neither of the two outcomes better (or worse) than other outcomes. According to the Set-Wise View, the considerations are set-wise; they function as features of the outcomes that make *A* and *B* better (or worse) than any other outcome *C* withina fixed set of compared outcomes.

I argue that, by contrast to the Pairwise View, the Set-Wise View rebuts the objection from intransitivity. It provides a set-relative way to rank multiple outcomes that preserves transitivity in fixed sets of outcomes: We assign set-relative values to the outcomes inproportion to the narrow person-affecting considerations; the outcomes are ranked according to their aggregated set-relative values represented by rational numbers; since "has a higher rational number" is a transitive relation, this guarantees the transitivity of better-than within a fixed set. The Set-Wise View

therefore avoid cyclical assessments and cyclical obligations within fixed sets and reduces the severity of money pumps.

I defend the Set-Wise View against objections. Comparisons of two outcomes can depend on further outcomes; but the outcomes can be shown to be relevant. The Set-Wise View seems to have absurd implications which, however, are perfectly acceptable if we take narrow person-affecting considerations seriously.

Panel 8A—Moot Court (XG03)

Dorina Patrunsu (University of Bucharest). *Is there any common sense morality? Egoistic vs. universalistic hedonism in Henry Sidgwick's The Methods of Ethics*

What I ought to do? is a question as striking in moral philosophy today as in Henry Sidgwick's time and by no means less controversial. Likewise, the assumption that the difficulty of answering this morally fundamental question would be mitigated if common sense were involved is as rhetorically pervasive and ethically significant as contested. However, ordinary folk, as well as moral philosophers, disagree about (i) what common sense is and, therefore, (ii) what actions and policies raise concerns about common sense morality. As a result, most of the philosophical works on Sidgwick's common sense morality-based utilitarianism, with some notable exceptions, revolve around either dualism or discrepancy between egoistic and universalistic hedonism.

In this paper, I aim to show that this is not a case of defeating universalistic hedonism. I therefore argue that the "profoundest problem in ethics" is not missed in Sidgwick's ethics, despite the failure claimed by himself and others.

There are two steps in arguing for this. First, I map out the logical space in which different conceptions of common sense are located, thereby clarifying the structure of the most salient contemporary views on Sidgwick's common sense morality.

Second, I highlight some shortcomings of the dualist interpretation, identifying the most prominent contemporary arguments against it. Their fundamental claim is that it is unnecessary to reject egoism-based hedonism for making a case favoring universalistic hedonism. So, the main objection to utilitarianism's "inevitable failure" is that the conflict between self-interest and morality is to be resolved, not dissolved.

In particular, following up on some dualism cons arguments (Parfit, Brandt, Lazari-Radek & Singer), I defend that there is no logical inconsistency between self-love and benevolence. The leading assumption in preserving the "harmony" between egoism and utilitarianism is that morally and socially correct actions matter because all individuals matter, whether selfish or not. This implies regaining sight of the distinctively ethical purposes of utilitarianism, given that no ethical requirement claimed by one can be greater than the one claimed by all, and any requirement imposed on others cannot exceed mine. The double-normative purposed argument for universalistic hedonism retains the key virtues of utilitarian conceptions while shedding their potential vices. It also allows for a stronger case favoring utilitarianism as the "ultimate moral" thing to do.

Don Habibi (University of North Carolina at Wilmington). *Sidgwick on subscription, secularism, and the problem of illiberalism*

Henry Sidgwick distinguished himself as the consummate scholar of utilitarian ethics and liberalism—the worthy heir to J.S. Mill. I argue Sidgwick also lived as a moral exemplar—making him a worthy heir to Kant. The most serious problem in his personal life, which rattled his conscience throughout the 1860's, revolved around the requirement of subscription. In Sidgwick's case, Cambridge University compelled him to subscribe (i.e., pledge allegiance) to the 39 Articles of the Church of England. Prior to 1860, Sidgwick had less problem with this loyalty oath. He was a believing Christian and even saw value in subscription. Yet, with exposure to secular ideas, science, and eventually, the influence of Darwinian evolution, he began to doubt several Articles. Sidgwick took his pledges seriously. He consulted respected thinkers, including Mill, who advised him to handle this dilemma in an expedient, consequentialist way. However, the advice of Mill, T.H. Green, and William Whewell, seemed to Sidgwick's moral compass a 'pious fraud.' Thus, Sidgwick's commitment to honesty, integrity, and freedom of conscience led him to refuse subscription—costing him his professorship at Cambridge.

Next, I argue that the pendulum has shifted. Eventually, the secular turn led to significant changes. Cambridge dropped the religious loyalty oath, as did most Universities in Europe and beyond. Religious perspectives became more tolerant, and liberal universities thrived in an atmosphere of free and open inquiry—marketplaces of ideas.

Unfortunately, in recent decades, the pendulum has shifted back to intolerance. Zealous, secular true-believers have approximated religious conviction, including pressures resembling subscription. Freedom of speech and inquiry have given way to compelled speech, ideological conformity, self-censorship, and fear of offending. Many universities expect student applicants and faculty jobseekers to submit statements professing their commitment to particular causes of social justice. Subscription has made a comeback. I conclude by discussing what the legacy of Sidgwick's dilemma teaches us.

Anthony Skelton (Western University). *On the Irrelevance of Sidgwick's Political Philosophy*

Henry Sidgwick published *The Elements of Politics* in 1891. In it, he aimed to clarify the utilitarian approach to practical politics. This is achieved, he thinks, through identifying the subordinate "principle or principles . . . capable of more precise application, relating to the means for attaining by legislation [and institutions] the end of Maximum Happiness". *The Elements* attempts to develop principles relating to property rights, contracts, conduct in war, international relations, among other topics.

While Sidgwick was working on the book, he worried about whether it was worth the effort. In a letter to John Addington Symonds in 1887 he wrote

> I am trying to absorb myself in my Opus Magnum on *Politics*. My position is that I seem to myself now to have grasped and analyzed adequately the only possible method of dealing systematically with political problems; but my deep conviction is that it can yield as yet little fruit of practical utility -- so doubt whether it is worth while to work it out in a book.

He appears to have been right to worry. *The Elements* simply does not register in contemporary discussions of political philosophy. The fortunes of his work in politics stand in marked contrast to the fortunes of his work in ethics.

In a recent article, David Miller reflects on why Sidgwick's political philosophy is ignored. I agree with Miller that this work is ignored. However, in this paper I argue he is wrong about why it is ignored. First, I argue *pace* Miller it is not true that Sidgwick's politics is ignored because (unlike some contemporary utilitarians) Sidgwick is too uncritical of existing expectations. Second, I argue that the failure to register cannot *contra* Miller be traced to Sidgwick's moral epistemology as Miller interprets it, since Miller's interpretation of it is mistaken.

Panel 8B—Hong Kong Alumni Room (221)

Dennis Pirdzuns (University of Manchester). *Supporting Singer's Seminal Points on Poverty*

Peter Singer got it right about poverty when he stated succinctly that: "suffering and death from lack of food, shelter, and medical care are bad. (…) Those who disagree need read no further." (1973, 231). Poverty implies a lack of important goods, which causes intense suffering, so everyone capable should provide relief, regardless of additional requirements.

While Singer's seminal essay 'Famine, Affluence, and Morality' is highly significant for the discourse on global justice, and his arguments for a cosmopolitan position are widely cited, he is also criticised for the political conclusions he draws from his analysis, mostly those around effective altruism. Less attention is paid to the criticism of Singer's conception of poverty and his normative claims to minimise deprivation, missing potential conclusions on utilitarian approaches to morality and justice. This inquiry revisits his arguments, reviews the criticism, and takes an affirmative stance in support of Singer's points about poverty.

Four points in particular deserve more attention and debate. The first is the conceptual claim that poverty is defined by deprivation as a lack of goods, which critics say ignores poverty's relational aspects; yet these can easily be incorporated. Secondly, the normative claim that generally everyone capable has a responsibility to provide relief, is criticised, but inconclusively, for neglecting individually incurred responsibilities. The third point worth revisiting is that justice demands to minimise deprivation and that goods held by those without urgent needs are an injustice; a position prominently endorsed by sufficientarian and limitarian accounts in contemporary debates on justice. Finally, the underlying value judgement for alleviating poverty is that it causes intense suffering; something Singer understandably takes for granted, yet which receives closer examination in recent works on ill-being, needs and the normative significance of suffering.

Christoph Lumer (University of Siena). *A Theory of Moral Duties: Progressive Norm Welfarism*

The contribution develops a new welfarist theory of moral duties, called "progressive real norm welfarism". The situation in the deontic part of welfare ethics is clearly more unsatisfactory than in the axiological part, not only because of the multitude of substantive problems, but also because the confusingly many new proposals usually only attempt to solve a small part of these problems and ignore the others. (1) The approach developed here therefore first collects the known problems of important conceptions of moral duties, above all of act utilitarianism. However, it also emphasises a problem that is usually ignored, namely the lack of bindingness in the sense of motivational pressure to comply with obligations. (2) From this list of

problems, adequacy conditions for an adequate welfarist theory of moral duties are then synthesised and additionally justified instrumentally, including rationality of compliance, minimal equity, limited moral optimisation, bindingness. (3) The constructive approach begins with the problem of bindingness. I argue that the most promising ontologically specified instrument for solving this problem are real, socially valid norms, in the sense of largely generally followed and sanctioned practices. In addition to formal, legal norms, informal norms must also be included, which are only supported by informal sanctions. Moral duty is then the observance of morally good, socially valid norms. (4) This approach already fulfils many of the adequacy conditions, but two problems remain: that of maintaining morally good norms and that of insufficient optimisation. The maintenance of good norms can itself be partly ensured by duties, in particular the duties of public officials. The rest must be guaranteed by a new type of moral behaviour: actions of norm support (informing, monitoring, punishing...), which in essence are only driven by the moral motives of the subjects. (5) The problem of insufficient optimisation is addressed in a theory of moral progress, which is concerned with morally improving the stock of socially valid norms. This part of the theory also relies on actions of moral-political commitment that are not morally obligatory but are essentially driven by one's own moral motives.

Satoshi Yamazaki (Kochi University). *Henry Sidgwick as the unsung leader of the old welfare economics*

For instance, Hutchison (1953) recalls that Sidgwick was the last great moral philosopher who did much for economics in Britain. Likewise, Bonner (1995), Myint (1948), and O'Donnell (1979) refer to Sidgwick's theoretical contribution toward the formation of the old welfare economics. Following J.S. Mill's analysis, Sidgwick has introduced a distinction between 'science' (what is) and 'art' (what ought to be) in economic thinking, and argued that the latter should subsume, as well as efficient production, just distribution principles allocating social resources among the member in order to maximize social welfare, which is considered to have rendered great hints for the formation of Pigou's welfare economics. Thus, those preceding studies point out that many of Pigou's essential concepts in welfare economics including its ethical foundation are to be attributed to Sidgwick's works. By contrast, however, through a detailed comparison of the two, I (Yamazaki (2011), unfortunately in Japanese) have presented several major differences such as Pigou's negation of hedonism and egoism in Sidgwick, and offered the following (somewhat bold) proposition: while Sidgwick holds egoism with the intention of protecting the individual against the whole, he does not seem to pay particular attention to the protection of the individual when it comes to maximization of total welfare. On the other hand, although Pigou rejects the principle of egoism, he places great emphasis on protecting the individual even in the context of social welfare maximization. The key to explain that, as I have indicated, is thought to be the notion of so-called 'basic needs'. Indeed, no doubt, it may be a tough problem how to consistently incorporate the need concept into the utilitarian economics framework. Still, spending a considerable space in *The*

Methods of Ethics and *The Principles of Political Economy*, Sidgwick does discuss Justice far more than Pigou. It about time we should explore his further aspects, not necessarily inherited by Pigou, which can be counted as new contributions to welfare economics and welfare state.

Panel 8C—Gideon Schreier Lecture Theatre (124)

Anne Brunon-Ernst (Paris-Panthéon-Assas). *Beyond anti-Imperialism and Pro-Colonialism—Bentham on Waste Lands*

Supported by the recent edition of *Bentham's Writings on Australia* (Bentham 2022), and its companion *Bentham and Australia* (Causer 2022), Bentham's writings on colonies haveattracted renewed interest triggering critical reassessment. The literature (Winch 1997, Boralevi1984) stressed the discrepancy, in his earlier writings, between Bentham's critical stance towards Western powers' colonial expansion in "Emancipate their Colonies!" (1830) and "RidYourselves of Ultramaria" (1995), and his later support of colonialism in *Colonization Company Proposal* (1831). Lately, Pitt and Arneil have offered a new take on this mixed discrepancy (Pitt 2011; Arneil, 2021). Thanks to Arneil's definition of colonialism as a productive form of management which turns idle paupers and waste lands into profitable ventures (2021), she shows that Bentham could be pro-colonies (domestic and settler alike) andanti-imperialist (criticising French and Spanish Empires) at the same time. The aim of the paperis to unpack this tentative distinction between colonialism and imperialism by adding Bentham's discussion on penal colonies to Arneil's typology. Indeed, Bentham's criticism of penal colonialism conflicts with his support for domestic and settler colonialism.

The paper argues that Arneil's notion of "waste land" in her definition of colonialism is key tounderstanding Bentham's apparent contradiction. Bentham praises domestic over penal colonisation, on grounds that the latter operates as a diversion of useful workforce who could labour on waste lands at home (as implemented by the former). The presence of waste land seems to be pivotal to his criticism of penal transportation. This contention is explored in his other writings. While in *Pauper Management Improved* (Bentham 2007) and *Colonization Company Proposal*, Bentham justifies domestic colonisation and settler colonialism respectively in terms of putting wasteland to use, this is not the argument he uses in relation toimperialism. In "Emancipate their Colonies!" and "Rid Yourselves of Ultramaria", the Frenchand Spanish overseas possession do not add to the aggregate amount of land turned to profit asland is not considered as waste, but rather as possessions to be defended at a cost, or as propertythat is owned and/or turned to profit thanks to produce grown, transformed and traded.Moreover, Bentham's project of domestic colonisation needs to be read in a continuum with his promotion of settler colonialism. Indeed, he advocates for settler colonialism as a final stageof reclaiming wastelands, as once all vacant lands have been populated and cultivated at home(thanks in part to domestic colonies), investors, middle class and pauper couples will be encouraged to turn to colonisation overseas.

Eleonora Buono (Lausanne University). *National character, government and civilization in James Mill's Thought*

What was the relationship between national character, government and civilization for James Mill? This article answers this question by focusing on Indian and British national characters. I explain how, for Mill, Indian national character was unsuited to trigger the progress of civilisation. Mill thus questioned how a society with a flawed character could be led towards improvement. This article underlines the importance of human agency for the progress of civilization in Mill's thought. In order to cause progress, individuals had to voluntarily guide society towards improvement by embracing the principle of utility. Governmental action should thus create the conditions necessary to bring about individuals of suitable character. The members of the middle class could foster progress better than the members of any other class. They could oppose the despotism of the aristocracy and be the driving force of civilisation by voluntarily shaping their ideas in accordance with the public good.

Georgios Varouxakis (Queen Mary University of London). *The imperialist liberal and the anti-imperialist liberticide—the international visions of John Stuart Mill and August Comte Reviewed*

There is an interesting paradox in comparing the visions of global order defended by John Stuart Mill and Auguste Comte. The two of them exchanged some intense correspondence for a handful of years and read some of each other's works. Although interpretations differ regarding the degree of influence that Comte may have had on Mill's thinking, there were certainly some important synergies and there were things in Comte's thought that the British thinker admired. Meanwhile, though their eventual disagreements were related to other matters primarily, to the extent that they did discuss the future of international organisation, they did not see eye to eye.

The paper proposes, in the first place, to compare the respective international visions of the two thinkers as they emerge from their correspondence with each other. Following the examination of the evidence from the Mill – Comte correspondence, as well as from Mill's comments on Comte's vision in "*Auguste Comte and Positivism*" (1865) and in correspondence with others, the paper will proceed to compare the respective international order visions of the two thinkers as they emerge from the rest of the work of each of them. Such an examination yields an interesting paradox in the light of recent historiographical trends. A concentration on the (undeniable) ills of past imperialism has led to categorisations of nineteenth-century thinkers along binary lines of imperialist versus anti-imperialist, and concomitant normative moral judgments, comparing the "anti-imperialists" favourably with the "imperialists". The paper wishes to complicate that picture by highlighting the unpalatable aspects of Comte's international organisation scheme, and comparing these latter with Mill's proposals for open-ended diversity and progress. The intention is not to defend Mill's

"imperialism" nor to underestimate Comte's "anti-imperialism" but rather to complicate – and hopefully enrich – our categories in assessing past thinkers and activists.

Panel 8D—Room LG17

Joel van Fossen (Hosei University). *Sidgwick, Rationality, and The Dualism of Practical Reason*

The most dramatic section of Henry Sidgwick's *the Methods of Ethics* comes at the end in his description of the dualism of practical reason (*ME* 497-8). However, Sidgwick has been thought by many to have made a mistake here. The purported mistake is that he should have recognized the methods of egoism and utilitarianism as two different ultimate sources of reasons, and reasons can be weighed against one another. I will argue that it is a mistake to interpret rationality in terms of reasons-responsiveness in *the Methods*, and the consequence is that the dualism is still a problem.

Traditionally, Sidgwick has been read as a philosopher whose conception of rationality is responsiveness to reasons. However, Sidgwick heavily relies on the concept of "ought," including in his definition of a method of ethics (*ME* 1). I argue that Sidgwick defends a version of what John Broome calls "enkratic reasoning" in which rationality consists in a demand for consistency between one's ought-judgments and intentions. This point is crucial in interpreting an important passage in Sidgwick's discussion of moral motivation in which he distinguishes between "desiring" from "adopting an end" (*ME* 37). The adoption of an end involves an ought judgment, and when a person makes such a judgment, Sidgwick argues, rationality demands that they pursue that end. However, it is not intending an end that gives them a reason to pursue it. Rather, he argues that what they subjectively ought to do is what they judge that they ought objectively to do (*ME* 207-8), and I read "subjectively ought" here as a demand for consistency among one's own mental states (i.e., ought-judgments and intentions). Therefore, taken together, both the methods of egoism and utilitarianism say that persons, insofar as they are rational, objectively ought to adopt incompatible ends (i.e., one's own happiness or happiness from the point of view of the universe). Given that both methods prescribe ends one ought to adopt, and do so categorically, the rational pursuit of one end conflicts with the rational pursuit of another. Thus, the dualism consists of an inability to carry out one's intentions coherently.

Yuqi Liang (University of Oxford). *Sidgwick's Dualism of Practical Reason, impartiality in ethics, and evolutionary debunking arguments*

Is Sidgwick right in claiming that both egoism and utilitarianism are rational, and that the conflict between them—known as the Dualism of Practical Reason—cannot be resolved at all? I shall advance three interconnected arguments to weaken the reasons in support of egoism, thereby attempting to dissolve the Dualism of Practical Reason in favor of its impartial side. Part I presents the first argument, which

concerns the underlying structure of the two conflicting theories. I believe that there is a distinction to be made in understanding precisely what the Dualism is about, but in the common understanding of the Dualism among scholars of Sidgwick, the importance of this distinction has been overlooked. With this distinction in mind, I shall argue that while the impartial theory is supported by an abstract and self-evident intuition, rational egoism is not, which casts doubts on its reliability. Part II is largely a response to a persisting challenge from the rational egoist—the claim that I have a reason to favor my interests just because they are mine; I shall provide what I think of as the most plausible and compelling reading of this claim, informed by Caspar Hare's egocentric presentism; and I shall confront this challenge by trying to ground genuine normative reasons—as opposed to merely motivating reasons--in values. Part III is built on Katarzyna de Lazari-Radek and Peter Singer's recent attempt to dissolve the Dualism using evolutionary debunking arguments. I show that my first argument about the structure of the Dualism can strengthen their argument against a line of objection in the literature, making it another set of reasons against egoism. All things considered, I believe that reasons against rational egoism are weighty (though not decisive) and that the Dualism of Practical Reason, though not fully dissolvable, need not lead one to total pessimism about the role of reason in ethics. Despite persisting issues, practical reason nevertheless shows promising signs of finding its unity in an impartial theory.

Piero Tarantino (University of Milan). *Bentham's Theory of Practical Reasoning*

An important but overlooked contribution to the understanding of practical reasoning in early modern British debates on the nature of morality and its binding feature is provided by Jeremy Bentham. In his writings we may acknowledge at least two separate treatments of human thinking about what to do and how to act. The first treatment is contained in *IPML*, in particular in chapters i-xii, which might be integrated by certain paragraphs in *Preparatory Principles*. The second treatment is worked out in *A Table of the Springs of Action*, *Deontology* and some sections of Bentham's texts on logic and language. These latter texts were written roughly between 1813 and 1815, when Bentham's theory of real and fictitious entities was full-fledged, whereas the printing of *IPML* dates back to 1780 and then in 1789 when some bits were added. My paper will focus on certain passages contained in the above-mentioned writings with the purpose of giving an overview Bentham's theory of deliberation through the investigation of its foundations, that is, its underlying connection between obligation and motivation, which is key to understanding the normative character of practical domains. According to Bentham, morality, law and religion can bind their subjects, and thus provide them with guidance, only by motivating those subjects to conform their behaviour to moral, legal and religious norms.