

# Edgeworth's Mathematization of Social Well-Being\*

Adrian K. Yee

(Lingnan University; Department of Philosophy, Hong Kong Catastrophic Risk Centre)

Forthcoming in *Studies in History & Philosophy of Science*

## Abstract

Francis Ysidro Edgeworth's unduly neglected monograph *New and Old Methods of Ethics* (1877) advances a highly sophisticated and mathematized account of social well-being in the utilitarian tradition of his 19th-century contemporaries. This article illustrates how his usage of the 'calculus of variations' was combined with findings from empirical psychology and economic theory to construct a consequentialist axiological framework. A conclusion is drawn that Edgeworth is a methodological predecessor to several important methods, ideas, and issues that continue to be discussed in contemporary social well-being studies.

## Introduction

"[F]or mathematical ethics, an elementary knowledge of the calculus, and the experience of common sense, suffice" (1877, 60).

- Francis Y. Edgeworth

The well-being<sup>1</sup> of humans and animals remains a widely studied topic in modern philosophy (Davis 1981; Sumner 1996; Alexandrova 2017; van der Deijl 2017), economics (Easterlin 1974; Adler 2012), and psychology (Kozma et al. 1990; Clark et al. 2018). As it stands, histories of well-being research remain comparatively scarce, despite well-being's tremendous philosophical, practical, and political importance, with but a few articles (Brülde & Bykvist 2010; Angner 2011; Kaminitz 2018) and a few books on related topics, such as subjective utility (White 2006; Moscati 2019). The notion that

---

\*Correspondence may be address to: [adrianyee@ln.edu.hk](mailto:adrianyee@ln.edu.hk). I acknowledge the following for constructive feedback on previous drafts: Craig Fraser, Christopher Green, Jamie Michaelson, Jin Hao Xie, Steven Zheng, audiences at the Mathematical and Numerical Inspirations in Philosophy conference held at Stanford University (April 1, 2020), audiences at the 2nd ASU Winter Institute for the History of Economic Thought (January 22-23, 2021), and three anonymous referees.

<sup>1</sup>Throughout this article, I use subjective well-being to refer to an individual sentient (i.e. human or non-human animal) and their actual lived experiences of hedonic affect, in addition to their broader outlook on how well their life is going for them.

‘social well-being’ - the well-being of a group of sentient (e.g. a community or country of humans or animals) - may be amenable to numerical analysis was a common theme in 19th-century economic thought, appearing for instance in the much discussed work of William Stanley Jevons’s *Theory of Political Economy* (1871),<sup>2</sup> however even less work has been done trying to understand the history of social well-being studies.

A neglected figure in the history of social well-being studies is the economist and statistician Francis Ysidro Edgeworth (1845-1926). In what follows, I argue that Edgeworth ought to be considered a sophisticated thinker in the early history of mathematical models of social well-being, whose theories elicited a degree of precision and argumentative rigor unparalleled for its time and which foreshadowed several contemporary methods and issues in social well-being studies. While Edgeworth is ostensibly not as philosophically deep as his contemporaries active in philosophy departments and intellectual circles, this is not so. Rather, Edgeworth’s *New and Old Methods of Ethics* (1877) (hereafter, *Methods*) reveals philosophical depth embedded within dense mathematical formalisms attempting to combine psychology, meta-ethics, eugenics, and the ‘calculus of variations’ into a unified vision culminating in what he called ‘exact utilitarianism’. This includes surveying and critiquing the utilitarian philosophies of Alfred Barratt and Henry Sidwick, as well as the psychophysical work of prominent scientists such as Gustav Fechner, Paul Langer, Hermann von Helmholtz, and Joseph Delboeuf, amongst others.

More specifically, this article aims to contribute to the history of well-being research by expositing and assessing Edgeworth’s ‘exact utilitarianism’, a framework seeking to determine the ideal distribution of a set of stimuli and resources amongst a group of sentient beings so as to maximize the greatest amount of pleasure for the greatest number. Most recent work on Edgeworth has concerned his views on economic methodology, such as his usage of models of physics (Yee 2021), his position on women’s wages (Chassonnery-Zaïgouche & Cot 2021), his idiosyncratic conception of how utility might be measured as a ratio scale (Mueller 2020), as well as his personal and broader intellectual life (Barbe 2010; Mirowski 1994). While Kaminitz (2013) has already provided a broader history of Edgeworth’s thinking on utilitarianism from the period 1877-1881, showing Edgeworth’s engagements with economists such as Jevons, and Creedy (1984) has connected Edgeworth’s work on utilitarianism to his views on wage negotiation and arbitration, little work has been conducted towards understanding his usage of mathematics, Edgeworth’s thoughts about evolutionary theory, and his views on the well-being of non-human, sentient beings, specifically as a unified theory.<sup>3</sup> And it is a systematic analysis and exposition of *Methods* in particular, in which exact utilitarianism is articulated, that is missing from current scholarship, while his other more famous text

---

<sup>2</sup>Jevons describes his project as an attempt “to treat Economy as a Calculus of Pleasure and Pain,” with the goal of seeking maxima and minima of functions representing our affective mental states, via the usage of differential calculus (1871, vi-vii; xxvi).

<sup>3</sup>To clarify, there is excellent work on some of these topics understood individually. See Baccini (2009) for commentary on Edgeworth’s views on probability theory and see Mirowski (1994) for a biography on Edgeworth’s technical contributions. But this paper will provide a synthetic account of these topics as unified in his theory of ‘exact utilitarianism’.

*Mathematical Psychics* (1888) has already received considerable attention (Wall 1978).

Contrary to a remark from Joseph Schumpeter (1954, 798) that “Edgeworth lacked the force that produces impressive treatises”, I argue that *Methods* is one such treatise and foreshadows several methodological ideas and issues in contemporary theories of social and subjective well-being, even if he appears not to have been directly cited by contemporary well-being scholars. In this sense, Edgeworth’s contributions were important and yet unduly neglected considering their relevance to methods and issues in contemporary scholarship.

This paper is organized as follows. Section 1 gives an account of Edgeworth’s views on psychophysics, value theory, and the notion of ‘ancestral habit’. Section 2 outlines his theory of ‘exact utilitarianism’. Section 3 give a systematic account of his views on mathematizing social well-being and section 4 outlines the core methodological insights Edgeworth had which foreshadowed key theories and issues raised by contemporary social well-being scholars.

## 1 Edgeworth on Psychophysics, Value Theory, and Ancestral Habit

Edgeworth wrote in a period in which British utilitarian doctrines were thriving, where even family members, such as his grandfather Richard Lovell Edgeworth and aunt Maria Edgeworth, were adherents (Creedy 1984, 610). Such doctrines may be broadly categorised as consequentialist, insofar as an action’s moral value was to be evaluated first and foremost with respect to its consequences, and secondarily with respect to agents’ intentions, the promotion of virtue, and other relevant ethical criteria. 19th and early 20th-century consequentialism was often combined with the view that moral value was not amenable to scientific investigation, as in the work of philosophers Henry Sidgwick, Alfred Ewing, Hastings Rashdall, and G.E. Moore (Hurka 2014, 1):

[They] were all non-naturalists, believing moral judgments can be objectively true rather than...just expressing emotions, and have a distinctive subject matter...neither reducible to nor derivable from ones about natural science, theology, or metaphysics; no ‘ought’ follows from an ‘is’.

While Edgeworth agreed that moral judgments can be objectively true, he disagreed with other aspects of these philosophers’ thinking insofar as he believed that (a) an ‘ought’ can follow from an ‘is’ given that (b) natural science could, and in fact must, be consulted for the purposes of ethics. In methodological disagreement with these philosophers, Edgeworth believed in the legitimacy of two sources of evidence - empirical psychophysics and a form of ‘proto-evolutionary theory’ inspired by extant eugenics reasoning - the study of which could lead to a fruitful scientific theory he called ‘exact utilitarianism’.

*Methods*’s initial discussion concerns the topic of what motivates human action, suggesting a causal impetus in our desire for pleasure and avoidance of pain. He notes his agreement with Alfred Barratt - author of *Physical Ethics* (1869), an influential text for

its time - who believed that the forces of material bodies acting upon our body's physiological senses generate all of our affective states. Barratt's meta-ethics was grounded in a naturalism in which ethics (defined as the study of 'What is the chief good?') required recourse to facts about natural science, particularly psychology and physiology. While Edgeworth agreed, he expressed concerns that the conscious experiences of others appear not to be knowable except with respect to studying how physical phenomena affect their consciousness from an external, intersubjectively verifiable set of measurement methods. Hence, subjective mental states cannot be studied directly<sup>4</sup> and must instead be inferred via other, third-person means.<sup>5</sup> While ultimately eschewing what is ostensibly a form of reductionism, opting instead for an account involving the notion of what contemporary philosophers call 'supervenience',<sup>6</sup> he nonetheless advanced the proposition that "All human actions take their rise in pleasure," given that "of such forces...simple organic movements are built up all, even the highest, human actions" (Edgeworth 1877, 3). For example, in accordance with common sense, my desire to drink coffee, and the subsequent act of brewing a cup, is intuitively caused by the nearby scent of beans (e.g. a physiological response). While he believes that all mental states supervene upon physical states of our brains and physiology, they are not wholly reducible in this fashion. What is more controversial is that he believes that this is all there is that causes our desires (10):

[T]o every phenomenon of reflection are conjoined certain phenomena of sensation: so that definite physical phenomena...are the cause of all human action...of pleasure as a conscious feeling...of non-hedonistic preference...duty, practical reason, moral sense.

Notice that his claim entails that even more sophisticated, non-hedonistic attitudes, such as relational mental states like 'sympathy towards others', are caused entirely by our body's physiological processes (14). This is contrasted with a view which posits that while an experience of sympathy may be caused by, for instance, seeing a homeless person in the street, the experience of sympathy itself is a higher-order mental phenomena that is emergent (i.e. not reducible to) an aggregate of micro mental states, such as the

---

<sup>4</sup>As it turns out, Edgeworth's position on the scrutability of mental states would fluctuate over the course of his career. For instance, Colander (2007) argues that Edgeworth's later work *Mathematical Psychics* (1881, 101) contains passages that can be read as a defense of (1) the scrutability of mental states from a third-person perspective, (2) that they are quantitatively measurable, and that (3) such measurements are objective. Mueller (2020) agrees and further argues that Edgeworth believed that such measurements could be amenable to ratio scale transformations. However, this was not the view of Edgeworth in 1877. Rather, the Edgeworth of *Methods* thought that (1) was false and that we could have at most a surrogate, but measurable, concept of 'utility', as a mental state, that could satisfy (2) and (3).

<sup>5</sup>William Stanley Jevons, a friend and intellectual interlocutor of Edgeworth, would also hold this view: "We can no more know nor measure gravity in its own nature than we can measure a feeling; but, just as we measure gravity by its effects in the motion of a pendulum, so we may estimate the equality or inequality of feelings by the decisions of the human mind" (Jevons [1888] 2010, 11).

<sup>6</sup> $X$  supervenes on  $Y$  if and only if changes in the states of  $X$  necessarily entail changes in the states of  $Y$ , but not necessarily vice-versa.

conjunction of ‘negative affect’, ‘desiring for the poor to be better off’, and the experience of seeing. After all, one does not literally sense sympathy with their eyes’ retinas; rather, sympathy is arguably a mental state constituted by a complexity of sub-states, each of which is either externally sensed or an internal emotion that is not directly observable. It is in this sense that Edgeworth is most charitably interpreted as holding the view that higher-order ethical and affective states can be caused by more basic physiological sensations which can generate composite states that are higher-order, without being fully reducible to the supervenience base layer.

He nonetheless concedes that the view is problematized by “cases where the physiological conditions of consciousness and unconsciousness are not yet distinguishable” (4). The idea here is that there are some actions with explicit reasons given for them (e.g. I turned on a lightswitch because I find the room to be too dark) and others which, from the perspective of the subject, have no discernible rationale (e.g. a person is attracted to someone else but they are unable to state why). Each action is grouped under the three categories ‘reflex actions and volitions’ (e.g. breathing), ‘hedonistic preference’ (e.g. explicitly desiring to drink coffee), and ‘non-hedonistic preference’ (e.g. shaking people’s hands due to cultural custom) (5). Edgeworth acknowledges that it can sometimes be difficult to discern the ethical concepts employed by agents with respect to these first and last categories. For instance, someone deciding to give a gift might have an underdetermined motivation, such as it being an act done out of altruism or merely out of social obligation (e.g. it is someone’s birthday party, and it is a convention to give gifts at parties). Despite introducing the distinction between conscious and unconscious motives, he merely acknowledges the distinction without going into further detail as to how to study it, so as to state his position within the space of views present at the time. In this sense, the study of the well-being of a society is partially dependent upon the study of that society’s ethical life.

Aware that the idea that physiology grounds mental life was controversial, he pre-empted several objections. One salient objection is that since social scientists, let alone human beings more generally, do not agree on a conception of morality, individual subjective well-being, or social well-being, we therefore cannot study it using mathematical techniques. And if attempts at mathematization are suspect, Edgeworth believed that the prospects of a proper science of well-being are similarly tenuous. After all, it would appear to be putting the cart before the horse if one developed a mathematical theory articulating subjective well-being without having a clear sense of what subjective well-being is to begin with. Edgeworth acknowledged that this methodological issue is to be contrasted with how Newton knew what sound was, and Bacon what heat was, prior to developing their respective theories of acoustics and elementary thermal physics (21). In these cases, there was a relatively clear and recognizable phenomenon that was to be the target of a scientific theory’s predictive and explanatory content; conversely, this seems not to be the case when it comes to human happiness and well-being. Nonetheless, to paraphrase, Edgeworth believed that this epistemic conundrum is consonant with how scientists overcame issues during much of the history of science, such as the history of temperature measurement, wherein an intuitive understanding of a phenomenon

guides an epistemic process of increasing calibration, cross-reference, and amendment, despite initial attempts at operationalizing measurement of a phenomenon appearing futile (Chang 2004). Hence, it is actually possible for a mathematization of social well-being to occur in theory, so long as the right epistemic process of construct creation was conducted in a legitimate fashion.

Edgeworth further responded not only by arguing against the view that mathematization removes the beauty and mystery of ethical life, but added that “[m]orality might be no more injured by physical science than music by acoustics” (22), and that “for mathematical ethics, an elementary knowledge of the calculus, and the experience of common sense, suffice” (60). That is, while the usage of mathematics might give the initial impression of an austere detachment from the complexities of human experience, rendering the theory potentially doomed from the start, he thought that it is only through the methods of natural science that the richness of human ethical life could be revealed and articulated in precise ways. Furthermore, he believed that so long as a reasonable approximation to the intuitive concept of subjective well-being can be operationalized into a measurement theory (e.g. by using psychophysical methods incorporating sensory data from sentient), that this would be sufficient for us to develop at least the beginnings of a science of social well-being.

More specifically, he argued that there is a sense in which all sources of hedonism may be compared numerically along a single scalar measure. Edgeworth further posited that such objects’ axiological value admits of a ratio scale, thereby allowing for not only the calculation of sums and differences of pleasure and pain, but multiplication and division of such states as well.<sup>7</sup> Additionally, differential and integral calculus can render such measurement procedures lucid, and positive and negative events can offset one another in some cases. This is revealed in one of his many characteristic poetic passages throughout *Methods* (26):

No multiple of the pleasure of eating tarts, says Lecky, can be equated to the pleasure of doing a generous action. Perhaps not, in the volumes usually compared. . . [O]ne might suppose that the so-called lower pleasures are related to the higher, somewhat as differentials to an integral, incommensurable indeed, yet capable of being equated after infinite summation. The permanent increase of material comforts and pleasures over an indefinite area of society, and through countless generations, may be set off against a definite and limited dereliction of moral beauty.

This passage reveals five core assumptions connecting value theory to psychophysics

---

<sup>7</sup>While he claims this, the actual justification for positing a ratio scale in *New and Old Methods of Ethics* (1877) cannot easily be found, if there is such an argument present at all, and is in any case not intuitive. This is because it is unclear what a non-arbitrary zero-point of subjective well-being, or of pleasure or pain, would be, such that operations of multiplication and division could be applied. By way of contrast, it is intuitive to understand the Kelvin temperature scale as having a non-arbitrary zero point at the theoretical minimum temperature where mean molecular motion ceases altogether. However, there are certainly more attempts at justification for a ratio scale of utility that Edgeworth provided in later works; see Mueller (2020, 714-716) for further discussion.

and mathematics in Edgeworth's thinking: (1) axiological value can be studied quantitatively in a non-trivial fashion; (2) many ostensibly axiologically incomparable goods can sometimes be compared given appropriate mathematical methods and contexts; (3) axiological value can be infinitely divisible and studied using the methods of calculus; (4) pain and pleasure can form a net aggregate effect in a sentient being's consciousness; (5) intergenerational well-being can offset a current generation's deficit in well-being. These assumptions were controversial for his time; for instance, his interlocutor Henry Sidgwick would deny exactly these kinds of arguments given that they rest upon the problematic premise of interpersonal comparisons of cardinal utility (Sidgwick 2000, 6):

It is assumed that all pleasures are capable of being compared quantitatively with each other and with all pains...and that this quantity can be known; so that each can be weighed in ideal scales against every other...I wish to point out that it is at any rate not verifiable by experience, and that very plausible objections may be brought against it on empirical grounds.

Unfortunately, Edgeworth gives no explicit argument in defense of assumptions (1) - (5) and merely provides examples intended to elicit and persuade the readers' axiological intuitions. After all, it is surely a substantial assumption to say that hedonic pleasures (e.g. the pleasant taste of ice cream) can be equivocated, under appropriate aggregational constraints, with higher-order states of well-being (e.g. consummating a marriage with someone). This would presuppose, among many other assumptions, the existence of an interval scale, according to which two phenomena  $x$  and  $y$  can be compared to one another in an arithmetic fashion, as opposed to merely an ordered ranking. For example, Edgeworth believed that the pleasant taste of ice cream can be represented by a real number  $x$  of pleasure and the consummation of a marriage as  $y$  of pleasure such that  $y = \int_0^\infty \phi(x)dx$ , for some integrable well-being function  $\phi(x)$  representing appropriate hedonic mental states that converges. This is a radical claim as it alleges that there could be some cases in which ostensibly qualitatively distinct sources of pleasure or pain can be compared between one another. He claimed that the form of  $\phi(x)$  would be determined by the specific content of the sentients in question and the situational context of their comparison. In this case,  $\phi(x)$  would be contingent upon sentient beings' sensory capabilities and the contingencies of culture. And yet, it is not at all obvious how this could be the case given that Edgeworth's methodological starting point is to reason in analogy with the experimental context and metrological foundations of classical mechanics. That is, the kinds of entities typically measured in the natural science of mesoscopic bodies, such as tables and chairs, admit of more clearly delineated boundaries and identifiable instances than such complex, socially constructed entities as marriages, such that it is unclear how one could compare the value of a marriage (or marriages) to the value accrued from consuming something like ice cream. In this sense, this is a reminder to readers of Edgeworth that he sometimes made omissions of justification, occasionally rushing off to the next topic rather than pausing to appease the reader's concerns, making it sometimes difficult to understand what exactly his views were.

Nonetheless, he continues along these lines and entertains whether a pleasure-function

might have a high value as a result of the stimulus itself or rather as a result of the capacity of the sentient being for attaining pleasure (Edgeworth 1877, 73):

How do we know that Pongo<sup>8</sup> has not a superhuman pleasure in simple sensations and muscular movements? How do we know, as Mr. Sully asks, that birds, attracted by bright colours, do not feel a simple pleasure more intense than we can realize? Can we be certain that does not lose in respect of  $\beta$  what does it gains in respect of  $k$ ? It may be hoped that advancing psychophysic [sic] will throw some light on this question.

The thought here is that some measures of an individual’s subjective well-being may sometimes be underdetermined between at least two competing hypotheses: the sensitivity of the animal or the power of the stimulus. In fact, different conscious beings (e.g. non-human animals) may possess superior capabilities of pleasure (or pain) than those of others. This is scientifically plausible insofar as, for instance, the mantis shrimp (*Haptosquilla trispinosa*) possesses the capability of experiencing many more color experiences than humans are, given upwards of 12 photoreceptors (Cronin et al. 2001). But Edgeworth grants that it was not clear from the ethological and psychophysical data of his time that we could empirically discern the relevant properties to answer this sub-question in social well-being studies. His point was that this was nonetheless possible given some future set of natural and social sciences, and that preliminary zoological studies suggested that this was a hypothesis worth theorizing about.

The last component of Edgeworth’s account of the foundations of sentients’ well-being is his notion of ancestral habit as the source of non-hedonistic preferences, an idea he attributes to the influence of Herbert Spencer. Such habits could be described in contemporary terms as a product of a combination of ‘human evolutionary psychology’ and extant eugenics thinking of Edgeworth’s time. He begins by noting that our being endowed with cognition of something *a priori*, such as the spatial geometry of our external senses, is not something grounded in our ancestors’ sensory history, but is rather endowed from birth, following philosophers such as Immanuel Kant on the transcendental nature of space (13). Instead, Edgeworth believes that there is a general tendency for not only our ethical views but also our aesthetic tastes to alter and potentially improve over our evolutionary history: “For that an approximation to the utilitarian desire is in the line of evolution is sufficiently probable, the sympathies of men being ‘widened with the process of the suns’” (31). Edgeworth’s idea is that the ‘ethical intuitionism’ of Henry Sidgwick, understood as a meta-ethical view in which some ethical propositions’ truth-values can be grasped by pure intuition alone, ought instead to be rendered naturalized via the proto-evolutionary psychological theories of the time.

The concept of ancestral habit furthermore connects directly with an even more controversial idea of Edgeworth’s in which he explicitly defends the Eurocentric and racist

---

<sup>8</sup>The satirical London-based periodical *Punch, Or the London Charivari* published an issue the same year of Edgeworth’s *Methods*, mentioning ‘Pongo’ as follows (Punch 1877, 242): “Pongo, the first Gorilla who had exposed himself, in the cause of discovery, science, and philanthropy, to the dangers of missionary enterprise in Europe, and had paid for his devotion with his life.”



view that aristocratic European countries have superior cultural habits and physiological capabilities for pleasure (Edgeworth 1877, 55):

Unto him that hath higher development shall be added more of this world's goods. This deduction agrees with common sense, as exhibited by approved dealings of men with animals, of civilized with savage races, in the privileges of aristocracy approved in ages when aristocracies really represent a higher order of evolution...For it was never intended by a sound utilitarian that "Mr. Pongo" was to count for one.

This is a product of extant eugenic views which were then widespread in many countries, especially the United States and Europe, and advocated by interlocutors such as Herbert Spencer (Levine 2017).

In summary, Edgeworth's meta-ethical and physiological views blended with a philosophy of science which posits that subjective well-being is intersubjectively impossible to directly measure and yet sufficiently measurable via an observable proxy, evolutionarily conditioned, and contingent upon one's objective cultural standing in some axiological ordering over the set of all human cultures. Having articulated these core methodological assumptions, I turn to Edgeworth's method of 'exact utilitarianism' which combines these assumptions with several mathematical formalisms into a full fledged theory of social well-being.

## 2 Edgeworth on Exact Utilitarianism

Edgeworth's primary contribution to social well-being studies in *Methods* was to articulate, defend, and criticize different components of Utilitarian doctrine using several advanced mathematical methods of his time. He follows Sidgwick in defining the spirit of utilitarianism as "the greatest quantity of happiness of sentient, exclusive of number and distribution - an end to which number and distribution are but means" (Edgeworth 1877, 35). He also follows Sidgwick's idea, common throughout his writings, that what counts as 'good' is what is essentially desirable from a universal point of view (Hurka 2014, 37). And yet, such universalism was to be relativized in some cases to the specific physiological facts of a given sentient's bodily and mental capacities for pleasure and pain.

This is so in two senses. The first is Edgeworth's idea that ethical value resides in facts about agents' physiological states, as opposed to merely their subjective preferences, and is emblematic of his purported naturalistic method. The second is that there will tend to be facts of the matter (though not always, as will be shown below) which determine what distribution of pleasure will be best (with a plurality of 'bests', depending on the nature of the distribuend). While he thought that there is a certain degree at which a person's subjective preferences ought to be taken into account in any adequate scientific study of both subjective well-being and social well-being, Edgeworth emphasized that it is possible for a person to be mistaken about what makes him or her well off (Edgeworth

1877, appendix note A, 80-81). This follows as a consequence of his naturalism about how to investigate the causes and structure of both subjective well-being and social well-being, by which external analysis of a sentient is required to have an objective perspective of well-being.

Before applying his psychological theory, he begins the outlines of what he calls ‘exact utilitarianism’ by articulating mathematically different ways in which one might seek to find the greatest quantity of a thing  $U$  in terms of another  $V$ . We can interpret Edgeworth as holding the view that such a formal articulation is indispensable insofar as it renders less ambiguous distinct meanings of the phrase ‘the greatest happiness for the greatest number’. If we suppose  $U$  and  $V$  to be functions of the same variables, and we would like to find local maxima, then this means finding solutions such that  $\frac{\partial U}{\partial x} = \frac{\partial V}{\partial x} = \frac{\partial U}{\partial y} = \frac{\partial V}{\partial y} = 0$  and that  $\frac{\partial^2 U}{\partial x^2} < 0$ ,  $\frac{\partial^2 U}{\partial y^2} < 0$ , and  $\frac{\partial^2 V}{\partial x^2} < 0$ ,  $\frac{\partial^2 V}{\partial y^2} < 0$ . Since Utilitarianism is concerned with the ‘greatest happiness’ (e.g.  $U(x)$ ) for the ‘greatest number’ (e.g.  $V(x)$ ), we seek to maximize a function  $W(U, V) = U + V$  insofar as, to use his example, one may “look for the workman with most strength and most skill, when we look for the *most efficient* workman” (36-37). Hence,  $x$  may be some parameter such as ‘increase the work day by  $x$  amount’ and  $y$  the ‘number of people per office cubicle’, so that we might seek to find the ideal length of the workday and office space which will maximize social well-being. This is but the first of many senses in which he thought that the general ethos of Utilitarianism may be articulated and defended.

As a second form of a Utilitarian maximization problem, Edgeworth invites us to calculate the maximum value of a function  $U$  *constrained* by another function  $V$  (37). A charitable interpretation of what Edgeworth is suggesting is to employ methods analogous to ‘Lagrange Multipliers’ in which conditionalizing upon some function, we seek the maximum that can be had when another function intersects its domain.<sup>9</sup> For instance, in accord with Edgeworth’s formal constraints on any well-being function, consider a function  $f$  whose first derivative is positive, 2nd derivative is negative, and is defined over the positive real numbers (53). This is a reasonable assumption to hold at the time insofar as many economists and psychophysicists of the 19th-century believed in diminishing marginal utility, whose ontological structure possessed these mathematical properties. For example,  $U(x, y) = \log(x) + \log(y) + k$ , for some constant  $k$ , is one such function. Consider now another function  $V(x, y) = 144 - x^2 - y^2$ , which is a cylinder of radius 12, perhaps representing a ‘time constraint’ on labor during a working day. The method of Lagrange Multipliers then asks us to take partial derivatives of  $U$  and  $V$  and then arrange and subsequently solve the following system of equations for  $x, y, \lambda$ :

$$\frac{1}{x} = 2x\lambda$$

$$\frac{1}{y} = 2y\lambda$$

---

<sup>9</sup>See Stewart (2010, 957-964) for an overview.

$$x^2 + y^2 = 144$$

Solving, we have  $\lambda = \frac{1}{144}$ , since  $x, y$ , are positive real numbers, and so  $x = y = \frac{12}{\sqrt{2}}$ .

Hence, the point  $P = (\frac{12}{\sqrt{2}}, \frac{12}{\sqrt{2}})$  is a solution to this maximization problem. One could interpret this example as a call for administering resources in a certain fashion, such as, for instance, choosing the optimal time spent on some task given a budget constraint. In this sense, Edgeworth took a method more commonly used in mathematical applications in engineering and applied the method to economic and value-theoretic problems in what were then novel ways.

As a third form of Utilitarianism, Edgeworth considers functions  $U = \int \phi(x, y, \frac{dx}{dy}, \dots)dx$  and  $V = \int \psi(x, y, \dots)dx$ , where the task is to determine the form of  $y$  so that  $U$  is a maximum and the limits of the integral are such that  $V$  is a maximum. This is essentially a problem in the ‘calculus of variations’, a commonly used method in modern physics, dating at least as far back as Newton’s work in *Principia Mathematica* in 1687 (Fraser 2009), and a method Edgeworth was familiar with from his studies of natural science. Indeed, he is known to have read John H. Jellett’s influential *A Treatise on the Calculus of Variations* (1850)<sup>10</sup> and the works of mathematical physicist William Rowan Hamilton, the latter a friend of Edgeworth’s mother Rosa (Barbe 2010, 36). Furthermore, recent scholarship has argued that Edgeworth’s interest in mathematical physics and statistics were core elements of his methodological views in economics more generally (Yee 2021, 7-9). We can interpret the class of problems he is concerned with in this third form of utilitarianism as involving ‘functionals’, which are functions mapping functions to real numbers. For instance, we might seek a function  $y$  such that  $U[y] = \int \phi(x, y, \frac{dx}{dy}, \dots)dx$  is a function where  $y'(x) < 0$ , implying that, given suitable boundary conditions, a solution to  $U[y]$  is a maximum. Similarly, given a function  $y$ , we seek the limits  $x_0, x_1$  of the function  $V$  such that  $V(x_0, x_1) = \int_{x_0}^{x_1} \psi(x, y, \dots)dx$  is a maximum.

What does Edgeworth make of these three different articulations of Utilitarian doctrine? While the first type of problem is readily acknowledged as commonly used in his time, he thinks this view is inadequate given that considerations of the greatest happiness are considered as taking values independently from the greatest number, and so admits many counter-examples. That is, while the function  $W$  takes  $U$  and  $V$  as input functions, they are linear sums and not functions whose values depend on each other, whereas many social well-being issues, such as political conflict, interpersonal issues, and workplace disputes, arguably involve relations between multiple variables that are non-linear (Edgeworth 1877, 72). He considers the second example employing Lagrange

<sup>10</sup>Edgeworth’s writings in *Methods* follows the symbolic formalism employed in Jellett’s *A Treatise on the Calculus of Variations* (1850); I have tried my best to preserve Edgeworth’s original formalisms while occasionally rewriting integrals using modern notation in cases in which either a typo is made in Edgeworth’s original passages or the notation is too inarticulate to be understood, relative to modern mathematical notational standards.

multipliers more adequate given that it captures the idea that the “greatest *product* of number and average happiness...should be sought” (39), emphasizing the non-linear properties of most scenarios involving social well-being. The third example he thinks is also an adequate articulation of sound Utilitarian doctrine and expresses the idea of a lexicographic ordering, insofar as “the greatest average happiness should *primarily* be sought; *secondarily* the greatest number” (39). Unfortunately, he falls short of providing a more systematic analysis outlining practical applications of the theory in *Methods*.

Having articulated these three different doctrines, he proceeds to connect psychophysical findings from research conducted by his contemporaries so as to enhance the empirical adequacy of his theory. For instance, he discusses Gustav Fechner’s ‘law of stimulus’, where the pleasure of each ‘sentient element’ is represented:

$$\pi_e = k(\log\gamma - \log\beta)$$

where  $\gamma$  is the strength of the stimulus (quantified as a real number),  $e$  an index representing a given ‘sentient element’, and  $k, \beta$ , are real number constants to be discerned by experiment. By ‘sentient element’ he means each entity that can have experiences, where it is reasonable to interpret Edgeworth as meaning that experiences are processes that certain entities can undergo only if there is ‘something it is like to be’ that entity (Nagel 1974). For instance, animals and humans are clearly sentient elements in this sense, and can therefore be subjects for whom other objects can elicit pleasure or pain, whereas tables and chairs have no conscious experiences. Social well-being calculations are therefore supposed to be potentially applicable to all sentient entities, demonstrating Edgeworth’s desire for a universally applicable theory that avoids anthropocentrism.

Furthermore, Edgeworth notes that in addition to the function  $\pi_e$ , one could just as equally consider other functions that have a positive first derivative and a negative second derivative, given that the primary criterion to be satisfied is that as the level of pleasurable sensation increases, its rate of increase decreases (Edgeworth 1877, 41). In addition to Fechner’s, he included the following three laws posited by other 19th-century psychophysicists of his time of the structural form of pleasure and hedonic functions:

Paul Langer:  $k\log\left(\frac{c\gamma^2 + b}{b}\right)$

Hermann von Helmholtz:  $\frac{a}{G - \gamma_0}\log\left(\frac{\gamma_0 + \gamma}{G + \gamma}\right) + C$

Joseph Delboeuf:  $k\left[\log\left(\frac{c + \gamma}{c}\right) - \log\left(\frac{m}{m - \gamma}\right)\right]$

Here,  $\gamma$  is the strength of the stimulus, and  $\gamma_0, a, b, c, , C, G$  are real-number valued constants to be determined by physiological facts of the subject receiving the stimulus.<sup>11</sup> He heavily criticized Langer’s formula as a “not very plausible formula” since the formalism is insufficiently sensitive to differences in individuals’ hedonic capacities (40). While he

---

<sup>11</sup>Unfortunately, as far as I know, Edgeworth does not elaborate at all on the role of these constants any further in the text and in other writings. I have tried my best to find other historical sources on these thinkers to elucidate the specific roles played by these constants, beyond their functional role as empirically discerned constants relevant to the calculation of well-being, but I have been unsuccessful.

was more agnostic about Helmholtz and Delboeuf’s accounts, at the very least acknowledging their consistency with reasonable constraints upon their 1st and 2nd derivatives, Edgeworth’s own preferred law is the quasi-Fechnerian law:

$$\pi_e(y, \beta) = k|f(y) - f(\beta)|$$

in which  $y$  is the strength of the stimulus,  $e$  an index specifying the specific sentient,  $\frac{d}{dx}\pi$  the rate of pleasure of a sentient element during a period of time,  $f$  an empirically discerned function whose first derivative is positive and whose second is negative, and  $\beta, k$  coefficients, where  $\beta$  is the threshold defined as “the lowest value of stimulus for which there is sense of pleasure at all” (42).<sup>12</sup>  $k$  is to be interpreted as a ‘capacity for pleasure’, with larger values meaning larger capacity (e.g. some species of bird may have visual experiences of more vibrant colors than others). Following the work of psychophysicist Wilhelm Wundt, Edgeworth adds that it is clear that past a certain value of  $y = z$ , not only does pleasure stop but that it can lead to displeasure if the stimulus is too great (e.g. too high heat can scald one’s skin) (41). This value of  $z$  is to be determined purely by experimentation and cannot be deduced *a priori*, a view yet again representative of Edgeworth’s insistence on a naturalistic approach to the analysis of well-being.

Lastly, Edgeworth even foreshadows<sup>13</sup> psychological theories positing *asymmetries* in hedonic value by noting that “minus pain is sweeter than plus pleasure, that what is given to the most miserable is given most felicitically,” citing Adam Smith and Wilhelm Wundt as predecessors who thought that an increase in pleasure is not as profound an experience as is the lack of pain. This he acknowledges as a thorn in the rose bush of Utilitarianism and concedes that this “reflection is one of the most important of the limitations which utilitarianism imposes on itself” (75). In this way, he made no attempts at advancing a perfectly, logically consistent theory but rather one that merely aimed to improve significantly upon extant proposals and to produce a properly interdisciplinary theory drawing upon a diversity of sources ranging from literature and philosophy to psychophysics and mathematics.

To summarize, this section presented Edgeworth’s mathematical articulations of Utilitarianism, his preferred views on empirical psychophysics, and how these ought to be synthesized into a holistic theory. The next section combines these inquiries into a broader theory as to how a set of commodities, services, or objects of pleasure and pain, may be distributed so as to satisfy different Utilitarian goals, culminating in a comprehensive theory of social well-being.

---

<sup>12</sup>It is especially unclear why he preferred an absolute value as a measure given the definition of the beta coefficient.

<sup>13</sup>This foreshadows the efforts of Krueger et al. (2009) and their ‘Unhappiness Index’. This is a utilitarian aggregational procedure designed to be sensitive to their assumption that unhappiness is not only distinct from happiness but can be easier to measure.

### 3 Exact Utilitarianism and Social Well-Being

Edgeworth's Exact Utilitarianism was a program specifically concerned with social well-being, as opposed to merely individual subjective well-being. The central problem of social well-being theory for Edgeworth is stated as follows (43):

“Given a certain quantity of stimulus to be distributed among a given set of sentient (with the condition that every element is to have some stimulus), to find the law of distribution productive of the greatest quantity of pleasure.”

Notice here that unlike many contemporary welfare economists, Edgeworth is not interested in the most efficient distribution in the general sense of Pareto optimality nor something like a Kaldor-Hicks<sup>14</sup> criterion of welfare distribution. This is because Edgeworth is less concerned (at this period in his thinking, at least) to advocate accounts of ordinal preferences rankings, as he believed that this would allow too much room for the subjective preferences of individuals to dictate distributions. In this sense, Edgeworth is not trying to do welfare economics as understood in neoclassical traditions.<sup>15</sup> Instead, he is interested in what an objective science of psychology could potentially tell us about what actually makes us have well-being irrespective of whether subjects themselves agree with it. However, he surprisingly foreshadowed a common assumption in modern welfare economics that an adequate account of social well-being ought to be impartial and anonymous with respect to the distribution, insofar as reference to specific members of the population is avoided, which is now a common assumption in social well-being formalisms (Adler 2012, 7).

He discusses several different scenarios in which social well-being calculations are made, specifically the following four:

**Case #1:** All members of the population have equal stimuli and capacity for pleasure. This is computed as  $\sum_{i=1}^n kf(y_i) - f(\beta)$ , with each  $y_i$  given. We wish to maximize:  $\phi = \sum_{i=1}^n kf(y_i) - \sum_{i=1}^n c(y_i)$ .

For instance, this could model a homogeneously distributed set of people of sufficiently similar cultural and physiological background, given that  $f$  is a function of a constant  $\beta$ . Since  $f''$  is assumed to be negative, he claims that the solution such that  $\phi$  is maximized is given by  $kf'(y_i) = c$  for all  $i$ , which entails that all  $y_i$  are also equal, meaning that the distribution of resources and stimuli should be equal too.

**Case #2:** Where only the stimulus varies but the capacity is the same.

---

<sup>14</sup>A distribution is Kaldor-Hicks efficient only if a Pareto improvement could be achieved by costless redistribution of the sum total of goods (Adler 2012, 98).

<sup>15</sup>While the concepts of Pareto optimality and Kaldor-Hicks efficiency postdate the time of Edgeworth's writing, a later paper of Edgeworth's discusses the related idea that a social welfare distribution may be objectively maximal while simultaneously admitting a redistribution, so long as the total quantity of welfare was not altered (Edgeworth 1923, 489-490). However, welfare in this sense is not to be confused with 'social well-being' in the sense in which concerns his theories and broader discussion in *Methods*.

Here, he believes the solution is simply analogous to Case 1.

**Case #3:** Where only the capacity for pleasure varies, and the stimulus is constant.

Here, we must maximize  $\sum_{i=1}^n k_i f(y_i) - \sum_{i=1}^n c(y_i)$ . He claims that the solution is given by  $k_i f'(y_i)$ , for all  $i$ .

**Case #4:** Similar to Case #2 but both the stimulus and the capacity vary.

Again, he claims this is analogous to the solution for case III.

A second family of social well-being problems he thinks we might attempt to solve is the following (44):

Given a certain quantity of stimulus to be distributed over some part of a given sensory tract (without the condition that each element of the given tract is to have some stimulus), to find the part and the law of distribution over it productive of the greatest quantity of pleasure.

He first solves this for the one-parameter case, using the simplified example of a sensory tract having square-shaped sensory input defined by  $x$ .<sup>16</sup> For instance, we might consider the social well-being derived from relieving a person who is loudly complaining about a wood splinter caught in their finger, and calculate the pleasure received from removal of that splinter at a specific part of one's finger. The pleasure of the tract is then given by:

$$\int_{x_0}^{x_1} k(f(y) - f(\beta)) dx$$

where  $y$  is the distribution over the required region,  $x_0$  and  $x_1$  are the given limits, and  $\int_{x_0}^{x_1} y dx = D$  is a given quantity. One then attempts to maximize  $\int_{x_0}^{x_1} [k(f(y) - f(\beta)) - cy] dx$ . This is essentially a generalized and hybrid version of the Lagrange multiplier problem and problems in the calculus of variations, as discussed in the previous section.

At this point it is worth asking what Edgeworth thought distinguished measures of an individual's subjective well-being from that of a social group's well being? The difference is that social well-being calculations require not only finding the pleasure functions of each individual but also understanding the nature of the *relations* between each individual in the group, representing what he thinks are often non-linear structures in society that condition our well-being. For example, the respective individual happiness that a friend and I have in isolation is positively and additionally affected by our relation to one another as friends, insofar as that friendship is healthy. Hence, these relations are not present when calculating the subjective well-being of each individual in isolation:

---

<sup>16</sup>While the idea in the abstract is clear enough, it is unclear what, if any, specific sentient creatures he has in mind that would have this sort of structure to one of its sensory modalities, as he doesn't say. However, we might charitably interpret Edgeworth as interested in providing a simple example for pedagogical purposes, and that the point is intended to generalize to more complex examples.

“[E]manations received...depend not only on the nature of his associates, but also upon their means, *their* shares of the distribuend” (58).

Therefore, the correct social well-being function should be represented by the multiplicatively structured functional:

$$S[y] = \int_{x_0}^{x_1} [F(\beta, k, \dots, y) \int_{\theta}^{\eta} f(\beta, k, \dots, y) dx] dx$$

where the outside integral represents the “nature and share of the individual” and the inner integral represents “the stimulus which the social unit centring receives from his associates...which vary with the individual’s order of evolution ( $\eta$  and  $\theta$ ) functions of  $x$ ” (59). Here, a function  $y$  is sought which maximizes the value of  $S$ , and  $x_0, x_1$  are limits of the integral to be determined by the context of application (e.g. perhaps an interval of time or a certain boundary condition on a set of resources to be distributed). This is further simplified in the following:

$$\int_{x_0}^{x_1} \pi dx = \int_{x_0}^{x_1} [\int_{\theta}^{\eta} XYxydx] dx$$

Here, the large  $X$  and  $Y$  represent ‘the nature and share of the social unit’ and the adjacent  $x, y$  denote the nature and share<sup>17</sup> of an individual sentient’s associates (i.e. the associates that the social scientist modeling the phenomenon thinks are relevant variables) (59). Edgeworth’s idea is that a proper social well-being function should take into account the share of a social unit’s contribution to the well-being of the individuals of which that unit composes. This is represented as a multiplicative factor of the individual’s own well-being.

As examples of social relations which can be articulated mathematically, Edgeworth considers the kinds of ‘useful qualities’ of other people to consist in a vast array of traits (58):

[P]hysical strength, inventiveness, industry, ambition...symphonious passion, to the qualities which irradiate the purest pleasures, to rational benevolence, and romantic love, and friendship, and even friendliness and good manners.

These differences between sentients are salient enough that Edgeworth believes that utilitarianism must be sensitive to differences in (1) an individual’s capacity for pleasure as treated as an isolated person; (2) an individual’s relationship to their group, and the net affect had with that group; (3) the extent at which societies have different rankings in terms of the net utilitarian effects (65). Furthermore, he is clear that naive summations of individual subjective well-being do not necessarily translate into adequate measures of social well-being. Rather, he believes that there is a sense in which social well-being is emergent from that of society’s individuals: “For in the arithmetic of hedonism, the

<sup>17</sup>If this is indeed the ‘share’ then it is unclear why Edgeworth did not stipulate that  $y$  be a percentage, and thus a real number bounded between 0 and 1 inclusive. Unfortunately, Edgeworth often neglected to elaborate on technical details and this is no exception.



whole is not to be equalled to the sum of the parts, but to the sum + the relations of the parts” (72). This relational account of well-being was therefore often heavily emphasized throughout later sections of *Methods*.

To summarize these previous three ideas in further detail, idea (1), of the contingency of affective experience upon one’s species, has already been discussed in previous sections of this paper. Idea (2) is that some people’s capacities for more sophisticated states of well-being, such as enjoyment, attention, and high regard, will reach their full potential for actualization only in certain circumstances, with Edgeworth using the example of a comedian requiring an appropriately sophisticated audience to fully appreciate their joke to its fullest potential (65). For Edgeworth, a deficit in potential well-being translates into a deficit in the “respite from the average amount of work expended upon the distribuend” (65). For instance, a comedian requiring more effort to get their audience to laugh will have lower values on a pleasure function and will require more labor to produce social well-being. Idea (3) is a consequence of his view that different human societies are at different stages of ‘evolution’ - where evolution<sup>18</sup> is here understood as economic, cultural, and moral development: “[W]e may now regard all living beings congregated on the mountain-tops; the highest and most favoured regions being occupied by the most capable” (52). Here, the idea of a stage of a society’s ‘evolution’ is a non-linear property of a group as contrasted with the order of evolution of the ‘race’ of an individual, which is a eugenic idea core to Edgeworth’s views on social well-being.

To illustrate how each of these different ideas intersect, he raises in a salient passage the question of enacting tariffs and restrictions on labor immigration noting that it may be foolish, depending on probabilistic calculations, to impose legislation prohibiting Chinese people to enter a country, on the grounds that the Chinese are allegedly inferior in several ways (76):

*[P]rima facie*, unequal legislation directed against the influx of Chinese labour might be justified, on the supposition that, if on a large scale Chinese competed successfully with Aryans, an inferior race would inherit the earth.

However, he is clear in cautioning the reader that he believes that it is merely a historical contingency that this is so, and that it is possible that the Chinese could “catch up [to] the superior in the race of evolution, and become ultimately as highly civilized” (76). Hence, he believed that the Chinese, or any particular race for that matter, were not intrinsically inferior as a result merely of their biology<sup>19</sup> but rather their current place in the history of their economic and cultural development was to blame. Social and historical contingencies were therefore factored into account in his formalisms by way of the following social well-being function:

---

<sup>18</sup>It is important to emphasize that Edgeworth sometimes used this term in a more colloquial sense, as in this passage, and yet in other contexts meant it in his technical sense of eugenic theory.

<sup>19</sup>Undoubtedly, he nonetheless believed what we would now call genetic factors had a highly non-trivial contribution to the sensory capacities of ‘races’ (a concept he believed in the existence of) as well as the individual idiosyncratic physiological capacities of individuals from each race.

$$\pi = nF(x, \frac{\phi(n, x)}{n} - T)$$

where  $n$  is the population size,  $x$  the ‘order of evolution’ quantified as a real number (i.e. a society’s value as measured against all other societies’ value),  $T$  the well-being threshold below which a sentient cannot exist,<sup>20</sup>  $\phi(n, x)$  the distribuend, and  $\frac{\partial F}{\partial x} > 0$ . Hence, the social well-being of a population is a multiple of the number of people and a function of the ‘order of evolution’ and the average distributed affect, minus the threshold value, so long as the function is positively increasing.

To expand on his example, he believed that the value for  $x$  will be higher for members of Victorian England than it would be for citizens of 19th-century Chinese society, adding the caveat that “As long as  $T$  does not increase with  $x$ , an indefinite progress in evolution is desirable” and that “if the threshold increased with evolution, then we should tend to a ‘stationary state’...not only wealth and number, but also...cultivation, evolution” (78). This hypothetical stationary state would be some idealized future state of society which attained a perfect balance between the needs, desires, and resources of a society. This state is stationary insofar as over time, this ideal state would remain in this equilibrium state so long as changes in these variables were not made by some exogenous force.<sup>21</sup> He concludes that “Not the most cultivated coterie, not the most numerous proletariat, but a happy middle class shall inherit the earth” and that this is a conclusion reached not only by utilitarianism but also purportedly by common sense (78).<sup>22</sup> The implication here is that the ideal state of social well-being is often one which tends to be, for instance, normally distributed with respect to wealth and the availability of resources, and whose mean value is the social well-being of the middle class. However, he does not restrict this to normal distributions and instead leaves open the exact mathematical structure of the relevant measure of dispersion from central tendency.

One might object at this stage that while Edgeworth’s views were sophisticated and may have some plausibility, they are impractical to implement. However, Edgeworth was self-aware that his proposal suffers from issues of practicality. Rather, he aimed to provide a theoretically adequate framework and starting point from which further

---

<sup>20</sup>Unfortunately, Edgeworth is silent on this detail, as it is not at all obvious how one might quantify as a real number the threshold of existence for a sentient being. However, it is clear from surrounding passages in the text that he does not mean that the sentient is so devoid of well-being (e.g. depressed) such that it commits suicide; perhaps he has in mind something more like extreme poverty so debilitating that the sentient cannot live at all due to illness. Nonetheless, it is ultimately unclear from the text what he has in mind.

<sup>21</sup>This notion of equilibrium states would be explored further in Edgeworth’s later work *Mathematical Psychics* (1881). See also Yee (2021) for further discussion on Edgeworth’s account of equilibrium in his applications of concepts from mathematical physics to economics

<sup>22</sup>It is worth wondering to what extent this passage suggests that Edgeworth was a methodological predecessor to the field known as ‘eupathics’, started in 1917 by Abraham Myerson (Myerson 1917), insofar as Edgeworth was concerned also with an analysis of the happiness of normally situated people, as opposed to those who had symptoms of some hedonic pathology. Indeed, Myerson was a eugenicist who had a view of happiness and pathology from which to contrast the feeble with the normal (Angner 2011). However, it is beyond the scope of this paper to investigate this possible connection.

inquiry can proceed (66):

With what success could mathematical calculation address itself to social phenomena, when it is unable to cope with the problem of three bodies...mathematics [is] capable of advancing victoriously, even while leaving impregnable fortresses in the rear. . . [I]t might be hoped that approximative methods would be attainable, if a sufficiently clear and appropriate conception of the data were obtained.

Notice here that he believed that it was merely a lack of data that is the primary reason for the inability to calculate social well-being functions' values. Other economists of his time, such as Jevons, would make similar remarks (Jevons [1871] 1970, 21):

There is no reason whatever why we should not have those statistics, except the cost and trouble of collecting them, and the unwillingness of persons to afford information. The quantities themselves to be measured and registered are most concrete and precise.

This leads Edgeworth to the last species of social well-being problem: find an allocation of resources such that those who have the greatest capacity for pleasure can be most cultivated and supported by others *less* capable. He employed a poetic metaphor noting how different lamps have different capacities for illuminating, given a specific quantity and quality of fuel (74):

[S]entients being regarded as so many lamps of different lighting power, the questions have been what lamps shall be lit, and how much material shall be supplied to each lamp, in order to produce the greatest quantity of light...a large portion of the material to be distributed is applied not to be burned by the lamp, but to construct and repair it.

It is not uncharitable to read into his idea the thought that those with a higher capacity for pleasure, such as society's elites, are to be afforded the privileges of greater resources, and that the poor and less mentally fit are supposed to maintain the privileges of the elite in this way. In this sense, Edgeworth was an advocate of eugenics and an apologist for the elite, reinforcing the *status quo* of unequal socioeconomic conditions in Victorian society.<sup>23</sup> As described in earlier sections, the variant of eugenics he advocated was one which applied across all animal species, including human beings, insofar as all sentient beings have different capacities for pain, pleasure, cognition, and happiness, and therefore capacity for, and mathematical structure of, their social well-being. This hierarchy

---

<sup>23</sup>It is worth noting however that Edgeworth was not wholly against social services and would eventually write a 1923 paper entitled 'Women's Wages in Relation to Economic Welfare'. This article defended circumstances in which men ought to unionize such that each provides a portion of their income to subsidize women, who he claimed tended not only to be incapable of making as much of an earning (due to their alleged, comparative physical weakness) but who tended to also be impoverished (Edgeworth 1923). See also Chassonnery-Zaïgouche & Cot (2021) for further discussion and analysis.

of order is conditioned by both nature and contingent circumstance. While it is an important area of scholarship what exactly Edgeworth's eugenic views were, it is beyond the scope of this paper to elucidate his views beyond what has been discussed above for the specific purposes of understanding his views on mathematizing social well-being.

## 4 Foreshadowing Contemporary Well-Being Studies

Having outlined Edgeworth's complete theory of how social well-being ought to be mathematized and applied to a wide variety of social well-being problems, I now proceed to outline more explicitly how his theories of well-being foreshadowed several contemporary models, concepts, and issues raised by recent scholars in the well-being literature. I focus in particular on ideas that were, as far as I know, unique to Edgeworth's vision of social well-being, as opposed to those which were generally shared with others of his time, and the extent at which they reappear in contemporary scholarship.

There were essentially four core insights that Edgeworth had that are relevant to contemporary work: that there could be a unified mathematical treatment of all sentient beings' well-being, both understood individually and understood in social groups (Universalism), that social well-being ought to be analyzed intergenerationally (Intergenerationalism), that social well-being is contingent on the physiological capacities and features of sentient beings (Relativization), and that social well-being metrics require non-linear mathematical terms to express the non-linear effects of social well-being, as distinguished from linear aggregates of individual subjective well-being (Non-Linearity).

Edgeworth's Universalism has been echoed in recent scholarship on social well-being, though his eugenics views, let alone any eugenics view, have not appeared in recent models of social well-being and are typically taken by contemporary scholars as either methodologically irrelevant, uselessly false, or socially harmful. Universalist philosophies of social well-being have appeared in philosophy in the recent work of Sumner (1996, 14): "[O]ur welfare vocabulary applies just as readily to children and infants, and to many non-human beings. It is perfectly natural for me to say that my cat is doing well...I am applying exactly the same concept of welfare to my cat that I habitually apply to my friends. A theory of welfare will therefore also be incomplete if it covers only them and ignores her." There are also recent traces of Universalism in the attempts by the Happy Planet Index organization (Jeffrey et al. 2016) and Bhutan's Gross National Happiness metric (Centre for Bhutan Studies & GNH Research, 2016) to incorporate the intrinsic value of the environment, insofar as it is an object of appreciation and utility for sentient beings, and should be calculated when considering the social well-being of sentients. Dasgupta (2021, 115-123) also provides an economic analysis of calculations of social well-being while incorporating the intrinsic value of other sentient beings, and surrounding ecosystems.

The ethos of Edgeworth's Intergenerationalism has appeared as a core assumption of the landmark 2018 *UN Inclusive Wealth Report*. The authors call for measuring 'inclusive wealth': the sum total of manufactured capital (human constructed materials, MC), human capital (knowledge, health, happiness, HC), and natural capital (earth's

naturally occurring and endowed physical materials,  $NC$ ). The measurement unit used is ‘shadow prices’: “the monetary measure of the contribution a marginal unit of that asset is forecast to make to human well-being” (Managi et al. 2018, 107). More precisely, they posit the following model (Managi et al. 2018, vii). Let  $V(t)$  be the intertemporal well-being of a society at a time  $t$ ,  $\delta$  a discount factor, where  $\delta, t$  are positive real numbers,  $U$  is a utility function, and  $C$  is aggregate consumption of  $MC$ ,  $HC$ , and  $NC$ .  $V(t)$  is then defined as:

$$V(t) = W(MC, HC, NC, t) = \int_t^{\infty} U(C_{\tau})e^{-\delta(\tau-t)}d\tau$$

for convergent values of  $V(t)$ .  $V$  is further defined as *sustainable* only if  $\frac{d}{dt}V(t) \geq 0$ , which symbolizes a steady state. The term in the exponent represents a discount factor denoting the extent at which we might care about the social well-being of future generations in a given calculation. Indeed, the very idea of a sustainable, steady state that a society might reach is something Edgeworth took seriously in his calculations of social well-being: “But if the threshold increased with evolution, then we should tend to a ‘stationary state,’ not only wealth and number, but also, what Mill hardly contemplated, cultivation, evolution, stationary” (1877, 78).

Edgeworth’s Relativization foreshadows a philosophical theory from Davis (1981). While he is principally concerned with non-relational states of happiness, Davis nonetheless developed upon the core idea behind Edgeworth’s ‘relational account of enjoyment’ by providing a mathematical theory which combines beliefs, desires, and thoughts (e.g. the happiness derived from thinking of a loved one) especially when directed towards others. His theory employs a simple aggregational framework which produces a net sum of an individual  $A$ ’s desires to have the thoughts and beliefs that  $A$  has over a period of time  $t$ . More recently, Alexandrova (2017) develops a view according to which there is no adequate, scientific conception of well-being except that which relativizes the notion of well-being to specific classes of people (e.g. child well-being, soldier well-being, parental well-being, etc.) and linguistic contexts (i.e. the referent of the term ‘well-being’ has different meanings in different situations). This is because there are non-linear effects pertaining to social well-being that are best understood contextually and locally, and that depend on the role each person plays in society. Though Edgeworth was similarly a relativist about social well-being, this did not mean he was a subjectivist per se. Rather, he believed that there were objective facts about the individual and social well-being of individuals as contingent upon purely physiological and social factors pertaining to the hedonic and mental states of sentient beings, independent of whether such sentient beings had preferences which were satisfied or failed to be satisfied. This distinction between objective and subjective well-being remains debated in contemporary literature from the perspective of economics (Krueger et al. 2009; Fleurbaey & Blanchet 2013), geography (Voukelatou et al. 2021), sociology (Wish 1986; Western & Tomaszewski 2016), and especially in philosophy (Alexandrova 2017; van der Deijl 2017).

Fourthly, I have shown in this paper that Edgeworth’s views on Non-Linearity are complex in that there are certain circumstances in which linearity of a well-being metric

is justified and others in which it is not. Recall that for Edgeworth, pleasure and pain within a single sentient can be offset with one another in certain circumstances, and that this is possible even for ostensibly incommensurable complex cognitive states such as the pleasure from receiving a job promotion to the aggregate hedonic pleasure from consuming multiple ice creams (Edgeworth 1877, 26). This idea has appeared in the work of psychologists Kozma et al. (1990, 123) who proposed the following method of looking at aggregate happiness, combining data from short term mental states (subscript ‘s’) and long term states (subscript ‘l’):  $\text{Happiness} = (PA_s - NA_s) + (PA_l - NA_l) + (E_s + E_l)$ , where  $PA$  and  $NA$  are positive and negative affect, and  $E$  is the measurement error. Notice that this account of aggregation presumes that positive and negative affect lie on a single, linearly ordered scale such that each can be compared with one another, and is therefore linear. It is therefore coherent, according to this view, that positive and negative affect could possibly ‘cancel out’ one another so as to produce a neutral affective state in a linear fashion.

However, Edgeworth believed that there is an intrinsic non-linear structure to social well-being, even if there might be certain linear properties to individual well-being in certain circumstances. It is worth making this more precise in noting that as concerns a single individual, there may nonetheless be nonlinear physiological causes that lead to an individual’s well-being if that individual’s cognitive and physiological system is structured in the right way (Cf. section 2). As argued recently by Wodak (2019), there are good reasons to think that individual well-being is not linear. Furthermore, depending on the social context in which each individual is operating, one can then either linearly sum each individual’s well-being (each of which may or may not have linearly constituted well-being) or non-linearly calculate their social well-being if these individuals engage with one another in some specific manner, such as humans forming friendships or animals forming herds in mutual advantage. For instance, recent empirical psychological work has suggested the non-linear effects of social interaction on social well-being (Ren et al. 2022) and mesoscopic studies on the social well-being of cities, such as the role played by the presence of green spaces and population size (Zhao et al. 2019). Magee et al. (2012) argue from a questionnaire study on people from the Middle East, South and South East Asia between 2006 and 2010 that the social relationships of cohesion, stability, and integration cannot be reduced to linear factors and are non-linear relationships in which the total is greater than the sum of its parts.

To close this section, despite significant traces of ideas and issues that Edgeworth endeavored to solve in *Methods* reappearing in contemporary scholarship on social well-being, I could not find a single modern study or scholar who specifically utilized the ‘calculus of variations’ for the analysis of social well-being. Furthermore, I was unable to find a single direct reference to Edgeworth in the contemporary empirical social well-being literature within the past 30 years, despite some discussion occurring in philosophy (Cf. Kaminitz 2013). This is not necessarily surprising and I offer several hypotheses for why this is the case.

Firstly, it is possible that the mathematics is prohibitively challenging for most without a background in modern physics, which is the area in which the ‘calculus of varia-

tions' is most commonly used; modern psychologists and philosophers most interested in well-being are not typically trained with this specific kind of mathematical formalism. Furthermore, it is rare that anyone with such a background would also have a background in philosophy and economics such that they would be able or interested in attempting to interpret, appreciate, or evaluate Edgeworth's work on the subject, especially considering *Methods* made approximately zero noticeable impact on contemporary social scientific thought compared with his more famous *Mathematical Psychics*, which impacted modern economic thinking.

Secondly, there are considerable practical problems with applying exact utilitarianism to real-life cases, as Edgeworth leaves it unclear how to operationalize each of the quantities in his formalisms, opting to instead leave his models as a branch of pure value-theory. Hence, those who may have discovered his work might have found it either impenetrable, useless, or simply false, rather than philosophically edifying. And yet, despite several practical and theoretical issues I have noted, Edgeworth had several redeemable ideas which foreshadowed contemporary thinking in social well-being studies. This therefore leaves it unclear why exactly he was ignored in this particular sense.

Thirdly, contemporary economists, in particular, have preferred to use other mathematical formalisms to study similar kinds of problems that Edgeworth was concerned with, even if they are not the exact type of scenarios Edgeworth envisioned. For example, in a comprehensive analysis of the mathematics of social well-being calculation, Adler (2012) defines 'social welfare functions' as sensitive to the distribution of resources, defends a subjective account of social well-being as preference satisfaction, and shows how we can make policy decisions using a combination of ordinally structured preferences, survey data from questionnaires, and ordinary arithmetical operations using utility functions to determine the distribution of resources amongst competing uses. This method has its virtues and is considered a mainstream framework in contemporary welfare economics. It is therefore highly unclear that the added mathematical complexity of using the 'calculus of variations' yields a net positive methodological impact compared to less sophisticated mathematical methods.

Lastly, *Methods* is a highly challenging text containing literary references, miscellaneous poetic phrases he employs in other languages (e.g. Greek, Latin, German, etc.), thoughts which are sometimes inconclusive and inarticulate, and has a rhetorical style akin to the late night musings of an erudite polymath than of a traditional academic treatise on either economics or philosophy. This is not to downplay the originality and depth of the text but it does provide steps towards answers as to why this text's contents has been neglected. This being said, most of Edgeworth's other ideas were at the very least implicitly discussed even if he was rarely, if ever, directly cited.

## 5 Conclusion

Edgeworth's mathematization of social well-being consisted primarily in using the 'calculus of variations' to calculate minima and maxima for functions of hedonic value definable via several operationalizable variables, while simultaneously consulting the psychological

and proto-evolutionary theories of his time. While no mainstream scholar of well-being advocates usage of the ‘calculus of variations’ for the analysis of social well-being, other aspects of Edgeworth’s philosophy are manifested in contemporary methodological discussions and literature. Edgeworth’s formalisms concern two distinct but overlapping projects: (1) the calculation of the ideal distribution of a given set of resources so as to ensure the greatest aggregate subjective well-being given certain conditions, using methods in the ‘calculus of variations’; (2) the calculation of the aggregate social well-being of a given group of sentients using a diversity of mathematical methods. Taking the view, now widely considered to be false, that biological evolution generates an intrinsically objective hierarchy of world cultures’ value, Edgeworth employed a quantitative parameter he calls ‘the order of evolution’ as a variable in his social well-being formalisms. Exact utilitarianism was understood to be a recursive procedure insofar as one can define a composite function consisting of variables representing others’ evolutionarily endowed sensory apparatuses, ancestral habits, and order of evolution, to produce a general theory of social well-being that he alleges could be empirically discernible through psychophysical procedures. Empirical psychology is the ultimate ground and evidential basis for measurements of social well-being, rendering his view diametrically opposed to the skepticism many philosophers of his time held towards naturalistic methods.

Edgeworth is most remembered by posterity for his theory of indifference curves, his contributions to statistical inference, his poetic metaphors and analogies of concepts, and his constant employment of theories from physics in his models of the economy and social well-being. And yet, hidden within the florid and sometimes cryptic passages of *Methods* lies an attempt at providing a comprehensive account of measuring social well-being that foreshadows several contemporary methods and issues in social well-being research. In these ways, Edgeworth was a sophisticated methodological predecessor to many of the methods in contemporary social well-being research in advocating such views as Universalism, Intergenerationalism, Relativization, and Non-Linearity, topics which we continue to study today.

## References

- [1] Adler, Matthew. 2012. *Well-Being and Fair Distribution Beyond Cost-Benefit Analysis*. New York: Oxford University Press.
- [2] Alexandrova, Anna. 2017. *A Philosophy for the Science of Well-Being*. New York: Oxford University Press.
- [3] Angner, Erik. 2011. The evolution of eupathics: The historical roots of subjective measures of wellbeing. *International Journal of Wellbeing*. 1.1:4-41.
- [4] Baccini, Alberto. 2009. F. Y. Edgeworth’s Treatise on Probabilities. *History of Political Economy*. 41.1:143-162.



- [5] Barbe, Luis. 2010. *Francis Ysidro Edgeworth: A Portrait with Family and Friends*. Cheltenham: Edward Elgar Publishing Limited.
- [6] Barratt, Alfred. 1869. *Physical Ethics*. London: Williams and Norgate.
- [7] Brülde, Bengt. Krister Bykvist. 2010. Happiness, Ethics, and Politics: Introduction, History and Conceptual Framework. *Journal of Happiness Studies*. 11:541-551.
- [8] Centre for Bhutan Studies & GNH Research. 2016. A Compass Towards a Just and Harmonious Society 2015 GNH Survey Report. Centre for Bhutan Studies & GNH Research.
- [9] Chang, Hasok. 2004. *Inventing Temperature*. Oxford: Oxford University Press.
- [10] Chassonnery-Zaïgouche, Cleo. Annie L. Cot. 2021. Sentiment and Prejudice: Francis Ysidro Edgeworth on Women’s Wages. *History of Political Economy*. 53.5: 799-832.
- [11] Clark, Andrew. S. Fleche. R. Layard. N. Powdthavee. G. Ward. 2018. *The Origins of Happiness*. Princeton: Princeton University Press.
- [12] Colander, David. 2007. Edgeworth’s Hedonimeter and the Quest to Measure Utility. *The Journal of Economic Perspectives*. 21.2:215-226.
- [13] Creedy, John. 1984. Edgeworth: utilitarianism and arbitration. *History of Political Economy*. 16.4:609-618.
- [14] Cronin, Thomas., Roy Caldwell., Justin Marshall. 2001. Tunable colour vision in a mantis shrimp. *Nature*. 411:547-548.
- [15] Dasgupta, Partha. 2021. *The Economics of Biodiversity: The Dasgupta Review*. London: HM Treasury.
- [16] Davis, Wayne. 1981. A Theory of Happiness. *American Philosophical Quarterly*. 18.2:111-120.
- [17] Easterlin, Richard. 1974. Does Economic Growth Improve the Human Lot? Some Empirical Evidence. Appears in *Nations and Households in Economic Growth. Essays in Honor of Moses Abramovitz*. 89-125.
- [18] Edgeworth, Francis. 1877. *New and Old Methods of Ethics*. Oxford: James Parker and Co.
- [19] ————— 1881. *Mathematical Psychics*. London: C. Kegan Paul & Co.

- [20] ————— 1923. Women’s Wages in Relation to Economic Welfare. *The Economic Journal*. 33(132):487-495.
- [21] Fleurbaey, Marc. Didier Blanchet. 2013. *Beyond GDP: Measuring Welfare and Assessing Sustainability*. New York: Oxford University Press.
- [22] Fraser, Craig. 2003. The Calculus of Variations: A Historical Survey. Appears in *A History of Analysis*. Ed. H. N. Jahnke. *American Mathematical Society*. 355-384.
- [23] Hurka, Thomas. 2014. *British Ethical Theorists from Sidgwick to Ewing*. Oxford: Oxford University Press.
- [24] Jeffrey, Karen. Hanna Wheatley. Saamah Abdallah. 2016. The Happy Planet Index. *New Economics Foundation*. URL: <https://happyplanetindex.org/hpi>
- [25] Jellett, John. 1850. *The Calculus of Variations*. London: William S. Orr and Company.
- [26] Jevons, William. [1871] 1970. *The Theory of Political Economy*. Middlesex: Penguin Classics.
- [27] ————— [1888] 2010. *Elementary Lessons in Logic: Deductive & Inductive*. Auburn: Ludwig von Mises Institute. Print.
- [28] Kaminitz, Shiri. 2018. Happiness Studies and the Problem of Interpersonal Comparisons of Satisfaction: Two Histories, Three Approaches. *Journal of Happiness Studies*. 19:423-442.
- [29] ————— 2013. Economics and Ethics under the Same Umbrella: Edgeworth’s ‘Exact Utilitarianism’, 1877-1881. *Utilitas*. 25.4:487-503.
- [30] Kozma, A., S. Stone, M. Stones., T. Hannah., K. McNeil. 1990. Long- And Short-Term Affective States in Happiness: Model, Paradigm and Experimental Evidence. *Social Indicators Research*. 22:119-138.
- [31] Krueger, Alan. et al. 2009. *Measuring the Subjective Well-Being of Nations*. Chicago: University of Chicago Press.
- [32] Levine, Philippa. 2017. *Eugenics: A Very Short Introduction*. Oxford University Press.
- [33] Magee, Liam. Andy Scerri. Paul James. 2012. Measuring Social Sustainability: A Community-Centred Approach. *Applied Research Quality Life*. 7:239-261.

- [34] Managi, S., K. Pushpam, P Dasgupta. et al. 2018. *Inclusive Wealth Report*. United Nations Environment Programme.
- [35] Mirowski, Philip. 1994. *Edgeworth on Chance, Economic Hazard, and Statistics*. 1st edition. Rowman & Littlefield Publishers.
- [36] Mueller, Thomas Michael. 2020. The Temperature of the Brain: Edgeworth's Thermodynamic Analogies of Utility Measurement. *History of Political Economy*. 52.4:709-740.
- [37] Nagel, Thomas. 1974. What Is It Like To Be A Bat? *The Philosophical Review*. 83.4:435-450.
- [38] Punch. 1877. *Punch, Or the London Charivari*. 73:iii-312.
- [39] Ren, Dongning. Olga Stavrova. Wen Wei Loh. 2022. Nonlinear effect of social interaction quantity on psychological well-being: Diminishing returns or inverted U? *Journal of Personality and Social Psychology*. 122.6:1056-1074.
- [40] Schumpeter, Joseph. [1954] 2006. *History of Economic Analysis*. Taylor & Francis.
- [41] Sidgwick, Henry. [1874] 1862. *The Methods of Ethics*. Toronto: Palgrave Macmillan.
- [42] ————. 2000. *Essays on Ethics and Method*. Edited by M. Singer. New York: Oxford University Press.
- [43] Stewart, James. 2012. *Calculus: Early Transcendentals*. 7th ed. Belmont: Brooks/Cole.
- [44] Sumner, Wayne. 1996. *Welfare, Happiness, and Ethics*. New York: Oxford University Press.
- [45] van der Deijl, W. 2017. Are Measures of Well-Being Philosophically Adequate? *Philosophy of the Social Sciences*. 47.3:209-234.
- [46] Voukelatou, Vasiliki. Lorenzo Gabrielli. Ianna Miliou. et al. 2021. Measuring objective and subjective well-being: dimensions and data sources. *International Journal of Data Science and Analytics*. 11:279-309.
- [47] Wall, Byron E.. 1978. F. Y. Edgeworth's Mathematical Ethics: Greatest Happiness with the Calculus of Variations. *Mathematical Intelligencer*. 1:177-181.

- [48] Western, Mark. Wojtek Tomaszewski. 2016. Subjective Wellbeing, Objective Wellbeing and Inequality in Australia. *PLoS ONE*. 11.10:1-20.
- [49] White, Nicholas. 2006. *A Brief History of Happiness*. Wiley-Blackwell.
- [50] Wish, Naomi Bailin. 1986. Are We Really Measuring the Quality of Life? Well-Being Has Subjective Dimensions, as Well as Objective Ones. *The American Journal of Economics & Sociology*. 45.1:93-99.
- [51] Wodak, Daniel. 2019. What If Well-Being Measurements Are Non-Linear? *Australasian Journal of Philosophy*. 97.1:29-45.
- [52] Yee, Adrian K. 2021. Econophysics: making sense of a chimera. *European Journal for Philosophy of Science*. 11:1-34.
- [53] Zhao, Yukun. Feng Yu. Bo Jing. et al. 2019. An Analysis of Well-Being Determinants at the City Level in China Using Big Data. *Social Indicators Research*. 143:973-994.