

Jeremy Bentham, ‘The Psychology of Economic Man’, and Behavioural Economics

Michael Quinn (Bentham Project, UCL. m.quinn@ucl.ac.uk)

Résumés

English Français

§ 1 briefly reviews first the received interpretation of Bentham, which sees him as having had little to do with the development of economics (excepting some passing mentions which recognize his deployment of the concept of utility or his reduction of human motivation to self-interest, and perhaps a note on his discussion of the concept of diminishing marginal utility); and second, the manner in which he applies his concept of rationality to political economy. In § 2, the central thesis of the paper is presented: it is argued that an examination of his insights into the psychology of individual choice supplies good reasons to identify him as an intellectual godfather of behavioural economics. In keeping with the normativity of his concept of rationality, Bentham would maintain that the way in which traditional economics continues to ignore the gulf between its model of human decision-making and the facts of human psychology weakens its usefulness both as a science and as a guide to public policy. Bentham anticipated several modifications to the standard model (for instance loss-aversion, the endowment effect, reference dependence, framing, the desire for cognitive ease, and status-quo bias) which have been introduced later by behavioural economics. § 3 introduces two problems concerning the normativity of economics, the first of which, at least for Bentham, rests upon a false premise, while his substantive notion of rationality insulates him against the second.

Français

§ 1 revoit en bref premièrement l'interprétation standard de Bentham, qui le voit comme ayant peu à faire avec le développement des sciences économiques (excepté des mentions passagères qui reconnaissent son élaboration du concept de l'utilité ou sa réduction de la motivation humaine à l'intérêt personnel, et peut-être une note sur sa discussion du concept de l'utilité marginale décroissante), et deuxièmement, la manière avec laquelle il applique son concept de la rationalité à l'économie politique. § 2 présent la thèse centrale de cet article: Il est proposé qu'une examination des perspicacités de Bentham vis-à-vis la psychologie de choix individuel fournit des bonnes raisons pour l'identifier comme un parrain intellectuel de l'économie comportementale. Conforme à la normativité de son concept de la rationalité, Bentham affirmerait, que la manière avec laquelle l'économie traditionnelle continue à ignorer le fossé entre le modèle traditionnel du processus de la prise de décision humaine et les caractéristiques de la psychologie humaine, affaiblit son avantage non seulement comme science, mais aussi comme guide aux politiques publiques. Bentham avait anticipé quelques modifications au modèle standard (par exemple l'aversion à la perte, l'effet de dotation, la dépendance à la référence, le cadrage, le désir de l'aisance cognitive, et la tendance au statu quo) qui ont été introduits plus tard par l'économie comportementale. § 3 introduit deux problèmes concernant la normativité de l'économie, dont le premier, au moins pour Bentham, dépend d'une fausse hypothèse, alors que son idée fondamentale de la rationalité le défend contre le deuxième.

Mots clés: Bentham, économie comportementale, rationalité, aversion de perte, cadrage, nudge.

Keywords: Bentham, behavioural economics, rationality, loss aversion, framing, nudge.

Introduction

Jeremy Bentham is celebrated as the founder of classical utilitarianism, but his influence on the development of economic thought has often been considered negligible (Schumpeter 1954, 128–9).¹ In general terms, Bentham’s deployment of the language of utilitarian calculation, his assumption of self-interested maximizing rationality, and his extensive discussion of diminishing marginal utility (Bentham 1843, i. 304-7; iii. 228–30) has led to some recognition of his (indirect) influence, through the later agency of Jevons (1970 (1871)) and Edgeworth (1967 (1881)), on the development of economics as a science grounded in the psychology of man the utility maximizer. There has also been some discussion, derived from examination of Bentham’s unpublished manuscripts, of the way in which he prefigured the later deployment of both willingness to pay, and of indifference between pleasures—or combinations of pleasures and pains—as indicative of equality in value, which lie at the heart of contemporary cost-benefit analysis and micro-economics respectively.

¹ Although Bentham invested considerable time and effort writing on political economy, it was not until the 1940s that any attempt was made to produce a critical edition of his works in this area (Bentham 1952–4). It is perhaps unsurprising that the most popular fruit of Werner Stark’s herculean labours in producing this edition should prove to be the two texts therein contained which are farthest from being coherent works of Bentham. Scholars are understandably drawn to Stark’s editorial constructions, ‘The Philosophy of Economic Science’ and the ‘The Psychology of Economic Man’ (Bentham 1952–4, i. 79–119 and iii. 419–50), despite the fact that Bentham wrote no such works, because they offer the prospect of accessing the essence of Bentham’s thought on political economy without ploughing through the mass of words in which these nuggets are contained. In order to create ‘The Psychology of Economic Man’ Stark drew on sixteen separate works, none of which were ostensibly concerned with political economy.

(UC n.d. xxvii. 36–7; Baumgardt 1952; Goldworth 1979; Guidi 2000; Warke 2000a).²

This paper will argue that an examination of Bentham’s insights into the psychology of individual choice provides good reasons to identify him as an intellectual Godfather of behavioural economics. Modern mainstream economics faces a problem in that, as demonstrated by Behavioural economics, its model of economic rationality does a very poor job of predicting human behaviours in a complex world. For Bentham, the ‘science’ of political economy was of no use unless it helped to inform the ‘art’ of legislation, and thus contributed to increases in well-being. Bentham’s political economy was always a matter of praxis, of the application of the best available theory to the resolution of practical problems. Frugality in the multiplication of theoretical presuppositions was generally a good thing, but not if it issued consistently in inaccurate predictions. In seeking to guide legislative practice, Bentham developed what he termed ‘axioms of mental pathology’, effectively generalizations about human cognitive functioning, which, as noted in § 2, feature anticipations of central conclusions of behavioural economics.

1. Bentham, rationality and economics

1.1 Rationality: formal or substantive

In modern economics, the rationality of economic choices is simply assumed. No attempt is made to measure utility, the presumed end of action, directly. Instead, the rationality of choice is inferred via the assumption of complete and transitive preference orderings between goods, derived from their predicted utility: ‘the proposition that people maximize utility was granted the status of a maintained hypothesis, which is used to constrain the interpretation of other facts but is not itself

² For fuller discussions of Bentham’s attempts to address the difficulties involved in measuring the

subject to test.’ (Kahneman, Wakker & Sarin 1997, 397) Bentham would have been content with this hypothesis, subject to two significant qualifications. First, he would insist that utility be understood in terms of the desire to attain pleasure or avoid pain: ‘my notion of man is, that, successfully or unsuccessfully, he aims at happiness ... in every thing he does.’³ (Bentham 1977, 67) Second, as the quotation makes clear, he would recognize that the achievement of this universal goal could not be unproblematically assumed: we often make mistakes.

The evolution of the concept of utility in the history of economic theory has been charted, and the reasons for the displacement of happiness as the substantive goal of rational action insightfully analyzed, by skilful commentators (Bruni & Sugden 2007; Sent 2004; Read 2007; Warke 2000a), and will not be addressed in detail here. As Bruni and Sugden have argued, part of Pareto’s endeavour was precisely to disassociate economics from psychology, and to ground it entirely on ‘principles of abstract rational choice’ (2007, 171), while the core characteristic of rationality is consistency: ‘In modern versions of the theory of choice, axioms of consistency of preferences are usually justified as formal principles of rationality.’ (Bruni & Sugden 2007, 170) Bentham’s model of rationality differs radically from the minimalist formal version offered by contemporary theory, in which the relation between rationality and consistency comes close to identity: ‘Rationality is logical coherence—reasonable or not’⁴ (Kahneman 2011, 411; see also Read 2007, 46; Morewedge in press). Bentham would agree that rational choice was characterized by logical coherence, so that rational choices are consistent and transitive (where consistent means instrumentally connected to the same end: happiness); but he would

value of pleasures and pains see Warke 2000b, Guidi 2007, Quinn 2014a.

³ This is the first extract deployed by Stark in ‘The Psychology of Economic Man’ (Bentham 1952–4, iii. 421).

⁴ For rationality as consistency plus maximization, see Becker 1996, 23.

not insist that preferences are simply given, and remain constant even while constraints change. Bentham certainly recognized that human individuals often make mistakes in their attempts at rational choosing: ‘such is human nature: any thing rather than consistency’ (UC n.d. lxii.188), but might also note that our failure to be consistently rational does not destroy the value of rationality as a standard of evaluation, or as a goal.

For Bentham, consistency was a necessary condition for rationality, but certainly was not a sufficient one, since rationality had an essential normative component. In contrasting his own approach to political economy with that of Adam Smith, Bentham argued (somewhat inaccurately)⁵ that Smith had written principally of the science, the *το ον*, or ‘what is’, whereas his own concern was primarily with the art, the *το πρεπον*, or ‘what ought to be’ (UC n.d. xvii. 14 (1952–4, i. 224)). Art and science for Bentham were two complementary descriptions of a single body of knowledge (UC n.d. lxix. 197). In keeping with his utilitarian perspective, the normative branch of the enquiry was the more important of the two (UC n.d. ci. 153 (1838–43, viii. 233)). What any would-be legislator might reasonably demand from Bentham’s art of legislation is an objective criterion for the evaluation of law and policy, to serve as a key to ‘the art of knowing what *ought* to [be] done in the way of internal government’ (UC n.d. lxix. 195).

1.2 Rationality and self-interest

Bentham’s is both a calculating rationality, and a maximizing one. For him, value – all value – depends on a single binary opposition between agreeable sensations (i.e. pleasures) on the one hand, and disagreeable ones (i.e. pains) on the other. Bentham takes the preference for pleasure over pain as the foundational premise of rationality.

By inference therefrom, a pleasure of greater value is preferable to one of lesser value: ‘Things are good or bad by comparison: good itself is bad, in comparison of better’.

(UC n.d. clxvi. 201) The assessment of quantities, calculation in a word, is for Bentham utterly integral to the meaning of rationality. The goal of the prudentially rational human agent is to maximize her net balance of pleasure over pain. Insofar as *homo oeconomicus* is assumed to be a self-interested utility maximizer, Bentham would recognize his own characterization of typical human motivation.⁶ Bentham asserted a connection between the pursuit of self-interest and the survival of the species (1983, 108), but did not assume that all motivation was egoistic. He recognized that sympathy for others, understood as desire for their well-being, could provide a motive for action, insisting only that since the pleasure I derived from knowledge of their well-being is mine, the desire to promote that well-being is motivated by the prospect of a pleasure which was likewise mine (Bentham 1983, 36). The range over which sympathy extended was variable, while the wider it extended, the closer would its dictates come to those of abstract utility. The task of the legislator was to harmonise the universal motive with the only acceptable moral end: ‘Self-regard actually and *properly* the universally *predominant* motive’; ‘According to utility, *proper* end: greatest happiness of greatest number. Actual end: each man’s own’ (Bentham 1983, 35, 37). The core task of the utilitarian legislator is to alter self-interested individual calculations of consequences to unite the two, by bringing interest into line with duty (Bentham 2010a, 353).

In relation to political economy, since wealth was an all-purpose ticket to the instruments of happiness, Bentham endorses the assumption that individuals in

⁵ For the view that Smith’s political economy formed a consistent whole with his moral and legal philosophy see Haakonssen 1981, and Winch 1983, 1992, 1996.

⁶ Of the thirty-four extracts contained in Stark’s compilation ‘The Psychology of Economic Man’, almost half contain the assertion of the prevalence of self-interest as a motive.

general desire to maximize their wealth (and thereby their happiness): ‘To get money is what most men have a mind to do: because he who has money gets, as far as it goes, most other things that he has a mind for.’⁷ (1787, 97 (1952–4, i. 157)) As already noted, the standard model takes the rationality of individual preferences as given, which is to say that individuals are always correct about what is best for them (about how, that is, to maximize their utility) (Thaler & Sunstein 2009, 10; Altman 2012, 92). For his part, Bentham follows Smith in endorsing this approach, at least to the extent of arguing at length that individuals are much more likely to be correct than government. In analysis of the best means to maximize national wealth, Bentham bases his conclusions on the relative position of individuals and governments in relation to the necessary conditions for voluntary human action, namely interest (I must want to do it), knowledge (I must know how to do it), and power (I must have the capacity to do it). With regard to interest, the conclusion is quickly reached (UC n.d. xvii. 211 (Bentham 1952–4, iii. 333)). By contrast, there may be cases where government is best placed to provide power, legal or financial, but the derivation of the latter from the pain-inducing source of coercive taxation places strict limits on the eligibility of this resource. There certainly are cases where government is in a position to promote the acquisition and diffusion of technical knowledge and empirical data (UC n.d. xvii. 244–6 (Bentham 1952–4, iii. 336–8)), and this exception will be further addressed below. However, the conclusion of Bentham’s analysis is that, if the goal of political economy is to maximize wealth, government is best advised to leave the vast bulk of the decision-making up to individuals (UC n.d. xvii. 218 (1952–4, iii. 337)).

⁷ See also UC n.d. xvii. 309 (Bentham 1952–4, iii. 323–4): ‘To add to his own particular stock ... is, with a very few exceptions, the constant aim and occupation of every individual, in every civilized nation.’ Of course, wealth did not constitute happiness, and the presumed desire to maximize wealth could be overridden by the value of other sources of enjoyment. As Bentham continued: ‘Leave men to themselves, each man is occupied either in the acquisition of wealth (the instrument of enjoyment) or in

§ 2 Bentham the behavioural economist?

Behavioural economics presents a challenge to the conventional model of economic man through the sheer bulk of the evidence it presents of the way in which human cognitive processes often bear little resemblance to the hypothetical reasonings of economic man: ‘the aim is to model economic agents in ways that take account of the affective responses that decision problems evoke in human beings and of the cognitive processes that are used in human decision-making’ (Bruni & Sugden 2007, 161). Economic theory ignores psychology at its peril. Camerer and Lowenstein remark in passing that many early economists ‘moonlighted as the psychologists of their times’ (2004, 5), but Bentham went further than most in explicitly recognizing the relations between psychology and political economy, and indeed between psychology and all human sciences. His own psychology was largely drawn from Helvetius, Beccaria and Hume, and was based on the associationist approach developed systematically by Hartley (1775). Bentham founded his prescriptions on psychological assumptions which he termed ‘axioms of mental pathology’. Mental pathology consisted in ‘the knowledge of the feelings, affections, and passions, and their effects upon happiness’ (Bentham 1843, i. 304), while its axioms were defined thus:

Axioms of mental pathology may be styled those most commonly applicable propositions by which statement is made of the several occurrences by which pleasure or pain is made to have place in the human mind:—as also the results observed to follow from the performance of such operations as have been

some actual enjoyment, which, in the eyes of the only competent judge, is of more value.’ (UC n.d. xvii. 242 (Bentham 1952–4, iii. 323–4))

performed ... for the purpose of effecting the augmentation of the aggregate of the pleasures, or the diminution of the aggregate of the pains (1843, i. 305)

Bentham's axioms bear close examination, not least because they reveal his anticipation of several assumptions incompatible with the standard model of economic rationality.

2.1 Loss Aversion, Status quo bias and Herd-following

Kahneman describes the phenomenon of loss-aversion as 'certainly the most significant contribution of psychology to behavioural economics' (2011, 300), while Camerer and Lowenstein recruit Adam Smith to the ranks of proto-behavioural economists on the basis of his recognition of loss-aversion as a widespread psychological characteristic (2004, 5).⁸ For his part, Bentham not only included in his axioms of mental pathology an explicit statement of loss-aversion, but utilized the concept as the very bedrock of his discussion of distributive justice in particular, and of the art of legislation in general. Loss aversion is the psychological root of the primacy which Bentham accords to the subordinate end of security. 'Mankind in general appear to be more sensible of grief than pleasure from an equal cause. For example a loss which would diminish the fortune of an individual by one quarter, would take more from his happiness than would probably be added by a gain which should double it.'⁹ (Bentham 1843, i. 307) Bentham traces the reasons for this psychological asymmetry to what behavioural economists call the endowment effect: 'A loss is almost always unexpected, because a man naturally hopes to keep what he possesses. This expectation is founded upon the ordinary course of things; for if we

⁸ Loss-aversion as a generic psychological propensity should be distinguished from individual attitudes (and particularly aversive attitudes) toward risk, variations in which are successfully incorporated into the standard model of economics.

⁹ Curiously enough, Bentham's rough estimate of the psychological rate of exchange between losses and gains turns out to be broadly vindicated by empirical findings: broadly speaking, people typically

look at the whole mass of men, they not only keep what they have acquired, but still further increase its amount.’ (1843, i. 306) And again, in arguing that, in legal cases between parties equally wealthy, happiness is more likely to be maximized by a distribution which favours the defendant against the plaintiff:

The loser experiences the pain of disappointed expectation: the other is simply in the condition of not having gained. But the negative evil of not having gained, is not equal to the positive evil of having lost. (If this were not the case, everyman would experience this evil with regard to every thing which he did not obtain, and the causes of evil being infinite, every one ought to find himself infinitely miserable). (1843, i. 307)

Bentham’s explanation of loss aversion thus depends on the special features of the pain of disappointment, which itself depends upon the centrality of expectation in his psychology (Kelly 1990; Guidi 2007).

This disposition to look forward, which has so marked an influence upon the condition of man, may be called expectation—expectation of the future. It is by means of this we are enabled to form a general plan of conduct; it is by means of this, that the successive moments which compose the duration of life ... become parts of a continuous whole. (1843, i. 308)

The pleasures of expectation are experienced utilities which arise from beliefs about future events. Expectation is the conviction that a future event will come to pass. Bentham recognizes that such beliefs are inescapably reference dependent, which means that my affective attitudes, and thereby my preferences, are influenced by my previous history and my current endowments. Since I very quickly become habituated to possession, or endowment, ‘[e]very thing which I actually possess, or which I

reject losses not compensated by the prospect of gains of twice the size (Kahneman & Shane 1990;

ought to possess, I consider in my imagination as about to belong to me for ever: I make it the foundation of my expectation' (1843, i. 310). This expectation is not only reinforced, but indeed created, solely by the force of law which penalizes theft. Expectation derived from possession is the explanation of our experience that 'the evil of loss is greater than the profit of gain' (UC n.d. xcix. 53), while loss contrary to expectation causes a specific pain of disappointment, which Bentham defines simply as 'expectation thwarted' (1795, 28 (1952–4, i. 290)).

According to standard models of economic rationality, loss aversion is irrational, whilst recognition of the prevalence of loss aversion allows behavioural economics to make more accurate predictions of behaviour than those models (Camerer & Lowenstein 2004, 4–5; Kahneman 2011, 283–3). Behavioural economists disagree on whether loss aversion is a cognitive error, which might be corrected with better education and reflection, or is a perfectly rational reaction to perceived threats to well-being (Rabin 1998, 34–5). For his part, Bentham does not question the rationality of the phenomenon, but simply accepts its reality to the extent of building it into the legislator's basic tool-kit of psychological presumptions. Late in life Bentham built this axiom into the 'Disappointment Prevention Principle' (Bentham 2011, 242 (1843, iii. 212)), but forty years earlier he had stated it under a different name: 'The principle of security ... directs that events, inasmuch as they are dependent upon the laws, should be conformed to the expectations to which the laws have given birth.' (1843, i. 308)

As noted above, Bentham explicitly recognized the phenomenon of diminishing marginal utility, and, in large part on the basis of its application to wealth,

Kahneman 2011, 284; Camerer & Lowenstein 2004, 15–16).

added equality to list set of subordinate ends of legislation, in addition to security, subsistence and abundance:

Thus it is, that if the effects of the first order were alone taken into account, the consequence would be, that, on the supposition of a new constitution coming to be established, with the greatest happiness of the greatest number for its end in view, sufficient reason would have place for taking the matter of wealth from the richest and transferring it to the less rich, till the fortunes of all were reduced to an equality, or a system of inequality so little different from perfect equality that the difference would not be worth calculating. (2011, 278 (1843, iii. 230))

Coercive redistribution in pursuit of equality is, however, ruled out by its pernicious effects in undermining security (and thereby abundance and subsistence). Indeed, with regard to the losers from retrenchments in public expenditure, Bentham's position was that they should be indemnified: 'no reform ought to be carried into effect without granting complete indemnity to those whose emoluments are diminished, or whose offices are suppressed' (Bentham 1825, 298 (1843, ii. 251)). Bentham thus anticipates later developments in welfare economics and cost-benefit analysis, but goes further than the received view, by insisting that to qualify as unequivocally beneficial, government policies should ensure not only that winners *could* fully compensate losers, as in the Kaldor-Hicks criterion (Bhagwati 1994; Scitovsky 1941), but that they *should* so do. The tenacity with which such losers will seek to defend their current endowments is anticipated by Bentham, and he recommends compensation as the route to overcoming it (Bentham 1825, 301 (1843, iii. 252)). The impact of loss aversion on Bentham's understanding of utilitarian justice, and on his strategies for

utilitarian reform, is pervasive and significant (Kelly 1990), and he would agree with Kahneman's that 'Loss aversion is a powerful conservative force' (2011, 305).

A further departure from economic rationality detailed by behavioural economists is status quo bias, which is closely connected with loss aversion and the endowment effect. The recognition of the ubiquity of loss aversion goes a long way to explaining the conservatism of some of Bentham's policy recommendations. Because of the centrality of expectations to the experience of pleasures and pains, and because the expression 'status quo' includes the entire constellation of existing expectations, practically every change in law and policy will involve the enhancement of some expectations and the diminution of others. The asymmetry between gain and loss makes it much more difficult than it might appear at first blush for the welfare gains from any proposed change to outweigh the welfare losses, and often the safest policy is to leave things as they are. Hence Bentham's advice to the legislator confronted with an established system of property rights: 'He ought to maintain the distribution which is actually established'. (1843, i. 311) It should be noted that there is an obvious tension in Bentham's thought between the conservative commitment to respecting established expectations, and the utilitarian commitment to reform of those expectations to deliver new benefits. The legislator has utilitarian reasons for respecting existing entitlements, but, insofar as she remains a utilitarian, she also has reasons for violating those entitlements, at least where they do not match the entitlements which would best serve to maximize utility. The policy implications of loss aversion and of diminishing marginal utility stand in direct opposition, while Bentham resolves the contradiction decisively in favour of the former.

Neither behavioural economics nor Bentham, however, advances loss aversion as a complete explanation of a general human preference for the devil you know. Both

identify two further related closely related cognitive biases, namely the preference for cognitive ease—itsself the consequence of the rebarbative, because pain inducing, nature of cognitive effort—and the unreflective nature of habit—the tendency to process information, that is to think, and therefore to act, automatically or by default—that is, literally, in the absence of thought. Kahneman notes the influence on human cognitive processes of the desire not to work too hard (2011, 59–70). For Bentham, the love of ease, prompted by aversion to painful labour, was one of the ‘most powerful, most constant, and most extensive’ of motives (1970, 155),¹⁰ which gave rise to what he called ‘the interest of the pillow’ (Bentham 1983, 85). Evidence of Bentham’s recognition of the influence of ‘the *vis inertiae* of the mind’ (UC n.d. lxxxvii. 175; 1827, ii. 218 (1843, vi. 446)), governed in part by our unreflective reliance on faulty, because lazy, reasoning, is widespread. To mention a single example, most of the argumentative work in the customary condemnation of lending money at high rates of interest was done simply by the attachment of the pejorative label ‘usury’ to the activity (Bentham 1787, 7 (1952–4, 130)). The infection of policy debates by the question-begging employment of dyslogistic appellatives was a phenomenon which Bentham saw everywhere, and which he viewed as a mischievous manifestation of the desire for cognitive ease, since it allowed the short-circuiting of proper investigation of facts. Unreflective endorsement of the status quo was an appeal not to evidence but to prejudice, that is precisely unevidenced assertion.¹¹ To defend a law, an institution, or a practice purely by an appeal to the authority of custom and habit was to abandon rationality completely (Bentham 1843, i. 241). Do

¹⁰ The equally universal motives were ‘physical desire, the love of wealth ... the love of life, and the fear of pain’.

¹¹ See Bentham 2015, 41: ‘Prejudice ... is a judgment, which being pronounced *before* evidence, is therefore pronounced without evidence.’

we tend uncritically to accept existing opinions when the public interest, and even our own private interests, are at stake?

Let Reason (that is utility) be fruitful and Custom barren is one of the fundamental aphorisms of the greatest genius that ever trod the earth.¹²

Unhappily custom, that blind guide, is the only one habitually consulted, the only one to whose suggestions men in general are disposed so much as to listen with complacency. *Man is a rational animal, say the logicians—true of one out of fifty thousand: man is an imitative animal—true of every body.*

(Bentham 2010a, 97–8n. My emphasis)

Unreflective habit thus garners further support from another human predilection well-established by experimental data, namely our ‘imitative’ nature, our desire to follow the herd. The fundamental desire for the approval of our fellows provided a foundational premise of Adam Smith’s moral theory (Smith 1976 (1759), 116), and Bentham too recognized it, insisting only that it was ultimately self-regarding: the benefits of fitting-in were not exclusively psychological, but consisted in the readiness of others to provide a flow of ‘good offices’, that is physical services. Popular opinion, and especially the opinion of that subset of the public with whom we interact on a daily basis (Bentham 2010a, 337n), wields the moral sanction, and punishes us for actions and attitudes which it disapproves: ‘A single individual is seldom able to withstand or change the laws established by public opinion’ (Bentham 1825, ii. 225 (1843, ii. 259)).

We have come a long way from the rational utility maximizer embodied in the standard model of economic reasoning, and from the economic model of rationality which, it has been argued, underpinned Bentham’s entire evaluative project.

¹² i.e. Sir Francis Bacon.

‘Happiness depends upon the correctness of the facts with which our mind is furnished, and the rectitude of our judgment’. (Bentham 1825, 69 (Bowring, ii. 213)) If our minds are furnished with spurious facts, and if our inferences from genuine facts are faulty, we will pay a cost in happiness forfeited. Does the legislator have a responsibility to furnish genuine facts, and to correct our judgments? Bentham sometimes sounded as if simply highlighting the divergence between popular opinion and the dictates of utility would suffice to eliminate it (1843, i. 324). However, in relation to a range of issues relevant to this paper—saving for the future for instance—he recognized a range of formidable obstacles to this happy outcome. His general answer to the question is, of course, that the propriety of government intervention depends on the outcome of a probabilistic calculation of the comparative costs and benefits of intervention.¹³ What is however clear, is that Bentham anticipated the insights of behavioural economics in relation not only to loss aversion, but to status quo bias, the desire for cognitive ease, and the prevalence of herd-following.

2.2 Wealth, Happiness, and the Easterlin Paradox

Easterlin’s paradox arises from the failure of large increases in national wealth to translate into roughly equivalent increases in national happiness, as recorded in responses to scientific surveys of opinion (Easterlin 1995; 2001). Although Bentham endorsed the notion that the end of political economy was the maximization of wealth, he noted also that ‘Happiness—enjoyment—not money, is or ought to be the ultimate object of the legislator’s care’ (UC n.d. xvii. 31). Whilst money was a general means to acquire the instruments of happiness, and whilst it was true that ‘of two persons having unequal fortunes, he who has most wealth must by a legislator be regarded as

¹³ For a related discussion of Bentham’s approach to dealing with oppositions between prejudiced

having most happiness' (Bentham 2011, 275 (1843, iii. 229)), it was no less true that the quantity of happiness would not increase in step with the quantity of wealth: 'It will even be matter of doubt, whether ten thousand times the wealth will in general bring with it twice the happiness. (Bentham 2011, 275 (1843, iii. 229))

Not only does wealth suffer from diminishing marginal utility, but once subsistence (the basic need for food, clothing and shelter) is secured, the battle for happiness is more than half won. Bentham advances two assertions in support of this proposition. First, the most productive sources of pleasure—crudely the pleasures of eating and drinking, of sex and of rest, and the vicarious pleasure of knowing that those we care for are enjoying the first three—are at the command of everyone whose subsistence is secure:

The principal enjoyments of which human nature is susceptible, constancy of repetition being considered as well as magnitude, are—those produced by the operations by which the individual is preserved; those produced by the operations by which the species is preserved; that cessation from labour which is termed repose; and that pleasure of sympathy which is produced by the observation of others partaking in the same enjoyments. (Bentham 1843, ix. 15)

Second, Bentham speculates that human nervous systems become less sensitive to sensory inputs as a result of habitual experience of them:

[T]he quantity of felicity actually enjoyed depends on the degree of sensibility to enjoyment, in each instance: and while in the labourer the sensibility is a maximum, the degree of sensibility in the monarch may be a minimum . . . for

popular opinion and the dictates of utility see Quinn 2014b.

by high doses of the exciting matter applied to the organ, its sensibility is in a manner worn out. (Bentham 1843, ix. 15)¹⁴

Further, in opposition to the standard model, Bentham once more recognizes the importance of context, and in particular of the reference points we use in evaluating our own situation. ‘Desires extend themselves with the means of gratification; the horizon is enlarged in proportion as we advance; and each new want ... becomes a new principle of action.’ (Bentham 1843, i. 304) Current endowments very quickly become taken for granted, they cease to deliver any anticipatory pleasures, and open up whole new fields for the pains of apprehension and of loss.¹⁵ The systematic distortion of recalled past pleasures and pains and anticipated future ones follows from the comparative vividness of present experience (Pellegrino ([2014])).

The combined impact of endowment effects and loss aversion explains the importance of relative comparisons to individual happiness. In his poor law writings, Bentham distinguishes between absolute and relative indigence. The relation in question is not between an individual’s holdings and those of others, but between a single individual’s holdings at different periods: ‘*Indigence* ... may be distinguished into absolute and relative: *absolute*, a deficiency in respect of the necessary means of preserving existence; *relative*, a deficiency of those means of comfort as well as subsistence to the enjoyment of which a man has been habituated.’ (Bentham 2010a, 255n) Bentham is further explicit that the crucial variable is the trajectory on which we find our wealth and income:

Absolute necessities being provided for, *comparative happiness as between man and man depends, even as far as wealth is concerned, not on the absolute quantum of wealth through the period of life, but on the ratio of the quantum*

¹⁴ For the development of a similar argument in nineteenth-century psychology (and psychological

of wealth at a more advanced period, compared with the quantum possessed at a more early period: that is, upon the mass of affluence being in a state of encrease, or at least not on the decline, in contradistinction to its being in a declining state. (Bentham 2010a, 311–12. My emphasis)

Bentham's insight has been confirmed by numerous studies on the relationship between wealth and happiness (Frey & Stutzer 2002, 409–16; Lowenstein & Sicherman 1991; Lowenstein & Prelec 1993; Frank & Hutchens 1993), while the preference for increasing income profiles 'appears to be driven in part by savouring and dread'—or, in Bentham's lexicon, the pleasure of expectation and the pain of apprehension—'and in part by adaptation and loss aversion' (Camerer & Lowenstein 2004, 26). Given the central importance of expectations which have become habitual, given the dependence of experienced pleasure or pain on changes or departures, either positive (pleasurable) or negative (painful), from the reference point of current endowment, Bentham's conclusion regarding the comparative happiness of the rich man who has just experienced significant loss, and the poor man who has just experienced a significant enhancement in income, should come as no surprise:

Habit tyrannizes the peasant not less than the prince. What a man has been used to have, he must continue to have, or he is unhappy. Be a man's income at a given period what it may ..., every accession it receives excites in his bosom the sentiment of affluence: every defalcation it receives excites in like manner the sentiment of privation and penury. The Prince who has £20,000 a year fixed income ... enjoys more happiness (be it admitted) than the Peasant who has £20 a year fixed income Be it so: but it is not the less unquestionable, that as far as depends on pecuniary circumstances, the Prince

economics) see Bruni & Sugden 2007, 150–4.

whose income has just undergone a reduction from £30,000 to £20,000 is less happy than the Peasant whose income has received an increase from £20 a year to £30. (Bentham 2010a, 312)

A second relativity also impacts on our current assessment of our happiness, and this depends upon social rather than temporal comparison: ‘Not happy enough to provoke our envy: not unhappy enough to warrant our sympathy: such is the condition in which, if it depended on each man, his fellows would be found. ... We can still worse bear that another should be too happy, than that he should not be enough so.’ (Bentham 2010a, 260) The impact of the rewards of others on the utility we derive from our own is completely overlooked by the standard model, but is recognized by behavioural economics, and by Bentham. It is true that he viewed envy as an irrational reaction to the good fortune of others (UC n.d. iii. 270–1 (1952–4, iii. 297–8)), but irrational or not, given its prevalence, it is a reaction of which it behoves the legislator to take account. Indeed, this is one reason why Bentham echoes Smith and presages Ricardo in endorsing the pursuit of growth. In a stationary or shrinking economy, with a stable or rising population, individual advancement becomes a zero-sum game:

For nations as for individuals, the happiest state is not that of having made one’s fortune, but that of making it: growing prosperity, that is happiness. Should the point be reached at which all employments are full, at which the earth has received all its developments, and industry has no more progress to make, what would then be the condition of human nature? A man could only make his fortune at the expense of another. ... Wages, the patrimony of the poor, that is of the great mass of the community, after having fallen gradually, will fix themselves at the point where they can furnish a man with bare

¹⁵ For the recognition of the insatiability of wants see for instance, Frey and Stutzer 2002, 414–15.

subsistence, with the most basic essentials. The state of work will be a state of war: all against all, fighting like gladiators at Rome—to the death. (Dumont n.d. 51, fo. 272 (My translation)¹⁶)

Partly because of these effects, Bentham would not have been surprised to learn that a doubling of national income has not been matched by a doubling of reported levels of happiness. Like the poor man's son of Smith's famous parable (1976 (1759), 181–6) we are all liable to be caught up on what Layard calls the hedonic treadmill (2011, 48–9). For Bentham, Easterlin's paradox is thus not remotely paradoxical, but depends in large part on psychological adaptation to increased prosperity.

2.3 Framing: Bentham the spin-doctor

At the top of Bentham's list of proposed reforms of government finance was a revival and extension of the law of escheat, by which vacant titles to property in land reverted to the state, whilst his enthusiasm for idea was derived precisely from its promise to offer 'Supply without Burthen'. The measure stipulated that on the death of property-holders not survived by children or other close relations, one half of their estate was to be 'appropriated to the use of the public' (Bentham 1795, 3 (1952–4, i. 283)), with the other half being disposed according to the will of the deceased. The dead have no expectations, while those of surviving children would be unaffected, since the transfer of property to them would continue unaltered. Meanwhile, the expectations of living but more distant collateral relatives could be modified without causing pain or alarm, provided that

¹⁶ MS orig: 'Pour les nations comme pour les individus, l'état le plus heureux n'est pas d'avoir fait sa fortune, mais de la faire: une prospérité croissante, voilà le bonheur. Quand on seroit au point que tous les emplois seroient pleins, que la terre auroit reçu tous ses développements, que l'industrie n'auroit plus de progrès à faire, quelle seroit alors la condition de la nature humaine? un homme ne pourroit faire sa fortune qu'aux dépens d'un autre. ... Les gages, ce patrimoine des pauvres, c.à.d de la grande masse de la Communauté, après être tombés graduellement, se fixeront au point où ils peuvent fournir à un homme sa simple subsistance, son plus étroit nécessaire. Un état de travail sera un état de guerre: tous contre tous, combattant comme les gladiateurs à Rome—jusqu'à la mort.'

sufficient notice were given of the change. The beauty of the proposal, for Bentham, lay in the manner in which it would avoid impacting upon existing expectations, but would rather prevent their formation.

For present purposes, the crucial passage occurs when Bentham compares his proposal to a tax on the inheritance of property. In so doing, he demonstrates an awareness of the influence of the frame of a proposal on the cognitive processing of the information which it contains:

Under a tax on successions, a man is led in the first place, to look upon the whole in a general view as his own: he is then called upon to give up a part. His share amounts to so much: this share he is to *have*; only out of it, he is to *pay* so much *per cent*. His imagination thus begins with embracing the *whole*: his expectation fastens upon the whole: then comes the law putting in for its *part*, and forcing him to quit his hold. This he cannot do without pain: if he could, no tax at all, not even a tax on property, would be a burthen:

(Bentham 1795, 32–3 (1952–4, i. 292))

A fifty per cent tax on successions, with exemptions for those inheriting from parents or near relations, would have exactly the same impact as Bentham's proposal on the post mortem disposal of the assets of the childless: the information contained in the two enactments would be substantively identical. However, Bentham the choice-architect recognizes that the public's evaluation of the two proposals will be very different, and that the difference arises from the role of law in giving rise to different expectations in the two cases.

Try the experiment upon a hungry child: give him a small cake, telling him after he has got it, or even before, that he is to give back part of it. Another time give him a whole cake, equal to what was left to him of the other and no

more, and let him enjoy it undiminished—will there be a doubt which cake afforded him the purest pleasure? (Bentham 1795, 33n (1952–4, i. 292n)¹⁷

Bentham was thus perfectly aware of, and happy to exploit, the importance of framing in eliciting preferences. In the classic Asian disease experiment, the only difference between the alternatives presented consists in the framing of their consequences in positive (lives saved) or negative (lives lost) terms (Tversky and Kahneman 1981). The conclusion that simply changing the frame can so decisively reverse the decisions made poses a major challenge to the notion of rational consistency. However, from a utilitarian perspective, since in this experiment the predicted consequences of the two alternatives are of equal value, there is no rational basis for preferring one to the other. What would alarm Bentham is the possibility that a better policy might easily lose out to a worse simply because of the frame in which they are presented.¹⁸ In general, he would agree with the behavioural economists that broad frames tend to lead to better, meaning more instrumentally rational, choices (Kahneman 2011, 371–4). However he would I think, cleave to the reality of pleasures and pains, and insist that observations about pain and pleasure were capable of bearing truth, and thus of offering the potential for frame-independent rational choice. Demonstration that the frame is decisive in identical examples is not equivalent to demonstrating that the ‘facts’ do not exist, and cannot make a decisive difference in non-identical examples.

2.4 Nudging and Defaults

¹⁷ Purity—the chance of being followed by further pleasures—was only one of seven dimensions of pleasure which together constituted its value (Bentham 1970, 38–41). Here, however, it seems that Bentham is asserting that the total value of the latter experience will be greater (i.e. will deliver a greater net balance of pleasure over pain) than that of the former, since it will not include any pain of deprivation.

¹⁸ The recent controversy in the UK about the ‘spare-room subsidy’, otherwise known as the ‘bedroom tax’, provides an example of how divergent descriptions of the same measure can give rise to different affective attitudes, and thereby opposite responses.

In terms of policy prescriptions, perhaps the most noteworthy contribution of behavioural economics has been the nudge theory of Thaler and Sunstein, who have introduced the notion of ‘choice architecture’ into contemporary debate: ‘A choice architect has responsibility for organizing the context in which people make decisions.’ (Thaler & Sunstein 2009, 3) In terms of Bentham’s logic of the will,¹⁹ the choice architect’s ‘nudges’ are distinguished from commands or prohibitions: ‘a nudge ... is any aspect of the choice architecture that alters people’s behaviour in a predictable way *without forbidding any options or significantly changing their economic incentives*’. (Thaler & Sunstein 2009, 6. My emphasis) Unsurprisingly, Bentham explicitly discusses neither choice architecture, nudges, nor default options. However, he did recognize the pervasive influence of the context of choice, and in particular the power, derived from our disinclination to think too hard, of the framers of a choice to increase the chances that a particular option will be chosen. In ‘Indirect Legislation’, with reference to the legislator’s exposition of accessory offences, he noted the danger of bringing ‘innocence into danger by hasty and inconclusive inferences’. His gloss presents an analysis of the effect of changing the default option from ‘opt-in’ to ‘opt-out’:

The description of an accessory offence can indeed scarcely be sufficiently secure without involving in it a clause leaving it to the discretion of the judge to determine concerning the conclusiveness of the inference. *In this case the creation of such an offence comes to nearly the same thing as to suggest the fact in question to the judge in the way of instruction*, under the character of an evidentiary circumstance. *There is only the difference that in the latter case*

¹⁹ Bentham’s logic of the will anticipates modern deontic logic, and analyses the varieties of imperation. For Bentham, the directive portion of any law describes an action, together with a prescription in relation to it which splits crudely into four: do it, don’t do it, you don’t have to do it, and you don’t have to refrain from doing it. (Bentham 2010b, 115–41)

the legislator puts it to the judge to make the inference, whereas in the former he makes the inference himself, reserving only a power in the judge to repel it if he should think proper Upon the latter plan the conclusion is most likely to take place: in the former it requires an effort to establish it: on the latter it requires an effort to overthrow it. That side of an alternative has caeteris paribus the best chance which has the vis inertia of the mind in its favour. Whichever scale the vis inertia of the mind is thrown into possesses a great advantage. (UC n.d. lxxxvii. 175. My emphasis)

Default options exploit precisely the *vis inertia of the mind*, that is, human aversion to labour. Thaler and Sunstein eschew direct appeals to well-being as the criterion of evaluation, but their rationale for nudging people towards certain options looks rather Benthamic: good nudges produce better choices and thus render ‘choosers better off, as judged by themselves’ (Thaler & Sunstein 2009, 5). Significant normative problems lurk behind this seductive formula, which will be addressed in § 3. For his part, it is difficult to imagine that Bentham would have hesitated to endorse their approach, including (at least in many cases) their caveat that the subject should retain the ability to resist the nudge without incurring significant costs.

For all his general endorsement of laissez-faire in political economy, Bentham made a significant exception in regard to the provision of knowledge. In discussion of indirect legislation he noted that ‘Sovereigns have hitherto been but little in use to stand forth in the character of instructors. It is a milder, however, and surely not less honourable function to enlighten one part of mankind by their wisdom, as to make another part feel the burthen of their power.’ (UC n.d. lxxxvii. 167) He recommended that government should directly address asymmetries of knowledge which allowed people to exploit the vulnerabilities of others. Standards of weights and measures,

officially approved indicators of quality on products (such as modern CE marks), registers of births, deaths and marriages, registers of property titles, and publication of current prices would go a long way toward preventing fraud. In addition, the legislator might consider ‘furnishing the people with cautionary instructions putting them on their guard against several modes of defraudment and other species of delinquency’ (UC n.d. lxxxvii. 153). Bentham considered whether government’s recommendations on the best ways to validate legal instruments, such as transfers of title, should or should not be mandatory, and noted:

Instructions given by government to the effect above suggested, even though each were left at liberty to adopt them or not as he thought, would not by any means be useless. *In matters of this sort people are ready enough in general to follow the patterns that are set before them by authorities much inferior to that of the legislature. Were such advice then to be given, the probability is that in general it would be taken:* especially if the compliance with it were to be enforced by some trifling penalty (UC n.d. lxxxvii.160. My emphasis)

Again, following a pattern set before me is what I do when I accept a default option.

As noted above, Bentham’s analysis of the subjective factors of production of voluntary human actions—namely knowledge, interest, and power—features prominently in his writings on political economy. Default options do address interests to a modest extent, and thus do address the will (by providing a very modest disincentive to rejecting the default: I have to tick the box on the form), but their central operation concerns knowledge, and addresses the understanding. Defaults contain two different kinds of information. The first kind is simply the message that the provider of the default option endorses the option, that is, they recommend acceptance: if they did not, they would not have made it the default. This message is

particularly powerful when it comes from an authority whom we trust, to whom indeed, we are looking to supply deficiencies in our own knowledge. The second kind of information contained in defaults is the message that ‘This is what most people do’, which appeals to our willingness to go along with the crowd. When the choice architect is a Benthamic legislator, it simply behoves her to set utility-maximizing defaults.

2.5 Saving for old age

Bentham would agree with Thaler and Sunstein that errors in human decision-making occur more frequently where such decisions are infrequent, and where feedback is slow (i.e. where the significant consequences of those decisions only impact on our experience of pleasure and pain at temporally remote periods). Unfortunately, many of the most important decisions we make—moving house, choosing a career path, saving for temporally distant goals—possess these features, which limits the extent to which standard economics and standard rationality are applicable to them (Bruni & Sugden 2007, 168–9). Thaler and Sunstein use what looks like a straightforwardly utilitarian calculation in advocating a strong nudge to the design of contributory pension plans: the possible bad consequences of saving too much are simply not as heinous as the bad consequences of not saving enough. Bentham himself was acutely conscious of the way in which ‘the wants of to-day eclipse those of to-morrow’ (Bentham 1843, i. 313), noting that: ‘In the youth, the idea of immediate good and evil occupies nearly all the sphere of reflection, excluding the ideas of distant good and evil.’ (1843, i. 315) In discussion of the utility of Friendly Societies (voluntary associations paying benefits in case of sickness or old age) he noted that many young men would decide that they had no need to make provision for such contingencies: which were ‘one of them distant in itself, and seen at an exaggerated distance (for

such are the optics of the mind), both of them such as nature does not love to look at' (Bentham 2001, 87). In relation to sickness, we are likely to underestimate the probability that we will be ill: 'to a man in the vigour of youth, who knows not as yet what sickness is, the danger of it is but a dream: he who hopes never to be sick, is to punish himself with present privation, in order to insure himself against this unexperienced, and consequently undervalued, misfortune'. (Bentham 2001, 87) As regards provision for old age:

If he lives long enough, indeed, he is sure enough of being old:—but is he destined to live long enough? *Will the powers of enjoyment remaining to him at that age be enough to pay him for the sacrifice he must make of the instruments of enjoyment, at an age when the power of employing them is in all its vigour? Everything of that sort lies wrapt in clouds.* (Bentham 2001, 87–8. My emphasis)

A degree of temporal-discounting was built in to Bentham's model of rationality. Of two otherwise equivalent pleasures, the first of which is to be experienced now and the second in the future, the rational agent will prefer the first. With regard to the present value of future pleasures or pains, that is pleasures and pains in prospect, that value is equal to the value which they would possess if they were indeed present, discounted according both to the degree of uncertainty that the sensation will occur, and to the distance in point of time which separates the individual from the experience of it. In regard to the temporal discount, Bentham relies on the rate of interest of money to supply the rate: if the interest rate is 5 per cent, a pleasure the enjoyment of which begins a year hence is worth 95 per cent of the same pleasure if that enjoyment were to begin now (Bentham 1998, 251–2). Even allowing for the temporal discount, however, the decisions made by young adults about whether, and how much, to set

aside from current income to facilitate the meeting of temporally distant needs looks irrational, in the Benthamic sense of failing to maximize the net balance of pleasures and pains.

What is Bentham's response to this failure of human rationality? One option would be simply to respect actual choices, and insist that autonomous individuals who choose not to save should be respected by allowing them to reap the consequences of their own actions, that is, potentially, to starve or freeze to death. For Bentham, this strategy is unlikely to succeed for two reasons. First, the interval between action and consequence is simply too long. Once I have arrived at an age when the question of securing provision for my declining years has acquired sufficient salience, it is probably too late to correct my mistake (assuming that I ever did earn sufficient surplus income to make the necessary savings). Second, if it is too late for me, the potentially salutary impact of the example of my suffering on the next generation, who are in danger of repeating my error, is likely to be minimal: 'How must this pretended lesson be weakened by the distance!—how small the analogy between an old and a young man!—how little does the example of the one operate upon the other! (Bentham 1843, i. 315)

Despite his general belief in the good consequences of allowing individuals the freedom to reap as they sow, Bentham clearly believed that abandoning the chances of survival when the capacity to labour was exhausted to the contingent chance of having had both the opportunity and the foresight to set aside a sufficient provision, was to expect too much of human rationality. In arguing that young people will make systematic mistakes in this area, he anticipates Thaler and Sunstein's plea for the architects of pension schemes and governments to recognize and to compensate for our systematic error (Thaler & Sunstein 2009, 118–19, 126–7): 'It is

sought to guide those who think little; and in order to draw instruction from such a misfortune, it is requisite that they should think much: of what use, then, I ask, is a political instrument destined for the least prudent class, if it is of a nature to be efficacious only upon the wise?’ (Bentham 1843, i. 315)

Bentham’s first thought was to promote the idea of saving simply by changing the frame of the decision. The image of senescence was both rebarbative in itself, and too alien from current experience to be appropriately salient, but neither negative applied to the prospect of marriage, which, by offering the reward of socially approved sexual activity, constituted an appeal to the positive rewards of the target state, rather than to the negative reward of avoiding temporally distant future pains. The combination of maximum capacity for labour with lack of responsibility for dependent others made young male adulthood the ideal time to foster habits of thrift, since at this time the necessary surplus of earnings was most likely to exist. Bentham’s proposed Poor Man’s Bank would provide the independent poor with the facility to save, and as the institution developed, the operation of the moral sanction, that is public opinion, would support it: ‘In process of time, under favour of the facility thus afforded to this species of thrift, a notion that may not unreasonably be expected to form itself ... is that of some certain sum, as a sum that a young couple ... *ought* to be in possession of, before they engage in marriage. It may come to be disreputable ... to set out with less’ (Bentham 2001, 80–1). The Poor Man’s Bank fosters the propensity to save by connecting it directly with a much desired state, and provides continuous prompting to the strengthening of the propensity, since every addition to the sacrifice of present pleasure shortens the period during which that state remains out of reach. Supported by an internalized moral sanction, or a desire to follow the herd, it punishes me for giving in to temptation to dissipate my surplus in

the purchase of the instantly attractive but permanently injurious gratifications of alcohol.²⁰

In another departure from the standard model, Bentham did not assume that adult human agents are unproblematically able to act on their preferences (Altman 2012, 35). Recognizing again that the context of choice was a crucial variable (and thereby anticipating the contemporary focus of capability theorists like Sen and Nussbaum (Sen 1985; Nussbaum & Sen 1993), Bentham directed his reader to the issue of power.²¹ Bentham was writing at a period when the inception of state provision of universal benefits for old age was long in the future. In passing, he did flirt with the notion of making the purchase of retirement-annuities compulsory for those whose earnings allowed the necessary deduction to be made without threatening current subsistence, whilst providing a pension sufficient for subsistence at public charge to those whose earnings were insufficient to make them ‘self-annuitable’ (Bentham 2001, 195–7). With reference to Thaler and Sunstein’s proposal to make

²⁰ The desire to provide a salient motive to thrift is also an important factor in Bentham’s defence of lotteries as expedients of finance. The obvious advantage of a lottery over a tax, as Bentham recognized (UC n.d. clxvi. 40), is its voluntary nature. However, lotteries also had an important collateral benefit, namely that of promoting frugality. To a modern ear, the idea of frugality-promoting lotteries sounds straightforwardly oxymoronic, but the contradiction is lessened, if by no means eliminated, when the price of entry to the lottery, that is the price of a ticket, rises to the point where a sustained period of self-denial is required to meet that price: *‘The sacrifice of the present to the future, of the pleasure of the moment to the security and satisfaction of the whole of life, is the great desideratum in morality The thoughtless are to be lured[?] to frugality by slow degrees. He who has prevailed upon himself to save his guinea for the purchase of a Lottery Ticket will learn in time to save it for the purchase of a distant Annuity for the decline of life.’* (UC n.d. clxvi. 40. My emphasis) The good thing about lotteries was that in order to put myself in a position to undertake an admittedly irrational gamble, I had to save up by resisting the daily temptations of, for instance, alcohol for an extended period. Such lotteries might work where advice would fall on deaf ears. Bentham admitted that the many losers in any lottery would suffer pain, but argued that it would very probably be outweighed by the value of their pleasures of hope in the extended period before the draw was made (Dumont n.d. 51, fo. 63).

²¹ See Bentham 2001, 86 ‘The prevalence of improvidence on the part of the Poor has not unfrequently been made the theme of lamentation, nor certainly without cause *But it is to little purpose to increase propensity where power itself is wanting: it is giving bitters instead of food. The first article to look to is power: that created or found existing, it is then and then only that to give existence or strength to inclination can be of use. This creation of power is oftentimes all that is wanting to be done on the part of government: for the nature of things may furnish inclination, and enough of it: but without power, all the inclination in the world is but uneasiness in waste. Oftentimes where inclination appears wanting, and perhaps really is wanting, it is only because opportunity is wanting—that power*

‘save more tomorrow’ the default for company pension schemes (2009, 113–27), it seems likely that he would have approved. Given the costs of coercion, in terms of pain, he might also endorse the liberty-preserving back-stop of good defaults: it remains open to the contributor to opt-out of the default. Both in reviewing the possibilities that government might supply itself with revenue (and thereby reduce the burden of coercive taxation) by selling financial services, and in discussion of existing Friendly Societies, he came close to reaching two conclusions found in his contemporary heirs, namely (1) simply providing information, without giving recommendations, is very unlikely to lead to good outcomes, and (2) more choices are not always best (Thaler & Sunstein 2009, 120, 151–63; Bentham UC n.d. clxvi. 169, 201)

§ 3 Choice, Liberty and Formal v. Substantive Rationality

3.1 Autonomy and the manipulation of choice

Libertarians have objected to ‘Nudge theory’ because of its alleged undermining of the liberty of the individual: for the state to seek to influence choice is for the state to disrespect the autonomy of the free individuals who are its citizens (Hausman & Welch 2010; Grüne-Yanoff 2012).²² Thaler and Sunstein argue convincingly that since every choice presupposes a context of choice, and since empirical studies demonstrate the pervasive influence of seemingly trivial or irrelevant aspects of that context, the idea of non-intervention, of allowing unmediated choice to do its job, is simply fantasy. Bentham might well agree, having argued consistently that the effort

and consciousness of power, without which inclination can not so much as bring itself into life:—and such is the case here.’ (My emphasis)

²² McQuillin and Sugden discuss a third possible source of normative value (in addition to happiness and freedom), namely well-being, which is assumed, in whatever it consists, to be the target of revealed

to erect political principles on the basis of a principle of liberty was a straightforwardly incoherent enterprise. Liberty—that is the absence of coercion—and government—that is coercion—are mutually exclusive concepts. Bentham rejected out of hand any political theory which began with a false premise of freedom, and developed instead the concept of security—precisely the product of coercive law. Security relates to the ability to project oneself and one’s interest into the future on the basis of expectations buttressed and defended by legal sanctions. No policy should be rejected on the basis of objections themselves grounded on the question-begging fallacy of its alleged incompatibility with the unproblematically autonomous choices of individuals, since the very notion of such choice was itself chimerical.²³ The incoherence of liberty as a foundational value directs attention precisely to the conditions under which persons come to conceive of themselves and their interests (Quinn 1997; 2011), or, in other words, to the context or ‘architecture’ of choice. Just such a focus is advocated by Hargreaves Heap, who recognizes the bankruptcy of the notion of an ‘antecedent set of individual preferences’ (2013, 989), and notes acutely: ‘people face a problem of how to decide what preferences to have, and this is something that policy can address through attention to the conditions under which people make such decisions without being committed to any view about what those preferences ought to be’ (996).

For Bentham, human beings simply *are* the type of creatures which will seek to control the choices of others, whenever there is the prospect of a pay-off in terms of pleasure or exemption from pain. We are all continually seeking to modify the behaviour of others in accordance with our will, while the assumption that the only

individual preferences (2012, 562–3). For Bentham, of course, no meaningful distinction can be drawn between well-being and happiness.

levers available to us in this task are those of force and fraud leaves the field wide open to those choice-architects capable—because of their possession of large resources, or expertise in cognitive psychology, or both—of exploiting the cognitive frailties of human beings. Government, as the guardian of the universal interest, owes its citizens not only the negative good of protection against the sinister interests of private choice-architects, but the positive assistance of default options and other nudges which help people with those frailties, that is all of us, make decisions which avoid consequences which we are all too lazy and or stupid to anticipate.

3.2 What does a good choice look like?

The challenge posed by behavioural economics to standard model of rationality, as utilised in the method of revealed preference, concerns the way in which revealed preferences often fail to maximize well-being. Failures of knowledge, of computational capacity and of will power issue in poor choices, which choosers later regret. As Thaler and Sunstein put it, the argument for seeking to nudge choosers in good directions is that ‘it will make choosers better off, *as judged by themselves*’. Hargreaves Heap poses the question ‘If choice is no longer a reliable guide to well-being, what is?’ (2013, 992) McQuillin and Sugden point out that Thaler and Sunstein’s approach requires a criterion by which to distinguish between rational and irrational choices, between true and false preferences: ‘The idea that choices are affected by mistakes ... implicitly assumes that people have reasonably coherent “true” preferences, even if these are not reliably revealed in choices.’ (McQuillin & Sugden 2012, 560) True preferences are those we would make if we had more perfect knowledge, indomitable will power, and first class computational abilities (Sunstein & Thaler 2003, 1162). In other words, they are the ones we would make if we

²³ Most basically, new human beings whose autonomy is respected, who are, that is to say, left alone to

measured up to the standard model of rationality: the rational choice is the one with the best chance of maximizing our utility. McQuillin and Sugden admit that there are a range of cases in which the existence of true preferences is relatively uncontroversial, as in the case of the bewildering complexity of tariffs for domestic utilities: ‘it seems unexceptionable to assume that consumers “truly” prefer to pay less rather than more for given goods and services, even though they sometimes fail to choose the lowest price tariff’ (2012, 561). However, there are many more cases where the criterion for the evaluation of the rationality of consumer choices is, to say the least, less obvious. Thaler and Sunstein have been criticized for seeking to manipulate the choices of individuals in the direction of a substantive conception of happiness with which those individuals may in fact disagree. Hargreaves Heap questions the legitimacy of nudge by rejecting the notion of objective happiness: ‘who is to say that leading a short, fat life should be discouraged, or that consuming more now and less in old age is a bad thing?’ (2013, 993) In essence, a substantive conception of the good is being imposed under the guise of a formal conception of rationality. In relation to ‘nudge’ theory, the charge has significant force: Thaler & Sunstein are notably reluctant to commit themselves to an explicitly utilitarian scheme of values. In relation to Bentham, however, there is simply no case to answer. As noted in § 1, his notion of rationality is substantive from top to bottom: rational choices are those which maximize the net balance of pleasure over pain.

Bentham, like Thaler and Sunstein, rejects the notion that people should simply be prohibited from making imprudent choices: since coercion is painful, the legislator should refrain from imposing a substantive conception of the good life on individuals who do not share it. Thus, what he termed ‘offences against the self’

develop and pursue their own interests, do not last long enough to self-define. Without interference, no

should not, generally speaking, be made penal (1970, 195–6, 289–91). ‘The cases in which it is not better for a man to suffer by his own will, than to be saved against his will, are neither many, nor very easy to determine.’ (Bentham 2010a, 606) In the sphere of indirect legislation, however, government’s role was potentially much broader. If advice and information fail to motivate individuals to more prudent calculations, no responsible government should neglect the opportunity to influence behaviour in a utilitarian direction. Bentham endorsed ‘sin taxes’, for instance on alcohol or ostentatious funeral services (UC n.d. clxvi. 39; 141 (1952–4, i. 395)), but opposed the imposition of taxes on ‘merit goods’, his favourite examples being medicines, books and newspapers (UC n.d. clxvi. 39; xvii. 317 (1952–4, iii. 368)). After all, as Thaler and Sunstein might point out, those who desire short fat lives will still be able to act on that preference: ‘Freedom to choose is the best safeguard against bad choice architecture’. (Thaler & Sunstein 2009, 12) For Bentham any frugal—meaning minimally pain-inducing—non-coercive interventions which can make it easier to make rational—meaning pleasure-securing or pain-avoiding—choices, can only be a good idea.

4. Conclusion

Bentham advanced a substantive, calculating and maximizing model of rationality. Although the substantive content distances his approach from that of contemporary economists, he shared several basic premises with them, and made early moves in the direction of the indirect assessment of utilities which would later be deployed in draining the substance from rationality, and divorcing economics from psychology. His own political economy was rooted in a psychology which recognized many of the

one would survive long enough to develop any complex interests (Bentham 2010a, 186–7n).

‘failures’ of rationality rediscovered by Behavioural Economists, and in the light of which it is plausible to regard him as one of their intellectual progenitors. The rational decision is that which maximizes expected happiness, and Bentham’s response to the cognitive biases revealed by behavioural economics would likely have been twofold. Insofar as both economists and legislators benefit from theories which make accurate predictions about human behaviour, he might well urge them to take the opportunity to incorporate the insights of behavioural economics into their respective theoretical and policy-making tool-boxes. With regard to economics, many behavioural economists make such a call, reflecting the perception that cleaving to an empirically flawed model of rationality, which appears incompatible with the empirical evidence about human psychology, is liable to produce significant errors in diverse parts of the field which economics claims as its own. (Bruni & Sugden 2007, 161–71; Camerer & Lowenstein 2004, 42–3; Binmore 1999). In relation to governments, there seems little doubt that Bentham would advise them to heed the lessons of behavioural economics by providing governmental nudges (except where the desired end could only be achieved through more coercive measures). Since Bentham’s enterprise encompassed not merely law and public policy, but deontology, or private ethics, there might well have been a second element to his response. Insofar as errors in prudential reasoning are widespread, he would see a role for behavioural economists as deontologists, in ‘pointing out to each man on each occasion what course of conduct promises to be in the highest degree conducive to his happiness’ (Bentham 1983, 123), as has been envisaged by Lowenstein and Ubel in their recommendation that the target audience for the discipline should be individuals rather than governments (2008), and, in varying degrees, by Thaler and Sunstein (2009) and Kahneman (2011). Finally, he would have viewed concerns over manipulation of choice and assaults on autonomy

as founded on a fallacious presumption of individual liberty, while his explicitly substantive notion of rationality would render him immune from charges that a fully developed and contestable standard of value was being disingenuously incorporated into an ostensibly formal evaluation of the rationality of choice.²⁴

²⁴ It is entirely arguable that the disjunction between experienced and remembered utility revealed in the seminal experiments by Kahneman, Wakker and Sarin (1997) presents Bentham's theory with a significant difficulty which Bentham himself never addressed, in so far as utility, the objective carrier of value, appears to speak with not one, but a plurality of voices. To which should the utilitarian cleave: decision utility, experienced utility or remembered utility? The choice matters because the different concepts will yield different prescriptions: should medical interventions aim to minimize total experienced pain or total remembered pain? However, proper investigation of the issue would require a different paper.

References

- Altman, Morris. 2012. *Behavioural Economics for Dummies*, Mississauga: John Willey.
- Baumgardt, David. 1952. *Bentham and the Ethics of Today*, Princeton: Princeton University Press.
- Becker, Gary. 1996. *Accounting for Tastes*, Cambridge MA: Harvard University Press.
- Bentham, Jeremy. No date. University College London collection of Bentham manuscripts (UC). (Roman numerals refer to the boxes in which the manuscripts are placed, Arabic to the folios within each box)
- Bentham, Jeremy. 1787. *Defence of Usury; shewing the Impolicy of the Present Legal Restraints on the Terms of Pecuniary Bargains*, London: T. Payne.
- Bentham, Jeremy. 1795. *Supply without Burthen; or Escheat vice Taxation: Being a proposal for a saving of taxes by an extension of the law of escheat, including strictures on the taxes on collateral succession comprised in the budget of 7th December 1795*, London: J. Debrett.
- Bentham, Jeremy. 1825. *The Rationale of Reward*, ed. R. Smith, London: John and H.L. Hunt.
- Bentham, Jeremy. 1827. *Rationale of Judicial Evidence, Specially Applied to English Practice*, 5 vols., ed. J.S. Mill, London: Hunt & Clarke.
- Bentham, Jeremy. 1843. *The Works of Jeremy Bentham*, published under the superintendence of his executor, John Bowring, 11 vols., Edinburgh: Tait.
- Bentham, Jeremy. 1952–4. *Jeremy Bentham's Economic Writings*, ed. W. Stark, 3 vols., London: Allen & Unwin.
- Bentham, Jeremy. 1970. *An Introduction to the Principles of Morals and Legislation*, ed. J.H. Burns and H.L.A. Hart, London: Athlone.

- Bentham, Jeremy. 1977. *A Comment on the Commentaries and a Fragment on Government*, ed. J.H. Burns and H.L.A. Hart, London: Athlone.
- Bentham, Jeremy. 1983. *Deontology, A Table of the Springs of Action, and Article on Utilitarianism*, ed. A. Goldworth, Oxford: Clarendon Press.
- Bentham, Jeremy. 1998. *'Legislator of the World': Writings on Codification, Law and Education*, ed. P. Schofield and J. Harris, Oxford: Clarendon Press.
- Bentham, Jeremy. 2001. *Writings on the Poor Laws: Vol. I*, ed. M. Quinn, Oxford: Clarendon Press.
- Bentham, Jeremy. 2010a. *Writings on the Poor Laws: Vol. II*, ed. M. Quinn, Oxford: Clarendon Press.
- Bentham, Jeremy. 2010b. *Of the Limits of the Penal branch of Jurisprudence*, ed. P. Schofield, Oxford: Clarendon Press.
- Bentham, Jeremy. 2011. *Jeremy Bentham: Selected Writings*, ed. S. Engelmann, New Haven & London: Yale University Press.
- Bentham, Jeremy. 2015. *The Book of Fallacies*, ed. P. Schofield, Oxford: Clarendon Press.
- Bhagwati, Jagdish. 1994. 'Free Trade: Old and New Challenges', *Economic Journal* 104 (No. 423), 231–46.
- Binmore, K. 1999. 'Why experiment in Economics', *Economic Journal* 109 (No.453), F16–24.
- Bruni, Luigino and Sugden, Robert. 2007. 'The road not taken: how psychology was removed from economics, and how it might be brought back', *Economic Journal* 117 (1), 146–73.

- Camerer, Colin F. and Lowenstein, George. 2004. 'Behavioural Economics: Past, Present and Future', in *Advances in Behavioural Economics*, ed. C.F. Camerer, G. Lowenstein and M. Rabin, Princeton: Princeton University Press, 3–51.
- Dumont, Pierre Louis Etienne. No date. Bibliothèque Genève, Manuscrits Etienne Dumont.
- Easterlin, Richard A. 1995. 'Will Raising the Incomes of All Increase the Happiness of All?', *Journal of Economic Behaviour* 27 (3), 35–47
- Easterlin, Richard A. 2001. 'Income and Happiness: Towards a Unified Theory', *Economic Journal* 111 (2), 465–84
- Edgeworth, Francis Y. 1967 (1881). *Mathematical Psychics*, New York: Kelley.
- Frank, Robert H. and Hutchens, Robert M. 1993. 'Wages, Seniority, and the Demand for Rising Consumption Profiles', *Journal of Economic Behavior and Organizations* 21 (3), 251–76.
- Frey, Bruno S. and Stutzer, Alois. 2002. 'What Can Economists Learn from Happiness Research', *Journal of Economic Literature* 40 (2), 402–35.
- Goldworth, Amnon. 1979. 'Jeremy Bentham: on the measurement of subjective states', *Bentham Newsletter* No. 2, 2–18.
- Grüne-Yanoff, Till. 'Old wine in new casks: libertarian paternalism still violates liberal principles', 2012. *Social Choice and Welfare* 38 (4), 635–45.
- Guidi, Marco E.L. 2000 'Bentham's Economics of Legislation', Dipartimento di Studi Sociali, Università degli Studi di Brescia, *DSS Papers* STO 2–00
<http://fausto.eco.unibs.it/~segdss/paper/guidi00>
- Guidi, Marco E.L. 2007. 'Jeremy Bentham's quantitative analysis of happiness and its asymmetries', in *Handbook of the Economics of Happiness*, eds. L. Bruni and P.L. Porta, Cheltenham: Edward Elgar, 68-93.

- Haakonssen, Knud. 1981. *The Science of a Legislator: The Natural Jurisprudence of David Hume and Adam Smith*, Cambridge: Cambridge University Press.
- Hargreaves Heap, Shaun P. 2013. 'What is the meaning of behavioural economics?', *Cambridge Journal of Economics* 37 (5), 985–1000
- Hartley, David. 1775 (1749). *Hartley's theory of the human mind, on the principle of the associations of ideas; with essays relating to the subject of it*, ed. J. Priestley, London: J. Johnson.
- Hausman, Daniel M. & Welch, Brynn. 2010. 'To Nudge or Not to Nudge', *Journal of Political Philosophy* 21 (1), 123–36.
- Jevons, William Stanley. 1970 (1871). *The Theory of Political Economy*, Harmondsworth: Penguin
- Kahneman, Daniel. 2011. *Thinking, Fast and Slow*, London: Penguin.
- Kahneman, Daniel and Shane, Frederick. 1990. 'Experimental Tests of the Endowment Effect and the Coase Theorem', *Journal of Political Economy* 98 (6), 1325–48.
- Kahneman, Daniel, Wakker, Peter P., and Sarin, Rakesh. 1997. 'Back to Bentham: Explorations of Experienced Utility', *Quarterly Journal of Economics* 112, 375–405.
- Kelly, Paul J. 1990. *Utilitarianism and Distributive Justice: Jeremy Bentham and the Civil Law*, Oxford: Clarendon Press.
- Layard, Richard. 2011. *Happiness: Lessons from a New Science*, London: Penguin.
- Lowenstein, George and Prelec, Dražen. 1993. 'Preferences for sequences of outcomes', *Psychological Review* 100 (1), 91–108.
- Lowenstein, George and Sicherman, Nachum. 1991. 'Do Workers Prefer Increasing Wage Profiles?', *Journal of Labor Economics* 9(1), 67–84.

- Lowenstein, George and Ubel, Peter A. 2008 ‘Hedonic adaptation and the role of decision and experience utility in public policy’, *Journal of Public Economics* 92 (8), 1795–1810.
- McQuillin, Ben and Sugden, Robert. 2012. ‘Reconciling normative and behavioural economics: the problems to be solved’, *Social Choice and Welfare* 38 (4), 553–67.
- Morewedge, Carey K. In press. ‘Utility: Anticipated, Experienced, and Remembered’, in *Blackwell Handbook of Judgment and Decision Making*, ed. G. Keren & G. Wu, 2nd edn, Oxford: Blackwell.
- Nussbaum, Martha and Sen, Amartya (eds.). 1993. *The Quality of Life*, Oxford: Oxford University Press.
- Pellegrino, Gianfranco. [2014]. ‘A Benthamite Solution to the Easterlin’s Paradox(es)’, unpublished ISUS conference paper, ISUS XIII, Yokohama National University.
- Quinn, Michael. 1997. ‘The Fallacy of Non-interference: the Poor Panopticon and Equality of Opportunity’, *Journal of Bentham Studies* 1 (<http://ojs.lib.ucl.ac.uk/index.php/jbs/article/view/4>)
- Quinn, Michael. 2011. ‘Post-modern moments in the application of empirical principles: Power, Knowledge and Discourse in Jeremy Bentham vs. Michel Foucault’, *Revue d’études Benthamiennes* 8 (<http://etudes-benthamiennes.revues.org/245>)
- Quinn, Michael. 2014a. ‘Bentham on Mensuration: Calculation and Moral Reasoning’, *Utilitas* 26 (1), 61–104.
- Quinn, Michael. 2014b. ‘Popular Prejudices, Real Pains: What is the Legislator to do when the people err in assigning mischief?’, in *Bentham’s Theory of Law and Public Opinion*, ed. X. Zhai and M. Quinn, Cambridge: Cambridge University Press, 63–89.

- Rabin, Matthew. 1998. 'Psychology and Economics', *Journal of Economic Literature* 36 (1), 11–46.
- Read, Daniel. 2007. 'Experienced utility: Utility theory from Jeremy Bentham to Daniel Kahneman', *Thinking & Reasoning* 13 (1), 45–61.
- Schumpeter, Joseph A. 1954. *History of Economic Analysis*, New York: Oxford University Press.
- Scitovsky, Tibor. 1941. 'A note on Welfare Propositions in Economics', *Review of Economic Studies* 9 (1), 77–88.
- Sen, Amartya. 1985. *Commodities and Capabilities*, Oxford: Oxford University Press.
- Sent, Esther-Mirjam. 2004. 'Behavioural Economics: How Psychology Made Its (Limited) Way Back Into Economics', *History of Political Economy* 36 (4), 735–60.
- Smith, Adam. 1976 (1759). *The Theory of Moral Sentiments*, ed. D.D. Raphael & A.L. Macfie, Oxford: Oxford University Press.
- Sunstein, Cass R. & Thaler, Richard H. 2003. 'Libertarian Paternalism is not an oxymoron', *University of Chicago Law Review* 70 (4), 1159–1202.
- Thaler, Richard H. & Sunstein, Cass R. 2009. *Nudge: Improving decisions about Health, Wealth and Happiness*, London: Penguin.
- Tversky, Amos and Kahneman, Daniel. 1981. 'The framing of decisions and the psychology of choice', *Science (New Series)* 211, No. 4481; 453–8.
- Warke, Tom. 2000a. 'Mathematical fitness in the evolution of the utility concept from Bentham to Jevons to Marshall', *Journal of the History of Economic Thought* 22 (1), 5–27.
- Warke, Tom. 2000b. 'Multi-Dimensional Utility and the Index Number Problem: Jeremy Bentham, J.S. Mill and Qualitative Hedonism', *Utilitas* 12 (2), 176–203.

Winch, Donald. 1983. 'Science and the Legislator: Adam Smith and After', *Economic Journal* 93 (3), 501–520.

Winch, Donald. 1992. 'Adam Smith: Scottish Moral Philosopher as Political Economist', *Historical Journal* 35 (1), 91–113.

Winch, Donald. 1996. *Riches and Poverty: An Intellectual History of Political Economy in Britain, 1750–1830*, Cambridge: Cambridge University Press.