Economics Letters 215 (2022) 110488

Contents lists available at ScienceDirect

Economics Letters

journal homepage: www.elsevier.com/locate/ecolet

Behavioural utilitarianism and distributive justice

Giorgos Galanis [a,*], Roberto Veneziani [b]

[a] Goldsmiths, University of London and Centre for Research in Economic Theory and its Applications, University of Warwick, United Kingdom
[b] School of Economics and Finance, Queen Mary, University of London, United Kingdom

ARTICLE INFO

Article history:
Received 21 February 2022
Received in revised form 27 March 2022
Accepted 29 March 2022
Available online 4 April 2022

JEL classification:
D63
D9

Keywords:
Utilitarianism
Inequality
Reference dependent preferences

ABSTRACT

What are the distributive implications of utilitarianism? Is it compatible with a concern for equality, as many utilitarians have argued? We analyse these questions in the context of a pure allocation problem. We consider an infinitely-lived economy and, drawing on the behavioural literature, assume that individuals have reference-dependent preferences: agents' utility is a function of current consumption and a reference point which captures consumption habits, or the agents' upbringing. Assuming a history of inequalities in consumption, we show that the utilitarian allocation is *equalising*: starting from an unequal distribution, inequalities decrease over time at the utilitarian optimum. However, even though agents are in a relevant sense identical, equality does not obtain at any finite time.

1. Introduction

Classical utilitarianism is undoubtedly one of the most prominent and widely adopted approaches in normative economics and in policy analyses. Yet, as critics have long pointed out, it has potentially undesirable distributive implications. On the one hand, the utilitarian planner is definitionally indifferent between alternative allocations given a certain level of aggregate utility. On the other hand, the very maximisation of total utility may require an extremely unequal allocation of both resources and utility, if agents have different preferences. Utilitarians have rejected these criticisms, or at least significantly deflated their relevance and have traditionally argued that utilitarianism is compatible with equality.

In Bentham, for example, the egalitarian implications of utilitarianism derive from the 'axioms of moral and political pathology', namely those empirical generalisations that are "expressive of the connexion between such occurrences as are continually taking place, or able to take place, and the pleasures and pains which are respectively the results of them" (Bentham, 1838–1843b, p. 224). Based on these axioms, which include claims on the marginal utility of money and wealth, Bentham (1838–1843a, p. 313) concludes that "We may observe, that in a nation which prospers by agriculture, manufactures, and commerce, there is a continual progress towards equality…This will be the result of different habits formed by opulence and poverty".

In this paper, drawing from the behavioural literature, we provide a novel perspective on this debate. In order to capture Bentham's notion of evolving "habits formed by opulence and poverty", and be consistent with a large behavioural literature on reference dependent preferences, we assume that in every period, each agent's utility depends both on current consumption and on their (or their parents', as a proxy of their upbringing) past consumption habits. Then, assuming the planner to inherit a society with a history of past inequalities, we ask, what will the distribution be at the utilitarian allocation? How will it evolve over time? Under what conditions, if any, is Bentham's conjecture correct?

We show that, starting from any initial distribution, if the common utility function is concave, then at the utilitarian optimum, inequalities decrease over time and disappear in the limit. While it is tempting to interpret this result as confirming Bentham's conjecture – as changes in habits lead to a convergence in consumption, – we shall argue that this is not the only, or necessarily the most persuasive interpretation of our result. For, noting that the reference point is endogenous, and that the agents' utility functions are, in a relevant sense, identical, our main result may be read as showing that at the utilitarian allocation, equality in the distribution of resources does not obtain at any finite time, *even with identical agents*. We argue that this raises some new interesting questions for utilitarians.

This paper is related to various strands of literature, in addition to debates in normative economics and social choice. The influence of habits can be understood as a special case of reference dependent preferences in line with Prospect Theory (Kahneman and Tversky, 1979), where the reference point evolves based on

* Corresponding author.
E-mail addresses: g.galanis@gold.ac.uk (G. Galanis), r.veneziani@qmul.ac.uk (R. Veneziani).

https://doi.org/10.1016/j.econlet.2022.110488

previous consumption levels. Hence, this paper provides a link between the literature on distributive justice and Prospect Theory and models of habit formation. As habits may also be interpreted in a more general sense to include one's history, upbringing, and family background, which determine one's norms and expectations. At a broad conceptual level, these intergenerational effects are reminiscent of the emphasis on family circumstances that is central in the modern theory of equality of opportunity (Roemer, 1998; Roemer and Veneziani, 2004; Fleurbaey and Maniquet, 2011). In this paper, however, we shall not assume an egalitarian planner to begin with but rather enquire on the conditions (if any) that may lead to equality at the utilitarian solution, whatever the agents' initial circumstances.

A similar problem has been analysed, in an older contribution, by Layard (1980). He considered the utilitarian distribution of a given amount of income in a static model in which agents care about both actual and *expected* income. Layard (1980) proved that if expectations depend on past incomes, then the utilitarian distribution is not egalitarian: "yesterday's rich should have higher net incomes than if all had the same expectations" (Layard, 1980, p. 746). However, he conjectured that if one allowed for expectations to adjust over time towards the level of income actually experienced and "there were no time discounting", then the distribution of a constant amount of income "should eventually become equal" (Layard, 1980, p. 746).

Finally, the paper also speaks to the recent literature on behavioural welfare economics (Bernheim and Rangel, 2007, among others). However, unlike in these contributions, we consider a rather mild deviation from the standard model with rational agents – namely, habits – and we explicitly consider normative issues in a dynamic context.

2. Results

Consider an infinitely-lived economy with $N \geq 2$ households, which can be interpreted either as N infinitely-lived individuals or as an infinite number of individuals, each living for one period and belonging to N family lines. As we are interested in the distributive properties of utilitarianism in a dynamic economy with reference dependent preferences, we will abstract from production and growth, and consider a pure allocation problem with a fixed population size.

In each period, a divisible consumption good must be shared among the N agents. Following Koszegi and Rabin (2006), we assume that individuals care about both absolute and relative consumption. To be specific, at time $t = 0, 1, 2, \ldots$, individual utilities $u^i = u^i\left(c_t^i, r_t^i\right)$, $i \in \{1, \ldots, N\}$, depend on a weighted average of current consumption, c_t^i, and consumption relative to the reference point r_t^i. We assume that agents have the same per-period utility function, $u^i = u : \mathbb{R}_+^2 \to \mathbb{R}$, for all $i \in \{1, \ldots, N\}$ but their reference points depend on past history and may differ, so that individuals may have heterogeneous preferences over current consumption. We assume that u is twice differentiable, strictly increasing in current consumption, and strictly concave.

Formally, the utility of individual $i \in \{1, \ldots, N\}$ at time t, is given by[1]

$$u[(1-\alpha)c_t^i + \alpha(c_t^i - r_t^i)] = u(c_t^i - \alpha r_t^i) \tag{1}$$

with $\alpha \in (0, 1)$. Following the habit formation literature (e.g. Carroll (2000)), the reference point is a weighted average of the history of previous consumption:

$$r_{t+1}^i = \lambda c_t^i + (1-\lambda) r_t^i, \tag{2}$$

with $\lambda \in (0, 1]$, given a vector of initial reference points $\mathbf{r_0} = \left(r_0^1, r_0^2, \ldots, r_0^N\right)$ and

$$\sum_{i=1}^{N} r_0^i = 1,$$

and $r_0^i \geq 0$, for all i. At any t, let $\mathbf{c_t} = \left(c_t^1, c_t^2, \ldots, c_t^N\right)$. The utilitarian planner solves

$$\max_{\{\mathbf{c_t}\}_{t=0}^{\infty}} \sum_{t=0}^{\infty} \beta^t \left[\sum_{i=1}^{N} u(c_t^i - \alpha r_t^i)\right], \tag{MP}$$

subject to (2), and

$$\sum_{i=1}^{N} c_t^i = 1, \ \forall\, t, \tag{3}$$

given r_0^i for all i, with $\beta \in (0, 1)$.

Proposition 1. *At the solution of MP, for $t \to \infty$, $c_t^i \to \frac{1}{N}$, for all $i \in \{1, \ldots, N\}$.*

Proof. 1. The solution of the MP is also the solution of the following Bellman equation:

$$V(\mathbf{r_t}) = \max_{\mathbf{c_t}} \left[\sum_{i=1}^{N} u(c_t^i - \alpha r_t^i) + \beta V(\mathbf{r_{t+1}})\right], \tag{4}$$

subject to

$$r_{t+1}^i = \lambda c_t^i + (1-\lambda) r_t^i,$$

and

$$\sum_{i=1}^{N} c_t^i = 1, \ \forall\, t.$$

The FOCs are

$$u'(c_t^i - \alpha r_t^i) - u'(c_t^N - \alpha r_t^N) = \beta\lambda[V_N(\mathbf{r_{t+1}}) - V_i(\mathbf{r_{t+1}})] \tag{5}$$

for $i = 1, \ldots, N-1$, where $V_i(\mathbf{r_t})$ is the partial derivative of the value function at time t, with respect to the reference point of individual i. The envelope conditions are

$$V_i(\mathbf{r_t}) = -\alpha u'(c_t^i - \alpha r_t^i) + (1-\lambda)\beta V_i(\mathbf{r_{t+1}}), \tag{6}$$

for $i = 1, \ldots, N$. Then, using the envelope condition of agent N, we get

$$\begin{aligned} V_N(\mathbf{r_t}) - V_i(\mathbf{r_t}) = \alpha &\left[(u'(c_t^i - \alpha r_t^i)) - u'(c_t^N - \alpha r_t^N)\right] \\ &+ \beta(1-\lambda)\left[V_N(\mathbf{r_{t+1}}) - V_i(\mathbf{r_{t+1}})\right]. \end{aligned} \tag{7}$$

Then, by substituting the RHS of (5) into (7), we get

$$V_N(\mathbf{r_t}) - V_i(\mathbf{r_t}) = \beta[\alpha\lambda + 1 - \lambda]\left[V_N(\mathbf{r_{t+1}}) - V_i(\mathbf{r_{t+1}})\right]. \tag{8}$$

Shifting (5) one period back, we get

$$u'(c_{t-1}^i - \alpha r_{t-1}^i) - u'(c_{t-1}^N - \alpha r_{t-1}^N) = \beta\lambda\left[V_N(\mathbf{r_t}) - V_i(\mathbf{r_t})\right], \tag{9}$$

which given (5) and (8), gives the following Euler equation

$$\begin{aligned} &u'(c_{t-1}^i - \alpha r_{t-1}^i) - u'(c_{t-1}^N - \alpha r_{t-1}^N) \\ &\quad = \beta[\alpha\lambda + 1 - \lambda]\left[u'(c_t^i - \alpha r_t^i) - u'(c_t^N - \alpha r_t^N)\right]. \end{aligned} \tag{10}$$

Note that $\alpha < 1$, so $\alpha\lambda < \lambda$ which also means that $0 < \beta[\alpha\lambda + 1 - \lambda] < 1$. Note that if

$$c_t^i - \alpha r_t^i = c_t^N - \alpha r_t^N, \tag{11}$$

for all t, then (10), is true.

[1] We follow much of the literature on habit formation and assume the reference point to enter additively the agents' utility function. Similar results can be derived if the habit takes a multiplicative form.

2. The summation of (11) over i, gives

$$N(c_t^N - \alpha r_t^N) = \sum_{i=1}^{N}(c_t^i - \alpha r_t^i) = \sum_{i=1}^{N} c_t^i - \alpha \sum_{i=1}^{N} r_t^i.$$

But from the budget constraint $\sum_{i=1}^{N} c_t^i = 1$, for all t, which also means that given (2) also $\sum_{i=1}^{N} r_t^i = 1$, for all t, hence

$$c_t^N = \frac{1-\alpha}{N} + \alpha r_t^N. \tag{12}$$

From (11) and (12), it follows that $V(\mathbf{r_t}) = N\frac{u\left(\frac{1-\alpha}{N}\right)}{(1-\beta)}$ solves Bellman's equation and since $\lim_{t\to\infty} \beta^t V(\mathbf{r_t}) = 0$ for all feasible sequences $\{\mathbf{r_t}\}_{t=0}^{\infty}$, we conclude that it solves MP.

3. If we express (11) in terms of the reference point given (2), we get

$$r_{t+1}^i - r_{t+1}^N = (1 - \lambda + \alpha\lambda)(r_t^i - r_t^N). \tag{13}$$

Note that $1 - \lambda + \alpha\lambda < 1$, hence for $t \to \infty$, $r_t^i = r_t^N$ for all i. Hence, for $t \to \infty$, $r_t^i = \frac{1}{N}$. Then also given (12), for $t \to \infty$, $c_t^N = \frac{1}{N}$. Also, given (11), for $t \to \infty$, $c_t^i = \frac{1}{N}$, for all i. □

Proposition 1 shows that consumption inequality between any two households is decreasing over time and disappears in the limit.

The mechanism underlying this result is quite intuitive. From Eq. (11), it follows that at the utilitarian solution, in each period consumption inequality between households is lower than the difference between the respective reference points — the ratio of consumption inequality to the difference between reference points is equal to $\alpha < 1$. Then, as consumption levels of one period feed into the reference points of the next, the households' consumption habits become more similar, which spurs the equalisation process further.

Proposition 1 thus seems to vindicate utilitarianism against egalitarian critics, and to confirm Bentham's conjecture. Even if the planner inherits a history of inequalities, leading to different habits, and thus to heterogeneous utility functions *over consumption*, at the utilitarian optimum inequalities in consumption (and thus habits) decrease over time and disappear in the limit. In addition to its theoretical relevance, this result would also lend empirical support to utilitarianism given the strong evidence of the pervasiveness of reference dependent preferences.

Yet, while this simple reading of Proposition 1 is legitimate, it is by no means the only plausible interpretation, and the implications of our analysis for utilitarianism are more nuanced than they may appear at first sight. A preliminary point to note is that while inequalities in consumption *per period* vanish in the long run, a history of past inequalities still implies potentially large inequalities if one considers N infinitely lived agents, or N households, over their *whole lives*.

Perhaps more interestingly, two features of the model should be highlighted which raise doubts on the simple interpretation of Proposition 1. First, although the history of past inequalities until $t = 0$ is given for the utilitarian planner, the dynamics of the agents' reference points at all $t \geq 1$ is actually endogenous. Second, at any time t, for a given set of (unequal) reference points, the agents' utility functions over consumption are indeed different. However, the basic structure of agents' utility function, as a function of *both* current consumption *and* consumption habits is exactly the same and, in this sense, one may argue that agents are structurally identical.

Indeed, given the endogeneity of the reference point at all $t \geq 1$, one may argue that agents living in each period after the first one are actually identical from the viewpoint of the utilitarian planner. To see this, consider the extreme case with $\lambda = 1$, in which the habit stock coincides with consumption in the previous period (or by the previous generation). In this case, the planner could obliterate both consumption inequalities *and* any effect of consumption habits on current preferences in a single stroke. Hence, in all periods after the first agents do have exactly the same utility function from the perspective of the utilitarian planner at $t = 0$. Yet, by Proposition 1, equality does not obtain, except at the limit, in the very long-run. Indeed, the structure of the utilitarian problem forces the social planner to ignore the dynamic effects of consumption allocation choices. As (11) shows, the solution of the intertemporal utilitarian problem coincides with the myopic utilitarian optimum, where the planner maximises the sum of the utilities in each period.

By way of illustration, Fig. 1 shows the dynamics of consumption over time at the utilitarian optimum in a society with two individuals (or dynasties) with initial reference points equal to 0.9 and 0.1, respectively, and $\alpha = 0.5$ and $\lambda = 0.2$.

But then, perhaps counterintuitively, Proposition 1 may be interpreted as contradicting the utilitarians' counterarguments against egalitarian critics. For it may be argued that, even though agents are, in a relevant sense, identical, the utilitarian optimum does not entail equality of consumption, at any finite time. Utilitarianism thus seems inconsistent with a concern for equality even in what has long been considered the most favourable scenario.

3. Conclusions

In this paper, we have examined the distributive implications of utilitarianism in a dynamic context, when individual preferences are reference-dependent. Assuming that all agents have the same utility function – which depends on current consumption and past consumption habits – but the utilitarian planner inherits a history of past inequalities, we have shown that at the utilitarian optimum, inequality decreases over time but it does not disappear except at the limit.

This result raises two interesting issues. First, if agents' preferences are interpreted as being fundamentally different, due to different reference points, then our result confirms – indeed strengthens – the relation between utilitarian thought and egalitarian principles: albeit not immediately egalitarian, the utilitarian allocation is equalising even in the presence of heterogeneity arising from past inequalities. If, however, one notes that, except for the very first period, consumption habits are endogenous and agents have utility functions with the same structure, then our result casts doubts on the relation between utilitarianism and egalitarianism highlighted by many economists and philosophers: even if agents have the same, concave utility functions, at the utilitarian allocation inequalities persist for an indefinitely long time.

Thus, the implications of our result for debates on utilitarianism and distributive justice hinge, at least partly, on a conceptual issue, namely what it means for agents to have the same utility functions in a dynamic economy with inherited inequalities and endogenous preferences.

Second, is it always ethically sound for the utilitarian planner to take agents' preferences as given? If current preferences – in our model, the reference level – emerge from a history of past injustices, then one may argue that the utilitarian planner should either disregard or discount actual preferences.[2] In the former case, the planner might evaluate the optimal allocation using ideal, or laundered utility functions, or even utility functions that actively correct for such injustices. However, this would be inconsistent with classical hedonistic utilitarian approaches – according to which only the agents' actual subjective preferences

[2] We are grateful to an anonymous referee for raising this point.

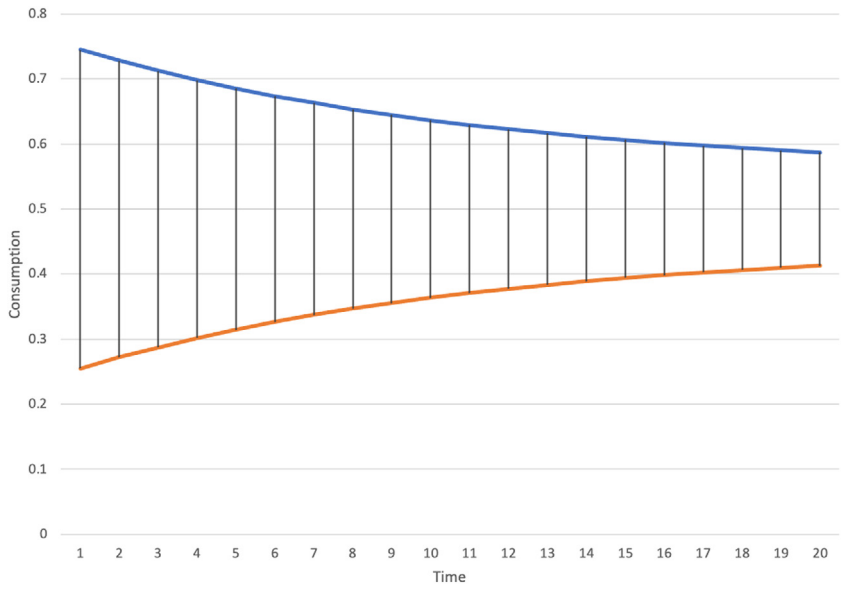

Fig. 1.

(whatever they are) matter – and it would raise some complex issues concerning the choice of such counterfactual preferences. In the latter case, one may depart from classical utilitarianism and assign higher weights to the utilities of individuals with lower past consumption, which raises the problem of the fair choice of weights. We leave these issues for further research.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We thank an anonymous referee for detailed and insightful comments. We are also grateful to Stephen Engelmann, Marc Fleurbaey, Peter Hammond, Karsten Kohler, Richard Layard, Peter H. Matthews, Herakles Polemarchakis, Jonathan Riley, John Roemer, Zvi Safra, Uzi Segal, Kostas Zachariadis, and audiences in Notre Dame, Warwick, Cambridge and London for useful comments on an early draft of the paper. The usual disclaimer applies.

Funding

No funding supported this work.

References

Bentham, J., 1838–1843a. "Pannomial fragments". In: The Works of Jeremy Bentham: Published under the Superintendence of His Executor, John Bowring. Vol. 3. Tait, Edinburgh.
Bentham, J., 1838–1843b. "Principles of the civil code". The Works of Jeremy Bentham: Published Under the Superintendence of His Executor, John Bowring. Vol. 1. Tait, Edinburgh.
Bernheim, B.D., Rangel, A., 2007. Toward choice-theoretic foundations for behavioral welfare economics. Amer. Econ. Rev. 97 (2), 464–470.
Carroll, C.D., 2000. Solving consumption models with multiplicative habits. Econom. Lett. 68 (1), 67–77.
Fleurbaey, M., Maniquet, F., 2011. A Theory of Fairness and Social Welfare. Cambridge University Press.
Kahneman, D., Tversky, A., 1979. Prospect theory: An analysis of decision under risk. Econometrica 47 (2), 263–291.
Koszegi, B., Rabin, M., 2006. A model of reference-dependent preferences. Q. J. Econ. 121 (4), 1133–1165.
Layard, R., 1980. Human satisfactions and public policy. Econ. J. 90, 737–750.
Roemer, J.E., 1998. Equality of Opportunity. Harvard University Press.
Roemer, J.E., Veneziani, R., 2004. What we owe our children, they their children,…. J. Public Econ. Theory 6 (5), 637–654.