The transformations of utility theory: a behavioral perspective

Ulrich Witt¹,²

Abstract The aim of this paper is threefold. First, it reappraises the major transformations which the utilitarian approach to human behavior has undergone in economics in search for a representation by utility functions and later by preference orders. Second, in the light of today’s behavioral and human sciences, an attempt is made to restore some elements of early utilitarianism that were abandoned in these transformations. Third, in line with the interest of the early utilitarians in both explaining behavior and elaborating on its moral assessment, the present paper also discusses some normative implications of the suggested restoration of utilitarian theory.

Keywords Utility · Preferences · Utilitarianism · Motivation · Needs · Reinforcement learning · Goal striving

JEL Classification A12 · B12 · B13 · B21 · D01 · I31

1 Introduction

In the early days of utilitarianism, utility was given a very concrete, sensory interpretation. This was the basis for explaining the reasons of human action and for deriving by implication what a ‘good life’ means. The result was a positive theory and a related normative discourse. Today this is different. Over the two centuries since its existence, the utilitarian approach has undergone a dramatic metamorphosis. If students of eco-

¹ Max Planck Institute for Science of Human History, Jena, Germany
² Griffith Business School, Griffith University, Gold Coast, Australia
nomics nowadays are introduced to the concept of ‘utility’, the sensory underpinnings and the connection with moral philosophy are almost entirely missing. Starting point is now an axiomatic theory of preferences from which a purely formal notion of ‘utility’ is logically deduced (see, e.g., Mas-Colell et al. 1995). It is the notion of an abstract magnitude, an index number, of which nobody seems to know what kind of psychic experience it represents. And utilitarian moral philosophy has adjusted to the language of subjective “preference orderings” in place of sensory perceptions and their hedonic quality (see, e.g. Sen 1987).

Outside economics, by contrast, a very different development took place. In the behavioral and human sciences, many material hypotheses about human behavior and its underlying motivations have been launched over the last decades—several of them in a remarkably similar spirit as the utilitarian conjectures that have been abandoned in economics. The evolutionary paradigm can be argued to offer an overarching framework able to organize the vast body of knowledge that has been accumulated about the motivations of human behavior (Burnham et al. 2016). In this paper, some of these developments will be reviewed with the intention to show how a behavioral theory of utility can be formulated on this basis, a theory that rehabilitates in part the explanatory and normative impetus of early utilitarianism.

The paper proceeds as follows. Section 2 summarizes basic views held in early utilitarianism and outlines the major transformations by which the approach was successively changed in economics into an abstract theory of choice behavior. Section 3 discusses a few attempts in economics that revived some elements of the early utilitarian approach or even its very sensory foundations, albeit with little impact on presently taught utility theory. Section 4 then reviews elements of a theory of action motivation developed in biology, the behavioral science, and psychology. They are relevant for a behavioral re-interpretation of utility theory, a re-interpretation that will turn out to restore the original utilitarian program in important respects. Section 5 addresses the normative question. This question resurfaces since the suggested re-interpretation of utility theory seems to impose limitations on what a ‘good life’ may, or may not, mean. Section 6 offers some tentative conclusions.

2 The early utilitarian program and its successive transformation

In 1789 Bentham published his “Principles of Morals and Legislation” (Bentham 1948). With its blending of a positive theory of action and a normative theory of justice it became the authoritative statement of the early utilitarian program. In his book the synonym for “utility”—a term which Bentham had taken over from Hume—is “happiness”. This points to the hedonistic underpinnings by which Bentham wanted to make the concept ‘operational’: “what happiness consists of we have already seen: enjoyment of pleasures, security from pains” (ibid., p. 70). Consequently, Bentham focused on the magnitude of pleasures and pains generated by someone’s action or by a “sanction” imposed on someone by other agents or by nature. Human action is intentional and motivated by the action’s consequences in terms of pleasures and
the avoidance of pains (ibid., Chaps. VII and X). With remarkable psychological intuition Bentham argued that the pleasures and pains—and, hence, the motivation to take action—vary, in turn, with their intensity, duration, certainty, and nearness in time. He expanded on a detailed enumeration of fourteen different sorts of pleasures and twelve different sorts of pains which are, in turn, subdivided further (ibid., Chap. V). Moreover, he argued that each of them can be sensed differently depending on yet another, long list of diverse circumstances which he extensively commented on (ibid., Chap. VI).

From today’s point of view, the different sorts of categories given by Bentham refer to quite a mixed bag of sensory perceptions, innate as well as learned responses, and complex contingencies. However, the basic ideas are quite modern.

(i) Utility is interpreted as a hedonic experience generated by the pleasurable or painful sensory perceptions resulting from an action like, e.g., the consumption of a commodity. Utility is thus an attribute of actions not of objects.

(ii) The utility that an action, the consumption of a good, say, generates can be caused by several different pleasures (or pains avoided) simultaneously.

(iii) As a sensory perception, utility is considered an observable and measurable magnitude.

With respect to the last point and the concrete measurement of utility, Bentham believed that pleasures and pains can be quantified, added up, and balanced—pleasures with positive values, pains with negative values (ibid. Chap. IV). With such an understanding it may appear only a small step to inter-personal comparisons of the magnitudes of pleasures and pains. Indeed, Bentham assumed that an outside spectator observing someone who enjoys pleasures or suffers pains can attribute money equivalents to the strength of the observed person’s sensory experience. Once quantified in ‘objective’ pecuniary terms, the utility of different agents can be added up and/or balanced. By striking a balance Bentham wanted to determine the moral justification of institutions like penal law or constitutions or individual actions. He considered an institution or action morally right (wrong), if the balance is positive (negative). Where the balance has its greatest positive value, the morally best solution is found—Bentham’s “greatest happiness principle”. Such inter-personal comparisons of utility involve (hidden) value judgements from which morally rather odd implications can follow. 2

As far as the positive part of Bentham’s sensory utilitarianism is concerned, his assumption (iii)—that utility, more precisely the magnitude of pleasures and pains,
can indeed be measured and observed—is one thing. (The early utilitarians took this so much for granted that they did not even think of possible procedures of observation or measurement.) That outside spectators can attribute money values to someone else’s sensory experience—thus creating a basis for inter-personal comparisons—is another thing. Since pleasures and pains are sensory perceptions, the former idea does not seem implausible. It has later been suggested that the Weber-Fechner law and the empirical measuring of sensory perceptions on which the law is based could provide both the foundation and the empirical method for utility measurement (see Stigler 1950, part II).

The metamorphosis of Bentham’s formulation of utility theory started with Jevons’ “Theory of Political Economy” of 1871 (Jevons 1879). Jevons took over Bentham’s pleasure-and-pain rhetoric, but his intentions were different ones. He aimed at a mathematical formulation of utility theory, a “calculus of pleasure and pain”, inspired by the calculus used in classical mechanics (see Mirowski 1988, Chap. 1). In classical mechanics—the then prevailing ideal of scientific analysis—laws are expressed in single functional relationships. They can be subjected to differential calculus and the corresponding extreme value analysis. Jevons’ project of constructing a corresponding “mechanics of utility and self-interest” (ibid., p. 23) aimed at a theory that “...consists in applying the differential calculus to the familiar notions of wealth, utility, value, demand, supply, capital, interest labor, and all the other quantitative notions belonging to the daily operations of industry. As the complete theory of almost every other science involves the use of that calculus, so we cannot have a true theory of Economics without its aid.” (ibid., pp. 4–5).

As Warke (2000) has argued, with its elaborate, naturalistic catalogue of pleasures, pains, sensibilities, and circumstances, Bentham’s interpretation was too complex for such a project. Jevons therefore introduced the following simplifications.

(i*) Unlike Bentham, he attributed “utility” not to actions (e.g., the act of consuming a certain quantity of a commodity) but to the commodities themselves and their quantities (Jevons 1879, pp. 40-2, 46-7).

(ii*) Where for Bentham an action (consuming a certain quantity of a commodity, say) can cause the simultaneous enjoyment of several pleasures, Jevons shunned the discussion of the plurality of causes by arguing that utility is a not further differentiated, compound “feeling”. He represents it in his calculus by a one-dimensional variable “u” (ibid., pp. 8-15).

(iii*) Jevons abandoned Bentham’s conceptualization of utility as an objectively measurable magnitude. He postulates instead that the compound “feeling” belongs to the subjective sphere of the decision maker which he considers “inscrutable”. (ibid., pp. 15-7).

Of these modifications, the subjectivist creed (iii*) is the most widely recognized one. Jevons claims that “every mind is .. inscrutable to every other mind, and no common denominator of feeling seems possible” (ibid., p. 15). However, the other, more subtle, modifications have significant consequences for the revision of utilitarianism, too. When Bentham’s divers causes and contingencies of utility are replaced by an undifferentiated, unobservable “feeling” (modification (ii*)), it is difficult, if not impossible, to explain how utility as a motivation for taking action relates to motiva-
tions like wants, needs, or desires. These constructs play a role in the behavioral and human sciences. Jevons’ modification creates obstacles for a dialogue with these sciences. His move also hampers the normative discourse within utilitarianism where the interpretation of utility in terms of desires and their fulfilment still figures prominently (see, e.g., Hare 1982). 3

Modification (i*)—attributing utility to quantities of commodities rather than the action of consuming them—is conducive to the use of marginal calculus. Its consequence is, first, that by replacing actions by quantities of commodities the time dimension can be left out of the calculus. While actions like, e.g., consumption develop their consequences over time, quantities of commodities do not necessarily have an intrinsic time dimension – which makes a static marginal calculus feasible. Second, when utility is attributed to actions, (expected) utility hinges on the agents’ knowledge of means-ends relationships. By attributing utility directly to commodities these contingencies of instrumental knowledge and the uncertainty they involve need no longer be made explicit in the utility assessment. 4 From the point of view of a normative theory, the major effect of replacing actions by commodity quantities is that the latter, unlike the former, have no moral connotations. Jevons’ substitution thus eliminates the common basis of the positive utilitarian theory of action and the normative utilitarian theory of the actions’ consequences. Subtle as the modification may appear, it may have contributed to an increasing alienation of the positive branch of modern utilitarianism from its origins in moral philosophy.

The development after Jevons focused on working out and refining the mathematical details of utility calculus. In the beginning twentieth century, writers like Edgeworth, Fisher, and Pareto wrestled with the problem of finding a representation of utility theory (or of what calculus had left over of it) in terms of a proper utility function. From the 1930s onwards, the transition from cardinal to ordinal notions of utility prepared the ground for further “progression in mathematical thought” (Samuelson 1947, p. 92). The problem was that the few empirical conjectures that the revised utility theory entails (decreasing marginal rates of substitution; the law of indifference implying a downward sloping demand curve) all hinge on the shape of the utility function. Yet, it was not clear what form of functions would be appropriate for expressing these conjectures in a consistent way, not to speak of whether well-behaved functional forms would conform to the empirical record of choice behavior.

One approach to solving that problem, suggested by Hicks and Allen (1934), was to deduce implications of the theory that should be observable at the level of market

---

3 The fact that utility derived from an action can have multiple causes does not exclude its representation by a one-dimensional variable. Recent research in the neurosciences has shown that in the assessment of choice alternatives, different sources of reward can be associated with the alternatives. The brain is able to spontaneously aggregate the different dimensions of reward in terms of one common ‘neural currency’ into a one-dimensional utility variable (see Glimcher 2016). A different question is, of course, how many different, multi-dimensional stimuli the brain is able to process at a time, see Witt and Binder (2013) for a discussion.

4 By conditioning utility by “understanding” and “consciousness”, Bentham (1948, Chap. VII) had been eager to emphasize the knowledge-based, cognitive and social contingencies in enjoying pleasures or avoiding pains. As will turn out below, these contingencies are important for understanding how preferences are formed through learning processes at both the sensory and the cognitive level.
demand curves, i.e. to derive, or rationalize, the law of demand. Utility theory is in this way transformed into a building block of price theory. A different approach, namely a logical reduction of utility theory to a theory of subjective preference orderings, was suggested by Samuelson (1938). In the positivist spirit of the time, the theory of revealed preferences, as he called it, was supposed to rely exclusively on operational terms—prices, quantities, and income—and a postulate of consistent choice behavior (the “weak axiom of revealed preference”). The theory should be logically equivalent to ordinal utility theory. It should therefore allow to deduce well-behaved demand functions without recourse to the utility concept. Moreover, by the preference revelation method that it implied, the theory should be able to empirically reconstruct the individual indifference curves on which ordinal utility theory relied.

Attractive as Samuelson’s theory appeared, it turned out to be neither as logically conclusive nor as powerful an empirical device as claimed. The consistency of choice behavior—whether in the weak or the strong versions of the axiom of revealed preference—is an empirical conjecture. However, the empirical validity of the hypothesis can only be proved by means of exactly the same observations that Samuelson’s procedure requires for reconstructing the individual preferences. Hence, whether or not the empirically revealed preferences are correctly derived hinges on whether or not one believes in the consistency of the individuals’ choice behavior. For this belief no independent evidence can be provided—a logical indeterminacy (cf. Georgescu-Roegen 1954a, Wong 1978, Chap. 4, 5). It made Sen (1973) wonder whether revealed preference theory represents more than an “elaborate pun”.

While the theory of revealed preference did not gain much momentum as an empirical program, its impact on microeconomic theory can hardly be underrated (see, e.g., Mas-Colell et al. 1995). It fostered a rigorous, abstract approach based only on the axioms of reflexivity, completeness, and transitivity. (The desired shape of logically implied utility functions and unique solutions in comparative statics analysis of budget changes also require the more technical assumptions or axioms of continuity, convexity, and non-satiation of preferences.) Given that all the axioms are highly idealized assumptions, the relevance of this theory for empirical work is controversial (Kreuzenkamp and Barten 1995). The theoretical rigor comes at a cost: utility theory has lost significantly in terms of its empirical content.

3 Partial reversals and the connections to other motivation theories

Over the past decades, attempts have been made to broaden the theory again or even to revoke some of the transformations. They gained a certain attention but did not

---

5 As later shown by Houthakker (1950), revealed preference theory is indeed observationally equivalent to ordinal utility theory, if the consumers’ preferences are transitive (i.e. if the “strong axiom of revealed preference” holds).

6 Originally envisaged tests of the theory by market data were not conducted. Some attempts were made to examine it in experiments, yet with rather inconclusive results, see, e.g., Koo (1963), MacCrimmon and Toda (1969), Koo and Hasenkamp (1972). Besides the consistency problem, the static framework of revealed preference theory ignores that individual preferences can change during, or even as a consequence of, the revelation procedure, a criticism already raised by Robinson (1962, p. 50).
develop a major impact on textbook utility theory. These attempts were in part driven by the elaboration of econometric aggregate demand theory, a huge literature of its own omitted here for space reasons. In part, however, the utility theory itself was the object of change. An example for attempts to broaden the narrow theory is Lancaster’s (1966; 1971) “characteristics approach”. Lancaster assumes that consumers derive utility from the “characteristics” or intrinsic quality features of the goods rather than from the goods as such. He introduces an ordinal utility function that has standard properties except that, as arguments of the function, the quantities of goods are replaced by a vector representing the quantities of characteristics which the goods possess. For example, the characteristics of a wrist watch are its color, its form and size, the number and specification of mechanical complications, the quality and weight of the various materials of which it is manufactured, and so on. Under the simplifying assumption that only one good is involved in each consumption activity (like the wrist watch in measuring the time), the consumption technology can be described as a linear transformation of the commodity space into the characteristics space. The consumption technology is assumed to be ‘objectively’ known to all consumers.

Lancaster’s approach has the merit of drawing attention to the largely neglected, but economically significant, quality dimension of goods and services (Wadman 2000, Chap. 6). Except for this extension, however, his approach is firmly grounded in assumptions (i*)–(iii*) that Jevons introduced. In contrast, an attempt to reverse a modification introduced by Jevons—namely assumption (i*) by which utility is attributed to quantities of commodities rather than actions—is made in Becker’s theory of the allocation of time and his household production theory (Becker 1965, see also Michael and Becker 1973). Becker argues that households obtain utility from “productive activities” in which purchased market goods and services are one sort of input and the household’s time the other. The households’ activities are assumed to generate utility via the produced “household commodities”. The latter enter a subjective utility function with standard features. Households maximize their utility function subject to given prices and marginal productivities of the inputs, an income constraint, and a time constraint. The consequence of replacing market commodities by household commodities via an analysis of household activities is that the time dimension is back as a crucial aspect of how utility is generated. Accordingly, alternative uses of time – i.e. decisions between alternative household activities which cannot be conducted at the same time—are the key feature of Becker’s theory and its applications.

---

7 See the literature on linear expenditure systems (Stone 1954) and distributed lags in consumption time series (interpreted as “habit formation”, see Houthakker and Taylor 1966). The statistical approach developed in these contributions was only loosely informed by utility theory. The “almost ideal demand system” proposed by Deaton and Muellbauer (1980) tried to change this. The authors derived constraints from utility theory for the demand systems that were subjected to empirical tests with aggregate data. Due to too many degrees of freedom, a direct test of the underlying theory turned out not be feasible in this way, ibid., Chap. 3.

8 As Steedman (2001, Chap. 2) has shown, the importance of time as an ultimate constraint and the marginal conditions for maximizing utility over alternative uses of scarce time were clearly formulated by Gossen as early as 1854. Steedman (ibid., Chap. 5) also makes clear that economizing on time cannot be framed other than as a problem of choice between actions differing in their time intensity.
In their already mentioned work under the programmatic title “Back to Bentham”, Kahneman et al. (1997) go a step further. In the light of modern psychological research they explain how the sensory perception of a hedonic experience arises from actions (Bentham’s assumption (i)), and how it can be observed and measured (assumption (iii)). They identify an immediate experience of outcomes (“instant utility”). Its values correspond to Bentham’s variable “intensity”. Bentham’s variable “duration” turns out to need a more differentiated representation. Kahneman et al. argue that the duration of hedonic experience is assessed in retrospect. Only what is remembered can therefore count for the assessment of the stream of immediate experiences varying in their intensity (“remembered utility”). However, they provide evidence that what is remembered does not seem to reflect the whole stream. Remembered utility rather seems to be represented by the average of the intensity of pleasures or pains at the peak of that stream and near the end of it—which means that remembered utility is not affected by the duration of the sensory experience. \(^9\)

By reversing modification (i*) and (iii*), Kahneman et al. return to Bentham’s interpretation of utility as a measurable, and perhaps inter-personally comparable, variable (see Binder 2010 for a discussion). Yet, they still treat utility as a rather undifferentiated “feeling” (Jevons assumption (ii*)). If one were to revoke this assumption too, the plurality of pleasures caused by an action would be back on the screen—and, accordingly, the motivating force which the expectation of these diverse pleasures has for choosing the action. It can then be discussed, more specifically, how non-utilitarian motivational theories of needs, wants, or desires relate to utility theory, and how they may contribute to explaining what causes pleasurable sensory experiences. Such an exercise can draw on theories of needs or wants which have a long, though today neglected tradition in economics.

The concept of needs and drives—in the economic literature often epitomized by the term “wants”—goes back at least to Plato. From the outset, it has been associated with the idea of a hierarchical order or the “principle of the subordination of wants” as Georgescu-Roegen (1954b) has called it. \(^10\) The principle can be interpreted to imply that, if a higher want always appears after a lower one has been satiated, an individual’s total demand or consumption will never reach satiation. Until the 1960s, a theory of wants or needs was frequently a part of explaining consumption behavior (e.g., in Duesenberry 1949, Chap. 1; Abbott 1955, Chap. 4; and Ironmonger 1972). In some cases it was explicitly combined with the utilitarian approach. An example

---

\(^9\) Kahneman et al. also reflect on normative implications, albeit in a way not comparable to the moral ambitions of the early utilitarians (see also Kahneman 1994). When the assessment of a stream of sensory experiences is left to the intuitive working of our memory, the result is “remembered utility”. Since this does not fully represent the temporal profile of instant utility, one could claim that people better base their assessment on the “total utility”, determined by the integral of the value of instant utility over time. However, this would be a normative claim which people do not intuitively follow. Depending on how this normative issue is decided, another (normative) question arises. Should people be forced—or at least be ‘nudged’—to make use of total utility where they tend to rely on remembered utility?

\(^10\) During the so-called marginalist revolution in economics this non-utilitarian theory was favored by Menger (1950, first edition 1871, Chap. 1). He argued that there is a demand for goods, because people have needs and have learned that they can be satisfied by these goods. Menger (ibid., p. 131) assumed a hierarchical structure as did Jevons—who called the assumption the “law of human wants” — although it was not consistent with his own approach.
The transformations of utility theory: a behavioral perspective

is Georgescu-Roegen’s (1954b) comparison between axiomatic preference theory on the one hand and the older literature on wants on the other—a remarkable antidote to the positivist attitude of his time.

Georgescu-Roegen avoids a precise definition of wants. But the way he uses the concept and, in particular, attributes indifference curves to wants makes it plain that an action that serves the satisfaction of a specific want is considered an action that causes a specific kind of utility. Since the consumers have a large, probably infinite, number of wants, there are as many potential sources of pleasures. With his “principle of the irreducibility of wants” Georgescu-Roegen claims that the multiplicity of wants (or sources of utility) cannot be lumped into just one catch-all “want”. However, precisely this has been done, he argues, “in a veiled passage” by “the founders of marginal utility theory” with their concept of utility representing “the unique want into which all wants can be merged” (ibid. p. 515). Obviously, Jevons’ modification (ii*) is attacked here, and the vehicle Georgescu-Roegen suggests to overcome it is his theory of wants.

Georgescu-Roegen’s approach was later taken up and varied by Ironmonger (1972). Interested in consumer innovations he recognizes that “without some distinction between various types of wants, there is no place for considering a change in the quality of a commodity or the introduction of a new commodity to the market” (ibid., 13). Like Georgescu-Roegen, Ironmonger acknowledges a multiplicity of wants which “...are assumed to be so ordered that at a given income and prices the consumer will satiate as many wants as possible, going down the order of priority from the most important to the least” (ibid., 23). But unlike Georgescu-Roegen he assumes that the number of units of satisfaction of all these different wants can be merged to give a homogeneous utility measure (see footnote 3 above). The measure induces a preference order over wants rather than commodities. More precisely, Ironmonger argues that because of the hierarchical order and the satiability of wants, the preference order is lexicographic. He goes on, on this basis, to determine optimal budgets by means of linear programming and to analyze, in the usual fashion, the comparative statics of choice and the effects of quality differentiation and new commodities.

Both Georgescu-Roegen and Ironmonger thus demonstrate how a theory of wants—even though they construct it ad hoc—can inspire material conjectures about the sources of utility. The theory of motivation has, of course, been developed more systematically and much further in the behavioral and human sciences. A reappraisal of the early utilitarian intuitions in terms of the present state of motivation theory can therefore draw on a set of richer hypotheses in which needs, wants, and (cognitive) desires or goals do play a role. The subject of the next section therefore is how these hypotheses can be made use of for better understanding what causes utility and under what conditions. In this way, utility theory can be integrated into the broader frame of a naturalistic theory of human behavior—something called for in recent pleas for the consilience of the sciences (see Wilson 1998, pp. 204–5).

4 A behavioral reappraisal of the utility and choice behavior

A straightforward starting point for discussing the reasons of action is to recognize the influence the human genetic endowment. Not only the physical traits of a species...
are innate, i.e. heritable, but also an elementary (instinctive) behavior repertoire and the basic learning mechanisms by which actual behavior adapts during a lifetime. Following the working hypothesis of evolutionary biology, particularly sociobiology, and evolutionary psychology, these inherited parts evolved most likely in ancestral times when humans were under fierce selection pressure. Behavior that enabled an advantage in accessing scarce resources like food, territory, status, mating partners then resulted in differential reproductive success and was selected for. In the case of humans, their emerging cognitive capacity is likely to have started influencing the process of selection. This has been conjectured to have given rise to a “dual inheritance” (Richerson and Boyd 2005), a cultural and a genetic one.

Regarding action motivation, the relevant feature of human behavior that seems to be inherited from ancestral times is the role played by need deprivation. This is an elementary motivator of action and its reduction also the basis of reinforcement learning by which the actions actually chosen are adapted to the environmental conditions. Regarding the physiological needs such as those for water, sleep, food, adequate body temperature, (a certain level of) physical activity, sex, etc. deprivation occurs as a consequence of some physiological tension, deficiency, or imbalance. In the case of the need for food, for instance, it is signaled by a feeling of hunger when the organism’s metabolism lacks energy in the form of nutrition. A motivation to act (e.g., to engage in foraging behavior) arises. The deprived state, and thus the motivation to act, vanishes when the action(s) lead to satiation of the corresponding needs. However, when an action is able to reduce deprivation with respect to several needs simultaneously, a motivation to take that action exists as long as satiation is not yet reached with respect to all relevant needs.

Deprivation also occurs and motivates action in the case of innate needs shared only by higher animals. Among these needs are the ones for arousal of the senses (or the cognitive system), for social status recognition, and for care and affection. Moreover, humans also share with the usual genetic variance specific cognitive needs such as the need for autonomy (Hagger et al. 2006) and the need for a positive self-image or self-esteem (Pyszczynski et al. 2004). In the utilitarian diction, need deprivation can in all these cases be interpreted as a sensory perceptions classified by innate neural processes as a form of pain. Its reduction is experienced as a relief from pain, a pleasant experience.

While need deprivation causes the motivation to act, reinforcement learning and conditioned reinforcement determined which actions come to be chosen as a result of the deprivation. By reinforcement learning (operant conditioning) the relative frequencies with which actions belonging to the behavioral repertoire are chosen are adapted as follows. Suppose that initially, actions are selected at random. If they are experienced to reduce need deprivation—a rewarding sensory perception—their relative frequency increases as long as deprivation is present or has reemerged. Conversely, an action followed by a sensory perception of pain is not only discontinued, but also motivates active avoidance behavior which, when successful in relieving pain, is reinforced (Herrnstein 1990). 11

11 Apart from relief of pain, a reinforcing effect has been observed for actions reducing deprivation with respect to innate needs such as the just mentioned ones. As innate or “primary” reinforcers they are
Conditioned reinforcement (conditioning learning) is a different, but related process (Leslie 1996, pp. 40-2). Imagine two actions. One of them triggers the rewarding sensory perception of reducing deprivation of an innate need while the other action triggers a neutral experience. Assume that over time both actions repeatedly coincide so that an association is learned between the two. Once such an association is established, the originally neutral action can result in a more or less rewarding instance in its own right. A conditioned or “secondary” reinforcer is established that works even if the previously coinciding primary reinforcement is dropped. The strength of a conditioned reinforcer fades, however, if the association on which it is based is not at least occasionally corroborated.

To distinguish it from innate needs let the motivational equivalent of secondary reinforcers be dubbed acquired wants. Given the powerful associative capacity of the human brain, an elaborate structure of acquired wants can emerge over an individual’s life time. Unlike the widely shared innate needs, this structure is idiosyncratic and varies greatly between individuals. Nonetheless, the influence of an individual’s cultural environment on what associations she happens to learn causes cultural contingencies in the acquired wants. They tend to be more similar within similarly socialized groups than between groups. With regard to the utilitarian program these considerations suggest:

**Hypothesis 1** Utility arises from the rewarding sensory perception associated with actions that reduce deprivation of innate needs or that satisfy acquired wants.

This hypothesis is in accord with core assumptions of the original sensory utilitarianism. In hedonistic terms, the rewarding sensory perception is a pleasurable experience (“instant utility” in Kahneman, Wakker, and Sarin’s terminology). It is connected to a variety of innate needs and acquired wants. As a sensory perception it is measurable in principle.

Turning to the explanation of the actual choice behavior, the motivation to act is often induced by different needs and/or wants simultaneously which can be deprived to different degrees. The actions available usually differ with respect to the extent to which they allow the rewarding experience of satisfying needs/wants. Hence there are multiple trade-offs. The question arises which of the actions are more effective in providing rewarding experiences. Such a situation requires the organism to spontaneously generate internal valuations of the available alternatives (Shizgal 1999, p. 509; see Witt and Binder 2013 for a discussion). They depend on the different degrees of deprivation, on earlier experience with the actions’ relative capacity to reduce deprivation, and on the current constraints. In the longer run, learning thus prompts an adaptation to both prevailing resource constraints and the organism’s ‘comparative advantages’ in generating rewarding sensory perceptions by choosing particular actions rather than other actions. The baseline behavior by which the double adaptation problem is solved is governed by reinforcement learning. Hence:

---

Footnote 11 continued
commonly shared among humans, see Leslie (1996, Chap. 2). Note that need satisfaction is not the only instance in which primary reinforcers occurs in a species, see Lea (1983).
Hypothesis 2  For behavior not controlled by cognitive deliberation, the relative frequencies with which alternative actions are chosen depend on two kinds of variables. The first is the relative strength of deprivation of the various innate needs or of the motivation triggered by the various acquired wants during the relevant period of time. The second variable is the relative reward in terms of need or want satisfaction resulting on average from the particular actions chosen in the past.12

The hypothesis denotes the main determinants of utility (abstracting for the moment from cognitive influences). These are the relative degree of deprivation the decision maker is exposed to with respect to her several innate needs; the learned association from which the acquired wants (conditioned reinforcers) derive their motivational strength; the average efficacy of the decision maker in triggering rewarding sensory perception by reducing need deprivation or satisfying acquired wants with the different available actions, given her constraints in time and resources. In relation to the last determinant it should be noted already that there are significant differences between the different innate needs in how the degree of their deprivation can be reduced when the same amount of resources is spent on their satisfaction (a point to be taken up in the next section).

In humans, the choice of actions is, of course, to a considerable extent involving cognitive activity, particularly in the context of actions requiring larger outlays. When cognition intervenes into making choices this can lead to modifications of the baseline behavior discussed so far. For that reason, the explanation of both the motivation to act and the way in which actions come to be chosen when a motivation to act exists has to account for cognitive influences (see Bargh et al. 2010 for the following). Regarding the motivation to act, cognitive activity can result in goal setting and a corresponding goal striving as a motivational force of its own. The strength of this motivation to act depends (among other things) on the decision maker’s strength of belief that an accomplishment of the goal is feasible and on her persistence (a personality factor). The accomplishment of the goal triggers a rewarding experience. Regarding the (adaptive) choice of particular actions, deliberation in the form of cause-effect conjectures and/or means-ends reflections can mediate the motivational force of deprived innate needs and acquired wants. Moreover, cognitive learning by inference, by own memorized experience, or by observing others (Bandura 1986, Chap. 2) can intervene into the non-cognitive adaptation patterns determined by reinforcement learning according to Hypothesis 2. For these reasons, it is necessary to add:

Hypothesis 3  Utility arises in addition from the rewarding experience associated with success in goal striving. If cognitive activities intervene into choice behavior, the relative frequencies with which alternative actions are chosen also depend on the subjective content of deliberation and the decision maker’s strength of belief in it and on her persistence.

12 Under laboratory condition it is possible to study the limiting case of one need in isolation whose degree of deprivation is held at a constantly high level. Hence, the first kind of variables is artificially held constant here. In such a situation, the relative frequencies with which the available actions are chosen have been found to be proportional to the average reward in terms of reducing need deprivation associated with the actions—the “matching law”, see Herrnstein (1990).
The first part of this hypothesis poses no particular problem. In terms of Kahneman’s et al. (1997) diction, the utility caused by the rewarding experience of success in goal striving is still “instant utility”. However, in the process of cognitive learning of how actions and utility are associated, the memorized experience takes the form of “remember utility” and cognitive deliberation leads to “predicted utility”. The two together form the basis of “decision utility”. Kahneman et al. emphasize that biases in both predicted and remembered utility can systematically bias decision utility and thus decision making.

The problem with the second part of Hypothesis 3 is that the subjective content of deliberative activities is highly context-dependent and idiosyncratic. It may lead to certain actions being delayed or chosen less frequently or being abandoned altogether, while other actions may be chosen more frequently. The modifying effect on the frequency distribution over actions that would result from the not cognitively controlled baseline behavior is therefore difficult to predict in generic terms. Cultural influences may, however, constrain in inter-personally similar ways the set of actions individually taken into account when there is a motivation to act and may bias the expectations regarding the outcome of the actions. This is a consequence, for example, of shared mental models often observed to exist within intensely communicating groups.

5 Motivational treadmills and the resurfacing of the moral dimension

As mentioned at the beginning, Bentham’s sensory utilitarianism was inspired by the twin idea of explaining the motivation underlying human behavior and of assessing its moral legitimacy—both to be accomplished simultaneously on the basis of his theory of happiness or utility. The question can therefore be raised what normative implications the behavioral perspective on utility and choice behavior summarized by the Hypotheses 1–3 may have. This question will now be addressed. As will turn out, its discussion is not affected by limitations due to the subjectivity of cognitive influences on behavior.

In a longer run perspective, adaptation of behavior by cognitive and non-cognitive learning is a constituent part of the process of cultural evolution. This process is empowered, and has been strongly accelerated over the past two centuries, by the huge increases in human productivity and per capita income—in short: by economic growth. The process has not left unaffected the vision of what happiness and a ‘good life’ mean. It gives reason to reconsider the moral appeal of these utilitarian notions in the light of the rapid development of modern economies.

Economic growth ultimately serves the end of a greater individual command of resources to satisfy the individuals’ innate needs, acquired wants, and cognitive goals. In the more recent past, the activities thus motivated have brought about an expansion of the human niche and a more intense exploitation of nature. Even though the multifaceted cultural evolution has not always been functional to this end (see, e.g., Diamond 1960, Chap. 31).

13 This is a reason for why decision subjectivism, which has a long tradition particularly in Austrian Economics, has a certain justification, see Wagner (2010) for a discussion. Attempts to explain the historical choices of a single decision maker therefore often resort to a reconstruction by means of “situational logic” as described by Popper (1960, Chap. 31).
economic growth that was made possible has eased the burden of poverty, drudgery, and malady for the human kind. This means that, on average, there is less deprivation of needs and a greater capacity for satisfying acquired wants (of which there is an increasing number) and accomplishing cognitive goals. In utilitarian terms, it is now possible to enjoy more pleasures and better be protected against suffering pains, i.e. reach a higher level of utility. Can it be expected that a continued economic growth will continue improving what is a ‘good life’ already?

There are certainly countries and people still suffering from a significant extent of poverty, drudgery, and malady for which economic growth would fulfill that expectation. But this fact only points to the necessity of paying attention to the level of income already reached when the question of further improvements is discussed. Once the income level is sufficiently high, other aims than overcoming poverty, drudgery, and malady seem to gain importance for what happiness and a ‘good life’ mean. For understanding these changes and their normative relevance it is useful to inquire in more detail into how the various motivational forces change with the resources available for their satisfaction.

Let us discuss this for the innate needs first. With a strongly increasing income it becomes feasible to reduce or even eliminate deprivation of several innate needs. (An obvious example is the need for food). The motivation to spend additional income on goods serving such needs declines comparatively quickly. By the same token, the additional instant utility that can be realized if consumption of such goods is expanded vanishes. However, there are also innate needs for which this does not seem to be the case. They are likely to cause a constantly upheld motivation to spend on their behalf when income grows further. Does this also imply that the expanding consumption of goods serving such needs continues to raise utility?

Consider the need for social status recognition that was mentioned above. If this need is deprived, a motivation to act emerges. A straightforward action available for trying to reduce deprivation is to buy goods that are supposed to have the property of signaling the desired status. Social status tends to be correlates with income, yet income is not directly observable. For that reason, the price of status symbols signaling the size of income is usually increasing with the average income in the respective social stratum. The purpose of associating oneself by accordingly priced status symbols with the social stratum one wishes to belong to (and/or the distinguishing oneself from groups with lower status) is to reduce deprivation in status recognition. Yet, with rising average income, lower income groups can sooner or later also afford such status symbols. As a consequence, the status-distinguishing character of the corresponding goods is lost and the intended need deprivation does not materialize or does not last. To continue to be able to signal the desired social status, other, usually more expensive, goods need to be consumed. This is an unstable situation (Hirsch 1978; Frank 2007, 2011)—a kind of motivational treadmill: If income is secularly rising, the motivation to act is constantly upheld. This leads to ever more expenditures on status symbols, but on average everyone just stays put in her status. This means that the intended reduction of deprivation can at best be realized transitorily. A possible utility gain derived by spending ever more is short-lived, if there is any at all.

Another example for an innate needs for which expenditures rising with the growth of income do not reduce deprivation in lasting ways is the need for sensory and
cognitive arousal. As already suggested by Scitovsky (1976), the reason is again an instability albeit one that differs from that of social status races. Need deprivation is reduced in this case through experiencing sufficiently arousing stimuli. If the need is deprived there is thus a motivation to spend on getting such stimuli. However, what is sensed as sufficiently arousing is subject to hedonic adaptation (Frederick and Loewenstein 1999). This adaptation effect tends to depreciate the arousal value triggered by the repeated experience of the same stimulus. Deprivation returns as does the motivation to reduce it. However, eliciting sufficiently high levels of arousal again requires different and/or stronger stimuli. They usually result in higher expenditures—just to find out that sooner or later hedonic adaptation destroys the utility gains again. Another round of this motivational treadmill begins.¹⁴

These examples of motivational treadmills seem to challenge the idea that a vision of happiness and a ‘good life’ could be specified independently of the level of income and opportunities reached in an economy. Once those needs that can be satiated by spending more on them have been satisfied, further rising income is increasingly spent on needs implying motivational treadmills. What is enjoyed then as pleasures are often only transitory enjoyments calling for being renewed by raising the corresponding expenditures ever more.

A motivational treadmill effect also occurs in the context of acquired wants and cognitive goal setting. The reason are the unlimited opportunities for a socially contingent learning of new wants and goals, following fashions and fads. Due to attentional limitations, newly learned wants and goals tend to replace previously pursued ones (which tend to be unlearned). As a consequence, the pleasures enjoyed are pleasures that had to be learned or thought up first, before they generated utility. But as with all learning there is a peculiar asymmetry. Had there been no continued income increases, no opportunities would have occurred to experience all the new action opportunities and to learn to appreciate them. None of the newly learned enjoyments of pleasures would be missed if the learning had not taken place. Once all the experience have been made, though, foregoing all the learned pleasures, e.g. because of a sustained decline in disposable income, would be felt as a harsh deprivation. Put differently, had the ‘better life’ that a higher income allows never been experienced, it would not have been known to be better than the ‘good life’ defined with respect to the present income level. Although this is hypothetical reasoning, a utilitarian should be concerned about its implications. They entail a rather relativistic assessment of the ‘good’ in the notion of a ‘good life’.

Moreover, questions of moral legitimation may arise. The modern version of the utilitarian program (or what is left of it) uses preference subjectivism as an argument not to engage in a discourse on the different motivations underlying choice behavior. With this argument such questions can be ignored. However, in the behavioral reinterpretation of the utilitarian program outlined here they cannot be ignored. Is additional income spent in pursuit of a motivational treadmill equally justified from

¹⁴ The consequences of this can be conjectured to express themselves in the consumption of consumer electronics, entertainment, tourism, and the social media. At the national level, expenditures motivated this way have been growing much faster with rising income than average consumption expenditures and are likely to continue to do so.
a utilitarian point of view as additional income spent on needs where deprivation can definitely be removed as in the case of true poverty, drudgery, and malady? Is additional income spent to obtain rewarding sensory perceptions from reducing deprivation in innate needs equally morally legitimate as additional spending to obtain rewarding experiences from acquired wants or accomplishment of cognitive goals? If, as in the hedonistic tradition, the moral measuring rod is the contribution to individual happiness, at least with respect to the motivational treadmill cases doubts seem warranted.

6 Conclusions

Since its inception, the utilitarian approach to explaining economic behavior has been subject to a radical transformation. The original, naturalistic interpretation laid out by Bentham was gradually turned into an abstract theory of preferences. Following a brief review of the early utilitarian program based on hedonistic conjectures about the motivation of human behavior this paper has highlighted some major modifications of that program. It has been argued that they resulted in a significant loss of material conjectures and, hence, explanatory potential regarding economic behavior. As discussed, this loss has recently led to a few attempts to extend the empirical basis of utility theory again and to revoke some of the modifications. Taking up this impulse, the paper has tried to reappraise core elements of sensory utilitarianism and to relate them to a more general motivational theory based on insights from today’s behavioral and human sciences. However, the aim of this endeavor was not only to obtain a richer theoretical foundation. As the early utilitarians were interested in both explaining behavior and elaborating on its moral assessment, the present paper has also explored some normative implications of the suggested behavioral restatement of utility theory.

Acknowledgments

Open access funding provided by Max Planck Institute for Science of Human History. Over the years in which this work developed, I have profited from inspiring exchanges on its topic with many colleagues. I should like to thank, in particular, Martin Binder, Wilfred Dolfsma, Paul Rubin, Christian Schubert, Ian Steedman, and William Wadman.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

Abbott, L. (1955). Quality and competition. New York: Columbia University Press.
Bandura, A. (1986). Social foundations of thought and action—a social cognitive theory. Negelwood Cliffs: Prentice-Hall.
Bargh, J. A., Gollwitzer, P. M., & Oettingen, G. (2010). Motivation. In S. Fiske, D. Gilbert, & G. Lindzey (Eds.), Handbook of social psychology (5th ed., pp. 268–316). New York: Wiley.
Becker, G. S. (1965). A theory of the allocation of time. Economic Journal, 75, 493–517.
Bentham, J. (1948). An introduction to the principles of morals and legislation. New York: Hafner.
Binder, M. (2010). Elements of an evolutionary theory of welfare. London: Routledge.
Burnham, T., Lea, S. E. G., Bell, A., Gintis, H., Glimcher, P. W., Kurzban, R., Lades, L., McCabe, K., Panchanathan, K., Teschl, M & Witt, U. (2016). Evolutionary behavioral economics. In D. S. Wilson & A.
Pyszczynski, T., Greenberg, J., Solomon, S., Arndt, J., & Schimel, J. (2004). Why do people need self-esteem? A theoretical and empirical review. *Psychological Bulletin, 130*, 435–468.

Richerson, P. J., & Boyd, R. (2005). *Not by the genes alone: How culture transformed human evolution*. Chicago: University of Chicago Press.

Robinson, J. (1962). *Economic philosophy*. London: Watts.

Samuelson, P. A. (1938). A note on the pure theory of consumer’s behavior. *Economica, 5*(61–71), 353–354.

Samuelson, P. A. (1947). *Foundations of economic analysis*. Cambridge, MA: Harvard University Press.

Scitovsky, T. (1976). *The joyless economy*. Oxford: Oxford University Press.

Sen, A. (1973). Behavior and the concept of preference. *Economica, 40*, 241–259.

Sen, A. (1987). *On ethics and economics*. Oxford: Blackwell.

Shizgal, P. (1999). On the Neural computation of utility: Implications from studies of brain stimulation reward. In D. Kahneman, E. Diener, & N. Schwarz (Eds.), *Well-being—the foundations of hedonic psychology* (pp. 500–524). New York: Russell Sage Foundation.

Steedman, I. (2001). *Consumption takes time—implications for economic theory*. London: Routledge.

Stigler, G.J. (1950). The Development of Utility Theory, *Journal of Political Economy, 58*, 307–327 (Part I) and 373–396 (Part II).

Stone, J. R. N. (1954). *Measurement of consumer expenditures and behavior in the United Kingdom* (Vol. I). Cambridge: Cambridge University Press.

Wadman, W. M. (2000). *Variable quality in consumer theory*. Armonk, NY: M.E.Sharpe.

Wagner, R. E. (2010). *Mind, society, and human action*. London: Routledge.

Warke, T. (2000). Mathematical fitness in the evolution of the utility concept from Bentham to Jevons to Marshall. *Journal of History of Economic Thought, 22*, 3–23.

Wilson, E. O. (1998). *Consilience—the unity of knowledge*. New York: Alfred Knopf.

Witt, U., & Binder, M. (2013). Disentangling motivational and experiential aspects of ‘Utility’—a neuroeconomic perspective. *Journal of Economic Psychology, 36*, 27–40.

Wong, S. (1978). *The foundations of Paul Samuelson’s revealed preference theory*. London: Routledge and Kegan Paul.