Taxing Utility

Terrence Chorvat*

*George Mason University School of Law, tchorvat@gmu.edu

This working paper is hosted by The Berkeley Electronic Press (bepress) and may not be commercially reproduced without the permission of the copyright holder.

http://law.bepress.com/gmulwps/art17

Copyright ©2005 by the author.
Taxing Utility
Terrence Chorvat

Abstract

In order to assess the efficiency of a tax, we should examine its effect on the behavior of individuals. In general, the less a tax affects behavior, the more efficient it is thought to be. The standard example of a non-distorting tax is a lump-sum tax, which does not change with the behavior of the taxpayer. However, this article demonstrates that behavioral distortions can and do arise from a change in even a lump-sum tax. The only truly non-distortionary tax would be one based on utility itself. Utility, which has been used as a norm for distributional analysis, is also the ideal base for efficiency analysis. In fact, any reasonable attempt to describe a minimally distortive basis of taxation will significantly resemble the notion of a tax on utility. Therefore, utility itself is the best basis for evaluation of the efficiency of a tax. Such a tax has many additional features which make it more useful for analytical purposes than lump sum taxes.
Abstract

In order to assess the efficiency of a tax, we should examine its effect on the behavior of individuals. In general, the less a tax affects behavior, the more efficient it is thought to be. The standard example of a non-distorting tax is a lump-sum tax, which does not change with the behavior of the taxpayer. However, this article demonstrates that behavioral distortions can and do arise from a change in even a lump-sum tax. The only truly non-distortionary tax would be one based on utility itself. Utility, which has been used as a norm for distributional analysis, is also the ideal base for efficiency analysis. In fact, any reasonable attempt to describe a minimally distorting basis of taxation will significantly resemble the notion of a tax on utility. Therefore, utility itself is the best basis for evaluation of the efficiency of a tax. Such a tax has many additional features which make it more useful for analytical purposes than lump sum taxes.

Income taxes and consumption taxes have assumed a very prominent place in public finance in large measure because they are thought to be relatively efficient bases for taxation. (Tresch, 2003) These taxes are considered efficient because it is believed that the behavioral response to these taxes is considered to be less elastic to marginal increases in the rate of taxation when compared with other bases of taxation such as excise taxes on goods and services. (Ramsey, 1927, Tresch, 2003) This lower behavioral response is commonly believed to result in lower losses to welfare than most other potential tax bases. However, as explored in this article, these conclusions are based on a number of assumptions about the utility functions of individuals which may not be correct.

* Associate Professor, George Mason University School of Law. The author would like to thank Michael Abramowicz, Joseph Bankman, Edward McCaffrey, Francesco Parisi, Mark Pingle, David Schizer, and Jeff Strnad for their comments on an earlier version of this paper. The author would like to thank the Lawrence Cranberg Foundation and the Law and Economics Center for their financial support.
The primary intent of this article is to put forth three very simple points. First, the standard arguments about how one measures the efficiency of income and other taxes are incomplete. This does not mean they are not useful, but they ignore many important aspects of behavior. Second, these lacunae can cause our analysis to be blind to important distortionary elements of a tax. If we do not understand what causes behavior, it will be difficult to prevent from distorting behavior unintentionally. Third, it is therefore necessary for any welfare analysis of taxation to study the actual utility functions of individuals. It is here that behavioral economics, experimental economics and other empirical economic research can have a significant impact on our analysis. In particular, this article will discuss how research tends to indicate that wealth or income enters the utility function not as single number (or scalar), but rather in multiple parts (i.e., as a vector).

As part of the discussion, this article demonstrates that the standard against which to measure the efficiency of any tax should be judged should not be a lump-sum or head tax, but rather it should be a tax based on utility itself. The article shows how such a tax could be crafted in principle, and demonstrates that the strong assumptions of cardinality commonly made in the discussion of such taxes (e.g., Feldstein, 1976) are not necessary in order to use a tax on utility as a standard for an efficient tax. Therefore, unlike the discussions of such taxes for distributional purposes, the usefulness of such taxes for efficiency analysis is not dependent on the interpersonal comparisons of utility. Of course, actually imposing a tax based on utility at the present time is perhaps even more unrealistic than a lump-sum tax; however, as discussed in section II, it is more analytically useful.
The article will also briefly discuss some of the challenges to the notion that individuals behave consistently with the idea that they have utility functions. This article argues that while there is much evidence that contradicts many of the standard notions of utility, in general the empirical evidence available does not argue against the entire notion of utility, but rather it informs us about the particular forms of utility functions which are most likely to represent preferences. This research may make the analysis more complicated, but it does not negate the analytical usefulness of a tax on utility.

I. Standard Neoclassical Argument for Income Taxation

The standard argument concerning the efficiencies of various potential tax bases begins by noting that any change in relative prices from that which would exist in the market is distortionary, unless it is compensating for an externality. (Tresch, 2003) Put in another way, we prefer that the tax system not affect behavior, unless the tax system is attempting to correct a market imperfection. For this analysis, we assume that individuals have maximized their utilities prior to the imposition of tax, and if taxes alter the way decisions are made, this will reduce societal welfare. Taxes by their nature will have income effects (for a discussion of a special case where this might not be the case, see Chorvat, 2003a). If in addition to its income effect, a tax has substitution effects, this will increase the behavioral distortion, therefore likely lowering overall welfare. (Tresch, 2003) This additional reduction in welfare is referred to as deadweight loss. (Tresch, 2003)

Income taxes are then generally thought to have lower deadweight loss than excise taxes on goods because they do not change relative prices between goods. Rather, they merely reduce the total income of the taxpayers (Varian, 1990, p. 118-119). In
addition, an income tax is also commonly viewed as a fair way to distribute the burden of the government, because in some sense, taxpayers are paying for government based on the value they receive from our modern society which the government has helped to create. (Tresch, 2003)

Actually, it is well known that income taxes do alter some relative prices. In particular, they distort both the relative prices of income producing activities versus leisure and the relative price of saving versus consumption. The utility one derives from leisure is untaxed, whereas the utility one derives from all monetary rewarding activities is reduced by the tax. (Silberberg, 1990, pp. 348-351) Similarly, the utility one derives from consuming income currently is untaxed, whereas part of the utility derived from saving (e.g., interest income) is taxed. Therefore, while income taxes are commonly thought to be less distortive than many other types of taxes, they still introduce their own types of distortions.

Consumption taxes on the other hand, do not distort between consumption in different periods. However, as with income taxes, consumption taxes distort the relative prices of monetary and non-monetary consumption, and monetary versus psychic income production. Furthermore, consumption taxes may ignore the psychic income of wealth. For example, a miser gains psychic income from stockpiles of wealth. In addition, a consumption tax will ignore the insurance value of wealth. The insurance value of wealth derives from the fact that we cannot be sure what our are monetary needs will be in the future. Therefore, a risk-averse person will save not only for planned consumption in the future, but also for possible risks that may occur in the future. (Tobin, 1958)

II. Non-Distortionary Taxation
A. Head Taxes

Economists have long understood that income taxes and consumption taxes can distort decisions between earning monetary and earning psychic income, and the income taxes can also distort between consumption in earlier and later periods. In place of income taxes, the paradigmatic non-distortionary taxes commonly used for analytical purposes are head taxes or lump-sum taxes. (Tresch, p. 308) Under these kinds of taxes, all individuals pay the same amount of tax regardless of their wealth or income. Since the amount of tax does not depend on behavior, it is normally argued that this kind of tax will not distort behavior. This analysis leads to the conclusion that the deadweight loss of such a tax is zero. However, these arguments ignore the fact that not all utility is derived from activities which require money and they also ignore the fact that income and wealth are not evenly distributed. To be specific, a lump-sum tax is likely to distort the decisions between the production of monetary versus non-monetary income. To the extent that there are some individuals for whom paying the tax may be a significant burden, these people would likely shift some of their productive activities to those that generate monetary income, instead of psychic or implicit income, and shift consumption activities from those that require a significant amount of money to those that do not. The distortion of a lump-sum tax derives from the fact that it increases the relative cost of engaging in activities that are pleasurable and require money to engage in compared to those that are pleasurable and do not require money. If one has enough money to pay the tax and comfortably exist, then perhaps the distortions induced by a lump-sum tax may be small. Even though such taxes have always been very unpopular and are unlikely to raise enough revenue to fund modern governments, these taxes are commonly used as a
benchmark against which the efficiency of other taxes is measured. While lump-sum
taxes are commonly argued to be perfectly non-distortive, we can see that such taxes can
distort relative prices between monetary and non-monetary goods, due to its requirement
that it be paid in money. Therefore, even a lump-sum tax has deadweight loss.

B. Utility as Tax Base

The only type of tax which would be completely non-distortionary would be one
on utility itself. This is not to say that we are likely to see such a tax at any time in the
near future, but rather this *conceptual* tax is helpful in considering the various distortions
resulting from alternative systems of taxation. This section will assume that a utility tax
is possible and follow the thread of the analysis.

How would such a tax operate? Taxes are generally determined by applying a tax
rate to a tax base. Here, the base would be the utility of the taxpayer, and for simplicity,
we will initially assume a flat rate of taxation. Under a tax on utility, no matter what the
activity, the relative prices would not be changed because in all cases the percent of
utility removed would be the same. To give an illustration, if one activity yielded 100
utils, and an alternate activity yielded 99 utils, one would choose to undertake the first
activity. This choice would not be affected by a tax which simply removed 20% of the
utils from each side. After-tax, the activities would yield 80 and 79.2 utils, respectively.
In other words, relative values of the activities would not change. The tax would be
assessed in such a way that paying it would always result in a reduction in utility of the
appropriate amount. The mathematical argument for this is that the effect of the tax is the
equivalent of simple linear revaluation of the utility function, which under expected
utility theory should not change choice behavior. Formally, if the utility from two
potential activities can be denominated \( x \) and \( y \), and by assumption \( x > y \), then it will always be the case that \( ax > ay \), for all \( a > 0 \). No one would ever find it advantageous to lower their own utility in order to lower their tax bill, because somewhat paradoxically, if they obtained utility from making themselves unhappy, this utility would be taxed.

One might dispute this idea by arguing that “utility” as such does not exist, rather it is a construct we have for predicting the behavior of individuals. Furthermore, the notion of “utils” is more of a pedagogical tool for the demonstration of basic principles rather than a tool useful for analytical purposes. In more formal terms, this argument is stating that utility should be thought of in ordinal terms (i.e., useful in analyzing the ordering of preferences into first, second, third, etc.) rather than in cardinal terms (i.e., directly comparable utility, such as 100 utils versus 78 utils, etc.). This ordinality of the utility function is derived from the fact that, in general, a utility function can represent observed preferences, any monotonic transformation of that function could also represent observed preferences. (Mas-Colell, Whinston & Green, 1995 pp.11-15) Therefore, in order to be of use as a basis of taxation, one has to impose stronger restrictions on the utility function.

One can argue for a tax on utility as the standard of efficiency even if one does not assume perfect cardinality of utility. The argument given above can be rephrased in ordinal terms, but restricting the form of the utility function to that used by expected utility theory (these are described in Part III.B). Because these utility functions are unique except for affine transformations (as described below), if we simply choose the utility function for each person such that the function will always have a value of zero when there is zero wealth, we can perform the calculations to demonstrate the invariance
of decision-making with and without such taxes. This concept can also apply if, as I suggest later, the argument of the utility function is a vector rather than a scalar. Part III of the article discusses some evidence which might undermine the argument that individuals can be even said to have utility functions.

Based on this analysis, a tax on utility is the ideal conceptual base, because it will not distort behavior. In addition, to the extent that any other tax departs from this standard, it can distort behavior, beyond its income effect, because any tax which differs from a utility base can alter the relative values of two or more items and therefore distort behavior.

Others have discussed using utility as a basis for assessing tax rules on distributional grounds (Feldstein, 1976). In fact, the notions of horizontal and vertical equity can be most simply stated in terms in of a comparison to a utility tax. Those making such distributional arguments have had to make strong assumptions about the cardinality of utility. These strong cardinality assumptions are not essential for an efficiency analysis of tax rules, because for efficiency analysis one does not have to rely on interpersonal comparisons of utility as you do with distributional analysis. Of course, to the extent that one does adopt these strong cardinality assumptions, the ideal base both for an efficiency analysis as well as for distributional analysis is then a tax on utility.

One could state the basis for taxation in way similar to the standard formulation of the economic definition of income, which is consumption plus net accretions to wealth. (Simons, 1938) If one conceives of the base for utility as occurring during a period of time, beginning at time $t_0$, and ending at time $t_1$, then the utility over this period would be:
where \( u(t) \) is the momentary utility at time \( t \) or if one treats the taxable period as a collection of discrete sub-periods, it would be of the form,

\[
\sum_n U(E_n), \text{ where } E_n \text{ is the wealth in the } n\text{th sub-period.}
\]

Given this base, the amount of the tax could then be

\[
\int_{t_0}^{t_1} F(u(t))dt, \text{ in the continuous case and}
\]

\[
\sum_n F(U(E_n)), \text{ in the discrete case}
\]

or alternatively either \( F(\int_{t_0}^{t_1} u(t)dt) \), or \( F(\sum_n U(E_n)) \), where \( F \) is the function which determines the tax based on the amount of utility.

The reason why such a tax will not distort decisions is the requirement that the level of the tax be such that it removes a constant percentage of utility from all possible actions. The dimensions of what this means and how this could be accomplished is discussed in the next section.

III. Utility Functions: Behavioral Economics and Neoclassical Economics

To understand how a tax on utility would operate, we first need to understand the tax base itself, i.e., the utility functions of individuals. The previous section discussed why a tax on utility is the only truly non-distortive tax. This section discusses some of the research that helps us to understand the nature of actual utility curves and what this research can tell us about what a non-distortionary or at least a minimally distortionary tax system would look like. This inquiry can aid us in understanding the distortions that our current tax system makes, as well the distortions of proposed alternative systems, allowing us to be better able to craft tax rules.
A. Deterministic Utility

The starting point for understanding utility functions is a standard utility curve. Under the most primitive form of the function, utility is simply a function of wealth (represented as U(W)). This level of sophistication is sufficient to make the standard arguments against distortions introduced by the tax system (Harberger, 1966). That is, individuals have already optimized their own utility given their wealth levels etc., and therefore altering these decisions can only reduce welfare, ignoring externalities and related Pigouvian reasons for a tax (Pigou, 1920). Of course, taxes as whole are intended to increase welfare, but not directly. They increase welfare by providing revenue to fund public goods. Public goods are assumed to increase welfare, while taxes generally decrease welfare. To properly analyze the effect of a tax system, one would have to examine the net benefits of public goods made possible by the tax revenues minus the net costs of taxation and then optimize the level of welfare using the tax as the choice variable. (Kaplow, 1994)

In addition, if we assume that the second derivative of wealth is negative \( (U''(W) < 0) \), then we can derive an argument for re-distributive taxation. If the marginal utility of a dollar to a rich person is less than it is to a poor person, (as follows from declining marginal utility of income) taxing the rich at a higher rate comes at a lower welfare cost than proportional taxation. (Edgeworth, 1919). As with almost all utilitarian redistribution arguments, this argument requires an ability to make interpersonal comparisons of utility. In addition, there is a need to balance the utility benefits of progressive taxation with the decreased incentive for the rich to earn income from taxing them at a very high rate. The analysis which derives from these simple assumptions and
the arguments which are derived from them provide the basis of optimal taxation theory. (Mirrlees, 1986) A large body of scholarship is based on the theoretical and empirical implications of these relatively simple utility functions.

Based simply on deterministic utility, it would be quite difficult to assess a tax on utility. Because observed behavior can be rationalized by any monotonic transformation of any other function which rationalizes behavior, it is impossible to determine a reasonable tax rate. That is, the tax would depend on the particular from of the utility function, which is arbitrary. Therefore, there would be no principled way to assess the tax.

B. Expected Utility

One important problem with the rather simple models discussed in section A is that they assume that there is no risk and that individuals possess perfect information. It is clear that much of human behavior is designed to deal with the fact that we do not know the future. The notion of risk is fundamental to any reasonably accurate description of human behavior. The key modification to the utility function in this case is use of an expected utility function in which the individual chooses a “portfolio” in order to maximize:

\[ \int p(x)U(x)dx, \text{ or } \sum p(x)U(x) \text{ in the discrete case,} \]

where \( x \in X \), and \( x \) is an outcome and \( X \) represents the set of all possible outcomes of an investment, gamble etc., \( p(x) \) is the either probability of the outcome \( x \) in the discrete case or the derivative of the cumulative distribution function in the continuous case, and \( U(x) \) is the utility of an outcome of \( x \).
Models of taxation based on expected utility theory commonly address the risk sharing and risk transferring aspects of taxation, as well as the appropriate valuation of investment and other property received either as compensation, in exchange for other items, or as part of a bequest. (Sandmo, 1985) These models often conclude that the behavioral effect, as well as the burden of a tax, may in fact be different than would be assumed under a riskless model (Chorvat, 2003a, Myles, 1995).

As discussed earlier, one of the beneficial features of the model is that it allows use to re-cardinalize utility to enough of an extent to allow us to be able to impose a tax on utility. Because expected utility functions are unique up to affine transformations, then if we establish a reference point (for example, it we assume a zero level of wealth and consumption has a utility of zero), utility values for any set of levels of wealth under different sets of utility functions are then proportional to each other. This causes any proportionality factor to wash out in the calculation of taxes. (Mas-Colell, Whinston and Green, 1995) So for example, if we say the a $5 level of consumption has a utility of 5 and a $10 level has a utility of 9.8, then if we instead denominate the $5 consumption as a 10 utils and $10 at 19.6 then a 20% tax will still not alter the decision between and the amount of the same whether we used the first set of valuations or the second. That is, $x > y$, then any for any positive proportional scaling $ax > ay$, and similarly if $(1-t)x > (1-t)y$, then $(1-t)ax > (1-t)ay$, where $t$ is the tax rate.

For expected utility theory to be correct, individuals would have to know precise probabilities for each event. Because this was perceived as being somewhat unrealistic, Leonard Savage developed a theory that is known as subjective expected utility theory, under which individuals behave as if they have a coherent subjective probability for each
potential event, even though not all persons might objectively agree to this probability (Savage, 1954) The calculations of subjective expected utility then use this subjective probability in the expected utility calculations described above.

While subjective expected utility would be based on subjective probability assessments, this does not in principle void the notion of using these utility functions for the assessment of tax. While the probability assessments are personal to the individual, they are still capable of assessment through studying the behavior of individuals.

C. Behavioral Utility

Since the formulation of expected utility theory and its close relative, subjective expected utility theory, many scholars have argued that the models based on the predictions of these theories do not accurately represent many features of human behavior. This literature is sometimes referred to as behavioral economics. Many types of behavior which are inconsistent with expected utility theory have become staples of this literature.¹ Many scholars have argued that by moving beyond standard expected utility theory, we can improve our understanding of utility functions if we introduce notions from behavioral economics such as mental accounting, under which the value of a dollar of wealth in one form is not necessarily the same as the value of a dollar of wealth in another form, as well as anomalies in how individuals perceive risk and uncertainty. This section discusses a number of empirical phenomena which many scholars have argued should be incorporated into our models of economic behavior.

1. Allais and Ellsberg Paradoxes

Almost as soon as von Neumann and Morgenstern formalized expected utility theory, Maurice Allais pointed out that individuals appear to violate its independence axiom.
when comparing bets that are certain to those which are not.² (Allais, 1953) Allais presented a “paradox” or more accurately, he suggested a number of sequences of gambles for which he argued that many people would violate expected utility theory. A commonly used illustration of the behavioral paradoxes is the observation of how individuals behave when confronted with choices similar to the following:

First, select one of two gambles

A: win $1,000 for sure and

B: you have a 10% chance to win $2,000, an 89% chance to win $1,000, and a 1% chance to win $0.

Now, choose which of the following two gambles you prefer:

C: an 11% chance to win $1,000, and an 89% chance of winning nothing versus

D: a 10% chance to win $2,000, and a 90% chance to win zero.

Under expected utility theory, the utility of a gamble is always the linear combination of the utilities of the various outcomes multiplied by their probabilities. Given this, it can be easily demonstrated that an individual who chose gamble A, should choose gamble C, nonetheless those who choose A commonly select D. The weight of experimental evidence indicates that many people do behave in the way predicted by Allais. (Camerer, 1995)

One of the ways this behavior is explained is to postulate that certainty is somehow valued discontinuously (Camerer, 1995). This would argue that individuals value sure bets over risky bets in a way that is not consistent with standard neoclassical analysis, at least as it has come to be known in either expected utility theory or subjective utility theory. Another explanation comes from the fact that it appears that often
individuals do not treat compound lotteries in the way predicted by expected utility theory (Kahneman, Knetsch, and Thaler, 1991). There are a number of other proposed alternative theories. (Camerer, 1995)

While behavioral phenomena, to the extent that they do not violate rationality, may make the computation of the utility functions more complicated, they do not negate the notion of a tax based on utility. In fact, they likely make the utility functions even more unique than even expected utility theory did and so increase the theoretical feasibility of such taxes, although it may complicate the practicalities of it.

About a decade after Allais described this behavioral paradox, Daniel Ellsberg published an article that criticized the assumption of subjective expected utility that individuals behave as if they have subjective coherent probabilities for events. (Ellsberg, 1961) He argued that uncertainty can be divided into risk and ambiguity and that these two can operate independently to alter choice behavior. This argument can be illustrated by the choice between two different gambles. Imagine that there is an urn which contains 30 red balls and 60 balls additional balls which are either black or yellow, but we do not know the distribution of those two colors of balls. You will select between two gambles:

A: you receive a payoff of $10 if a red ball is selected or
B: you receive a payoff of $10 if a yellow ball is selected.

If you select A, then under subjective expected utility theory, this indicates that you believe the probability of red (P(red)) is greater than the probability of yellow (P(yellow)), or P(yellow) < 1/3. Because by assumption, twice the P(red) is equal to P(yellow) + P(black), this implies that P(black) > P(red), or numerically P(black) > 1/3.

Now imagine you asked to select between two additional gambles:
C: you receive a $10 payoff if red or black is selected

D: you receive a $10 payoff if yellow or black is selected.

If you selected A and then selected D, you have violated the assumption of coherent probabilities. If you believe that \( P(\text{black}) > \frac{1}{3} \), then \( P(\text{black or red}) > \frac{2}{3} \). However, by assumption \( P(\text{black or yellow}) = \frac{2}{3} \). Therefore, if one chooses A, you should choose C. If you chose A and D, you have indicated that you are averse to gambles with ambiguous probabilities.

Experiments that have been conducted following up on Ellsberg’s famous article have shown that many individuals do indeed behave the way that Ellsberg predicted. (Camerer, 1995) It is commonly concluded that this implies that the lack of knowledge or ambiguity in probability also seems to affect utility functions (see chapter 6 of Mas–Colell et al., 1995). This distinction between known probabilities and unknown probabilities had been discussed earlier by Frank Knight, (Knight, 1921) as well as John Maynard Keynes (Keynes, 1921).

By adding ambiguity to the utility function, it again complicated the function, but does not diminish the theoretical possibility of such a tax. Ambiguity seems to introduce a new kind of uncertainty, but one which is likely to behave in ways similar to risk. In fact, the behavior of individuals with respect to ambiguity may be closer to standard expected utility theory than with regard to risk. (Smith et al. 2002)

2. Mental Accounts

In addition to the Allais and Ellsberg paradoxes, there have been a number of other empirical phenomena which appear to violate expected utility theory. One example is that individuals appear to divide various forms of wealth into separate mental accounts
rather than including all wealth in one joint calculation. (Thaler, 1999). Under standard expected utility theory, the argument of the utility function is simply wealth. Mental accounting would imply that the argument of the utility function is a vector, rather than a scalar.

The notion of mental accounting was developed to explain certain inconsistencies between expected utility based models and observed behavior. Mental accounting has been used to explain the empirical evidence which indicates that the marginal propensity to consume is not uniform among different types of income, but varies with the form of income. For example, the marginal propensity to consume income from increases in the value of one’s stock market portfolio is far smaller in comparison to that for labor income than would be predicted by the permanent income hypothesis. One study conducted by Case et al. looked at the effect of stock ownership and other forms of wealth on the marginal propensity to consume. They found the marginal propensity to consume earnings from stock was very small (below 1% or a marginal propensity to consume of .01) and the marginal propensity to consume income out of housing wealth was higher (about .09). This was contrasted with the marginal propensity to consume for labor income which is estimated to be between .6-.8. (Case et al., 2001, Cheng and French, 2000). The authors of this study argue that this occurs because wealth in the form of one’s house is viewed as being more stable. These findings would indicate that the utility function of individuals are therefore not simple functions of wealth, but have at least three variables in the argument (i.e., stock market income, housing income, and labor income). While one might rationally expect the marginal propensity to consume out of stock wealth to be less than labor income (because the first might arguably be a one-time
shock and the second may represent a permanently higher stream of income), almost any reasonable theory based on the permanent income hypothesis cannot explain a difference of this magnitude, nor can it explain the difference between housing wealth and stock market wealth. In addition, recent experimental economic evidence suggests that individuals keep wealth from risky assets such as stock in a separate mental account from wealth in less risky forms of wealth, such as bonds. (Rockenbach, 2004).

Another example in which individuals seem to form mental accounts is the behavior exhibited by contestants on game shows. (Gertner, 1993) In one study, actual bets on a game show studied averaged $3,200. When the cash stake available from betting increased by $1, bets increased by about $.60. But when the amount of earnings that could not be bet increased (e.g., when a contestant won a car) by $1, bets increased by only a penny. While the contestant may not have valued the prize as much as the stated value, it seems highly unlikely that they would value it at only 1% of stated value (they could almost certainly sell it for substantially more than that). Similar results have been found by others. (Thaler and Johnson, 1990, Biswanger, 1981) This behavior is inconsistent with the prediction from neo-classical economics that contestants should integrate all assets (bettable cash and unbettable prizes) then bet based on their integrated assets. In particular, based on the behavior of individuals, non-liquid assets appear to be valued at a substantial discount to liquid assets of equivalent market value.

B. Anomalous Time Discounting

It has been observed that individuals do not seem to follow the standard exponential discounting for determining the value of future payments. (Strotz, 1956). One of the commonly used ways to incorporate this concept into utility functions is the $\beta$-$\delta$ model of
David Laibson, in which there is a distinction between consuming now and consuming later in addition to standard exponential discounting. Under standard exponential discounting the value of any payoff which will be received $n$ periods from the present is of the form $\delta^n U(w_n)$, where $\delta$ is the discount rate per period, and $w_n$ is the wealth that will be received in the $n$th period. Under the Laibson model, if you are to receive the payoff now, its value is $U(w)$, but if you are to receive it $n$ periods in the future it is $\beta \delta^n U(w_n)$, with $\beta < 1$. There is now some neurological evidence that individuals conceive of trade-offs between now and the future in a manner consistent with this model. (McClure, Laibson, Lowenstein, and Cohen, 2004)

To the extent that individuals discount for the time value of money in a non-exponential fashion, financial models based on exponential discounting will not predict the behavior of individual investors. It has been argued that we can better explain many of the behavioral anomalies such as the equity premium puzzle by introducing models based on these non-exponential time discounting models (Benartzi and Thaler, 1996).

Another way in which individuals appear to violate standard exponential discounting models is the violation of what is referred to as the permanent income hypothesis. The standard model of the permanent income hypothesis is that the marginal propensity to consume out of all expected lifetime wealth, whether in the form of stocks, real estate, or any other source, should be the same relatively small number. That is, ignoring the time value of money and uncertainty, if an individual expects to live $N$ periods and will have a total lifetime wealth of $W$, the individual will want to consume $W/N$ each period. An increase in wealth in one period will increase lifetime wealth by only a small percentage and so should increase current consumption by only a small amount. The amount of
lifetime wealth (W) needs to be discounted for when the wealth arises, as well as the uncertainty surrounding it and the value of future consumption will also have to discounted as well. The empirical studies of this question tend to find that while there is a fair amount of smoothing of consumption over a lifetime, there is also an excessive relation between consumption and the actual realization of income. (Flavin, 1981, Campbell and Mankiw, 1989) Hence, a short-term increase in income received increases the amount of consumption more than the permanent income hypothesis predicts.

Interestingly, consumption patterns at odds with the permanent income hypothesis can also be found in to day-to-day consumption patterns. One interesting finding is that people spend more on payday than on other days. (Stephens, 2002a, Stephens, 2002b) It does not appear that this phenomenon is significantly related to liquidity constraints because in fact, the studies have found little correlation between credit cards and increased spending on payday (although wealth and other socioeconomic factors did play a key role).

3. Loss Aversion

One important way in which individuals appear to depart from expected utility is the method of accounting for potential losses, which may be related to the explanation of the Allais and Ellsberg paradoxes. An area which has been a particular focus of study has been the disposition of portfolio assets. Traditional expected utility theory, in conjunction with the portfolio theories, such as the efficient market hypothesis which is based upon expected utility theory, predict that investors should be indifferent between selling or retaining any particular portfolio asset, because it is fairly priced. Under a tax system which does not assess a tax until there has been a sale (or more accurately a
realization event), investors should sell their investments that have lost money and retain those which have increased in value. This is the most rational thing to do if you assume that the current price is the correct value of the asset, and you can defer the tax on the any gain by retaining those assets which have increased in value. (Odean, 1993 and Sheffrin and Statman, 1985) However, empirical evidence indicates that, in fact, individuals sell the stocks that have increased in value and retain those which have lost value longer than they would if they were behaving “rationally”, even ignoring tax incentives. This phenomenon extends beyond common stocks to include stock options, and residential housing. (Grinblatt and Keloharju, 2000, Heath et al., 1999).

d. Functional Forms of Behavioral Utility

Just as there are many behavioral phenomena which appear to violate standard expected utility theory, there are also a number of theories which attempt to correct for these deficiencies. One of the first was put forward by Harry Markowitz, who argues that it is possible that utility functions are best viewed not as based simply on wealth, but rather changes to wealth. (Markowitz, 1952) In addition, individuals may be risk averse as to gains (i.e., they have a concave utility function for gains) and risk preferring as to losses (they have convex utility function for losses). Of course, these concepts are also key elements of behavioral economic models such as prospect theory (Kahneman and Tversky, 1979).

There is a fair degree of evidence from the emerging field of neuroeconomics that individuals do not perceive gain and loss in the way predicted by expected utility theory. This evidence also helps to support some of basic notions of prospect theory and its variants, while adding intriguing differences. One study has found that risky situations
involving losses and risky situations that involve gains are perceived in two different regions of the brain. In addition, situations with ambiguous probabilities are perceived in yet another region of the brain. (Smith et al., 2002) Interestingly, this study found that risky gains and losses were valued along the lines of prospect theory, while gains and losses with ambiguous probabilities were valued in ways more akin to expected utility theory. (Chorvat, 2002) This evidence helps to demonstrate that Ellsberg’s distinction between risk and ambiguity in fact has a neurological basis. The neurological evidence indicates that the reference point from which calculations are made would appear to be the expectations of wealth or gains rather than the amount of wealth prior to the event. Therefore, to extent that future research findings are consistent with these results, a key element of any understanding of investor behavior will be the nature of how expectations are formed. Interestingly, prospect theory discusses how the nature of the “framing” of the problem can result in a different behavior. One recent study on the differences in framing then can significantly affect perception of an event (McClure, Li, Tomlin, Cypert, Montague and Montague, 2004)

These experiments do not necessarily directly contradict expected utility theory for at least two reasons. First of all, the argument of the utility function could easily be something other than terminal wealth, such as changes to wealth. Second, the tasks involved in the overwhelming majority of such experiments only test very short-term processing rather than time frames normally considered in economics and therefore the implications for longer decisions are unclear. Nonetheless, the results of these experiments certainly paint a different picture of what is occurring in the decision-making processes of individuals.
4. The Existence of Utility Functions

In recent years, there have been a number of attacks not only on expected utility theory and its variants, but on the notion that individuals can be said to have consistent utility functions at all. The idea of a utility function has been skeptically received from its inception. Samuelson’s revealed preference theory was an early attempt to get away from the notion of a utility function. (Samuelson, 1938) However, as is well-known, if there are only a finite number of possible choices and a rich enough choice set, then revealed preference theory and utility theory are mathematically equivalent. (Mas-Colell, Whinston and Green, 1995) The phenomena discussed in the earlier sections such as non-exponential discounting and mental accounting can be interpreted as modifications to utility theory which can be accommodated by making the utility function more complex. However, to the extent that experiments demonstrate that individuals not only have some odd risk preferences (See e.g., Rabin, 2000, Rabin and Thaler, 2001) but actually intransitive preferences (see, e.g., May 1954, and Lichentstein and Slovic (1971)) it is difficult to discuss the notion of a utility function at all. Transitive preferences are an essential element for the existence of a utility function (Jehle and Reny, 2001)

Another example of this type of challenge can be found in the distinction, championed most recently by Daniel Kahneman, between experienced utility and behavioral utility. (Kahneman, 2003) This research indicates that there is a difference between the happiness we experience at the time of an event and the utility that influences our behavior in later situations. If these are in fact two different things, then the notion of single utility function for individuals may be problematic. Standard utility
theory assumes that we make choices so as to optimize our well-being given our constraints.

Another problem is the likely heterogeneity of the functional form of utility functions among individuals, as opposed to merely heterogeneity in the parameters of the functions such as risk aversion. That is, individuals might not only have different tastes for wealth or other goods, but the functional forms themselves (e.g., standard expected utility versus rank dependent expected utility) might be heterogeneous. There is evidence from many experiments that some individuals follow expected utility theory, while many others do not. (Camerer, 1995) While this would complicate the application of a utility tax, in that it might imply the tax levied on different individuals might have different utility functions, in principle it would not alter the conclusion that a tax based on the utility function would be the most efficient tax.

D. Applications to Taxation

The notion of utility as the most efficient basis for taxation can be applied to the analysis of potential systems of taxation in a number of ways. Conceiving of the norm as a tax on utility causes us to focus on the actual utility curves of individuals. Using a tax on utility as a standard can help to clarify notions such as the taxation of psychic income and related items such as the potential taxation of value of owner-occupied housing, which many countries in fact tax. (Ault & Arnold, 2004) Such ideas are much more readily apparent from a utility tax base than from a monetary income base.

A utility analysis may to help to explain a number of tax law doctrines, which are nearly universal and would otherwise be quite anomalous. One such doctrine is the realization doctrine. Under this doctrine, income or losses from an investment in
property are not included in taxable income until the property is sold and the income from it is “realized”. This doctrine is often argued to be the cause of many of the problems of the tax system, because it fails to track income as it is earned. (Chorvat 2003b) As discussed earlier, if utility is based on wealth, the failure of the tax laws to track the economic definition of income (Simons, 1938) will result in distortions to behavior which are caused by the tax system, resulting in deadweight loss. However, if the realization doctrine is in fact more in accord with utility curves of many individuals, then it may actually be increasing the efficiency of the tax system.(Chorvat, 2003b) Based on the econometric evidence which shows that individuals do not consume very much out of increases to stock market wealth, one can argue that it is at least plausible that unrealized stock market wealth is not valued as highly as other increases to wealth. This implies that unrealized income may not be valued as highly as realized income, and hence should not be taxed as severely.

A utility analysis of the tax system might also help us to think more positively about our current hybrid income tax system (which is neither a consumption tax nor pure income tax). One might argue that actually the current income tax structure comes closer to taxing utility than either a pure income or pure consumption tax. This is true because neither consumption nor income is not necessarily the same as utility. As noted, income might simply be saving for future consumption and so taxing it as it is earned would be too early, while a consumption tax would not reach significant portions of economic income, such as the insurance value of wealth. While one might argue that the current tax system does not necessarily reach utility particularly well either, such an analysis
does provide an argument for a system which is different than either a pure consumption or pure income tax, but rather a combination of the two.

Another example of where utility might be helpful as an ideal base would be the analysis of provisions which impose lower rates of tax on retirement saving. To the extent that one invests in saving for retirement, you will not enjoy the increases in value in your retirement account for decades. If individuals utilize hyperbolic discounting or other non-exponential methods of accounting for the time value of money, then they will likely value increases in our retirement accounts significantly less than increases in salary or other forms of wealth. (Laibson, 1997) It is a common experience that we know we must save for retirement, yet we do not enjoy it very much. One might argue that it is then proper to tax earnings on our retirement accounts at a lower rate because we do not enjoy such earnings as much as other income. Therefore, if we were to tax them at the same rate, they would be overtaxed. 4

Another application for which one might use the utility analysis described above is to incorporate the research on concepts such as prospect theory or other income based expected utility theories. These models argue that it is changes to wealth, rather than wealth itself that people value. To the extent this is true, income taxes would more closely track utility than wealth taxes. In fact, taken to its logical extreme this line of inquiry would argue for a transactions tax, because many experiments indicate that this might more closely approximate how individuals appear to perceive utility. However, the strength of this conclusions would depend on precisely how individuals assess gains, and the extent to which they group transactions together.
Using utility as a base would also be helpful in understanding the question of whether we ought to give regional adjustments to income tax rates. The argument is that, in reality, a dollar of income in Los Angeles is not really the same thing as a dollar of income in Little Rock. (Knoll and Griffith, 2003). To the extent this is true, the use of utility as the tax base would pick up these differences and tax them appropriately.

Others have also discussed the impact of behavioral economic considerations to the analysis of tax rules. This has mostly been to explain how the rules we observe are the result of voters not necessarily fully understanding or caring about the impact of tax rules on them (McCaffery, 2000, McCaffrey and Baron, 2004). This research helps to explain why we observe the tax rules that we do. Furthermore, it may help to choose the optimal tax rules, by allowing us to better understand the choice set we face.

It is not the intention of this article to argue in depth for any of the particular applications discussed, but rather merely to show that this analysis can be applied to the most basic questions of the tax law. In addition, because these arguments are empirical in nature, it is of course true that all of these arguments depend upon the evidence that further research will discover. It is therefore incumbent upon us to go out and discover how individuals actually behave.

CONCLUSION

Any reasonable attempt to describe a minimally distortive basis of taxation will significantly resemble the notion of a tax on utility. Therefore, utility itself is the best basis for evaluation of the efficiency of a tax. This article argues that the paradigmatic
base against which taxes should be measured should not be a lump-sum tax, but rather a tax based on utility.

There are three fundamental points that this article has tried to demonstrate. First, that the standard arguments about how one measures the efficiency of income and other taxes are incomplete in part because they generally assume that all forms of wealth, no matter how derived, are equally valued. Second, because wealth and income do not enter into most individuals utility functions in so simple a manner, standard income, consumption or wealth taxes can be cause significant distortions. Third, therefore, we need to consider actual utility functions in order to craft optimal tax rules. Here, behavioral economics, experimental economics and other empirical economic research should play a significant role in our analysis impact on our analysis.

It has long been acknowledged that humans do not think precisely in the way some standard economic models have considered. This was acknowledged by the early marginalist scholars. It is now clear that utility functions, to extent that they can be argued to exist, are likely to be quite complicated. In order to truly understand the effects of a tax, we need to understand how individuals actually respond both to improvements to their welfare as well as to the tax itself.

Bibliography


John Maynard Keynes, *A Treatise on Probability* (1921)

Frank Knight, *Risk, Uncertainty, and Profit* (1921)


Agnar Sandmo, *The Effects of Taxation of Savings and Risk* in the 1 *HANDBOOK OF PUBLIC FINANCE* (Auerbach and Feldstein eds. 1985)


---

1 This literature is vast. For a discussion of many of these “anomalies” see the chapters in *CHOICES, VALUES AND FRAMES*, (Daniel Kahneman et al. eds, 2000).

2 Under the independence axiom, if Gamble X has a value greater than Gamble Y, and if gamble Z is independent of X and Y, then the value of Gamble (X+Z) must be greater than the value of Gamble (Y+Z). This axiom, rather like Euclid’s fifth postulate, is less obvious than the other axiom’s of von Neumann-Morgenstern expected utility (and consequently Savage’s subjective expected utility). Because if this it has drawn more criticism than the other axioms. The Allais Paradox was first discussed in Maurice Allais, *Le Comportement de l’Homme Rationnel devant le Risque: Critique des Postulats et Axiomes de l’Ecole Américane*, 21 ECONOMETRICA 503 (1953)

3 In order to make a prediction about how someone will react to changes in the market, we have to first understand how the individual will perceive those changes in the market. See Vernon Smith, *Microeconomics Systems as an Experimental Science* 72 AMERICAN ECONOMIC REVIEW 923 (1982). In particular, the one purpose of institutions is to generate messages which are perceived by the individual. The purpose of economics experiments is to determine how individuals behave given particular environments and particular institutional settings.

4 It is of course possible to argue that earnings in the retirement simply allow individuals to save less now.

5 For example Alfred Marshall stated “In all this they deal with man as he is: not with an abstract or “economic” man; but with a man of flesh and blood. They deal with a man who is largely influenced by egoistic motives in his business life to a great extent ... but who is also neither above vanity and recklessness, nor below delight in doing his work well for its own sake, or in sacrificing himself for the good of his family, his neighbors, or his country, a man who is not below the love of a virtuous life for its own sake.” Alfred Marshall, *PRINCIPLES OF ECONOMICS* 22 (1890).