

Jack L. Knetsch

The Curiously Continuing Saga of Choosing the Measure of Welfare Changes

Abstract: The results of the vast array of willingness to accept compensation/willingness to pay (WTA/WTP) disparity studies provide strong evidence that people value many losses and reductions of losses, more, and often much more, than otherwise commensurate gains or foregoing of gains. These findings also make it clear that people commonly value many changes not as final states as standard theory assumes, but as positive or negative changes relative to a neutral reference state. Consequently, not only are losses to be most accurately assessed with the WTA measure, but most positive changes that reduce losses are as well. Current practice, which rarely takes such reference dependence into account, is therefore likely to substantially understate the value and importance of projects, policies, and programs that reduce losses. Failing to take the possibilities of valuation disparities into account also appears to undermine other kinds of analyses as well, including, for example, the estimation of elasticities and setting effective levels of Pigouvian taxes.

Keywords: reference state; willing-to-pay; willing-to-accept; valuation disparity.

JEL classifications: D03; D61; H43; I31; Q51.

Introduction

In the past, the choice of a monetary measure of the change in welfare associated with an action, program, or policy change seemed an easy one: following the guidance provided by standard economic theory, all positive changes were taken as gains bringing about welfare improvements, and were assumed to be accurately assessed by the maximum sums individuals would pay (the willingness to pay [WTP] measure) to receive them; and all negative ones yielded losses, and would be just as precisely measured by the minimum amounts they would demand to accept them (the willingness to accept compensation [WTA] measure). Standard theory eased the burden still further by adding the assertion that the two measures would yield the same value, except for the possible impacts of any income or wealth

Jack L. Knetsch: Simon Fraser University, Canada, e-mail: knetsch@sfu.ca

effects, and as these would almost certainly be inconsequentially small in nearly all cases – “... we shall normally expect the results to be so close together that it would not matter which we choose” (Henderson, 1941, p. 121). The empirical assertion implied by this interpretation gave full license for analysts to responsibly ignore the issue completely and use whichever measure was most convenient. By and large, the WTP measure became, in practice, the metric of choice for essentially all changes – positive and negative, gains and losses, those that forego gains, and those that reduce losses.

Recent and not so recent empirical evidence, including but not limited to the hundreds of studies (and surveys and experiments included in them) reviewed by Horowitz and McConnell (2002) and by Tuncel and Hammitt (2014), not only calls these easy prescriptions for the choice of measure into serious question, but indicates the seriousness of the distortions that continuation of the current choice of measure practice is likely imposing on individuals and on the wider community. While these studies demonstrate that not all valuations are subject to significant disparities between positive and negative changes, they also show the pervasiveness and wide range of changes that people do value differently – and also how ones with larger differences tend strongly to be ones that are the subject of benefit–cost analyses, risk assessments, and other more explicit weighing exercises when the choice of measure is likely to be of substantial practical importance.

These behavioral findings have also made it clear that people value many gains and losses, not as final states as assumed in standard theory, but in terms of positive and negative changes relative to a neutral reference state, which may, or often may not, be the status quo or necessarily ones determined by explicit extant legal entitlements. An important implication of this reference dependence for the choice of measure, is that a positive change may be either a gain if in the domain of gains (above, or beyond, the reference state) or a reduction of a loss if in the domain of losses (below the reference state); and a negative change may be either a loss if in the domain of losses or foregoing of a gain if in the gains. And to the extent that the point of estimating the monetary measure of the change in welfare is to assess impacts of the change on community welfare, which is usually the case, and that changes in the domain of gains are usually best assessed with the WTP measure, and those in the domain of losses with the WTA measure, as is normally the case, the current practice of near universal use of the WTP measure for any and nearly all changes will frustrate proper accountings of welfare changes in the likely many cases where sizable disparities between the measures can be expected. In choosing between a project that would provide a gain and an equally costly one that would reduce a loss, for example, it is assessment of the former with the WTP measure and the latter with the WTA measure that will best ensure accurate guidance.

WTA and WTP

The two traditional compensating variation monetary measures of welfare changes are the WTP for a gain and the WTA for a loss, in each case the sums that leave the individual at the same level of well-being as the before-change reference level to which it is compared – graphically, the amount that leaves them on the same indifference curve. Paying the WTP sum returns the individual to the welfare level of the reference state and is therefore the equivalent of the welfare increase stemming from the change at issue. Compensation of the WTA sum returns the person to the original reference level and is therefore an accurate measure of the value of suffering a loss.

The two less traditional, but equally valid, equivalent variation monetary measures of welfare changes, those given prominence by the reference-dependent nature of many valuations, are (1) the WTA valuation of actions that reduce or eliminate a loss and (2) the WTP valuation of foregoing a gain. Being a change in the domain of losses, the former is the minimum sum the individual will demand to forego a change to the higher level of well-being of the reference state (the WTA measure).¹ Being a change in the domain of gains, the latter is the minimum amount the individual would give up to avoid a change to the lower level of the reference.² In both of these cases too, the sums are derived from comparisons to reference state levels of well-being. In the case of the WTA to forego the higher reference level, it is the sum that would be required to return to this level and is therefore the equivalent of the value of the reduction (or elimination) of the loss. In the case of the negative change in the domain of gains, it is the sum that people would be willing to pay rather than move to the lower reference level of well-being.

Traditional criteria for the choice of measure typically do not consider the reference dependence of people's valuations of changes. Consequently, reductions, or eliminations of losses, are likely to be systematically undervalued. Taking reference states into account in distinguishing when the WTA and the WTP measures are appropriate should result in more accurate assessments and less misleading guidance.

1 It seems more likely, for example, that people would regard being free of a physical assault as their expected, or reference, state, even if they are being beaten at the time. Given this to be the case, a positive change of stopping a physical assault is then a reduction, or elimination of a loss, rather than a gain. To the extent that this is indeed the case, the value of such an action is then appropriately assessed with the WTA measure and not the WTP measure. There appear to be many parallel cases, ones that also call for WTA rather than WTP measured assessments.

2 An early relating of the standard welfare measures to more contemporary assessment issues was suggested by Zerbe (2001), with more details and added comment provided, for example, by Knetsch, Riyanto and Zong (2012).

If there is little or no difference in the valuations of changes by either the WTA measure or the WTP measure, as traditionally assumed in expositions of standard theory and as reflected in essentially all analysis practices, the sorting out and use of appropriate measures for particular changes have limited practical importance beyond protecting against misguidance stemming from the use of the inappropriate measure – the final sums and the guidance provided by them would be the same. The choice of measure issue takes on significant practical importance to the extent that results of WTA and WTP valuations differ in statistically significant and economically meaningful magnitude. And for many changes, the array of evidence appears persuasive that this is the case – and has for many years.

There were a few early interesting and important studies and informed speculations of the issues raised by possibilities of significant gain versus loss valuation disparities, ones not due to any income or wealth effect nor to any other explanation offered by standard theory. These included the early study by Hammack and Brown (1974), which found duck hunters were willing to pay an average of \$247 to preserve a marsh area important for maintaining duck numbers, but would demand an average of \$1,044 to agree to its loss – a finding most analysts credit as providing the first empirical demonstration of the disparity. Most observers at the time, seemingly including the authors of the study, appeared to attribute this difference more to possible unspecified weaknesses in the survey design than to any possible reflection of real preferences differing from specifications of standard theory.

It was the publication of the much-cited prospect theory paper by Kahneman and Tversky (1979) that many people take as signaling the start of the continuing contemporary focus on tests for such differences and implications of the many that have been empirically confirmed. This now³ includes a very wide array of hypothetical stated preference surveys, real exchange experiments, and a variety of real life revealed preference natural and field experiments – some providing more persuasive evidence than others. Three that involve people making real, non-trivial, decisions might be taken as illustrative of the common findings in this literature.

In one, teachers were offered different incentives to increase the numbers of their students passing a standard examination (Fryer et al., 2012). In one “treatment” the teachers were offered a bonus of up to \$8,000, dependent on the number of students passing the examination. This offer of a gain had very little, and no statistically meaningful, positive impact on outcomes. In another treatment, the bonuses remained the same, but teachers were given the maximum sums they could

3 This was not always the case as those controlling the content of economics journals, in particular, commonly exercised a very high level of caution before allowing such detractions from standard theory to be read from their pages – a “tradition” that a few continue even now. As some have noted, this reluctance in itself was a clear demonstration of the gain/loss disparity they were presumably trying to protect against.

earn in advance and told that they would be required to pay back a portion of their bonus for every student who did not pass the examination. The prospect of a loss had a dramatic positive impact on student outcomes. Clearly, the possibilities of gaining or of losing significant sums of money provided very different valuations of the bonuses – their loss was weighed much more than their gain.

In another well-known case, employees' contributions to their pension schemes were increased dramatically when they were changed from being a reduction from their current wage, to foregoing part of a gain of future increases in their wage (Thaler & Benartzi, 2004). In what is a common drill, particularly with incoming employees, they are told their wage. They presumably take this as a firm expectation and it becomes their reference state for judging the importance of anything having an impact on it. When then asked how much of that "reference" wage they would like to give up for their pension fund, they apparently tended to see this as a loss from their reference income and responded with what was widely regarded as an unsatisfactorily low average of 3.5%. When employees were instead asked, not how much they would give up, but how much of an increase (that is, gain) in their wage they would forego to add to their pension fund, this less onerous giving up of a gain resulted in average contributions of 13.4%. A similar change to the elicitation of employee pension contributions has now been carried out in numerous employment venues, with essentially the same results in each case.⁴

In a completely natural experiment, stroke records of professional golfers competing in PGA tournaments revealed that they putted significantly more accurately when attempting to save par rather than score a bogey (one over par), than they did when trying for a birdie (one below par), even though final placement of competitors (and the prize money) is based on total strokes over all rounds of the tournament (Pope & Schweitzer, 2011). The apparent greater value associated with the loss of a stroke by incurring a bogey over the lesser value of gaining a stroke by scoring a birdie, prompted this greater, or at least more effective, effort.

A good bit of the comment on the findings of disparities in people's valuations of gains and losses, has centered on such differences being due to people's mistakes, because of their use of heuristics, thinking too fast, or whatever, rather than being due to the valuations reflecting their real preferences. While this may not be a wholly settled matter in all cases, it is telling that in the case of the professional

⁴ Given the differing views of economists on the credibility of many findings of behavioral economics research, particularly of the gain/loss disparity, perhaps an interesting study might have been carried out to find out how many of a wide cross-section of them would predict that the change in how pension contributions are elicited would *not* work, that is, would not have had any impact on contribution rates. The same might also have been of interest for the teacher-bonus study – or, for that matter, for any of a large number of other clear demonstrations of a reference effect that is not plausibly explained by standard theory.

golfers, who presumably know what they are about and repeatedly take part in tournaments where large sums of money are present to concentrate their minds, they were not at all surprised when confronted with the recorded data indicating the disparity in their putting behavior. They uniformly said they were aware of the difference in their putting strategies, and justified their deliberate choice to play in this manner by their “not wanting a bogey” – choices based on preferences that included unequal weightings, rather than ones resulting from mistakes.

Although much of the literature that has been developed on the gain/loss valuation disparity has resulted in the voluminous collections of disparity observations, there have also been some presenting contrary results, critiques of some of the evidence, as well as studies that help define some of the conditions in which differences are to be expected and those in which they are less likely. Many of the seemingly contrasting results appear to result from a lack of appreciation of the extent to which people’s valuations depend on the reference state, and the extent to which various experimental manipulations shift these reference states and that it is then these shifts that give rise to the observed valuations rather than an absence of reference dependence (see, e.g., Koszegi & Rabin, 2006 and Knetsch & Wong, 2009). An example of the more useful explanatory suggestions, backed by persuasive experimental results, is that in the case of common consumer goods used to test for valuation differences, a goodly portion of the reluctance to give up such goods may well be due, not just to a feeling of loss as generally suggested, but to individuals’ hesitation to accept a sum they know to be less than the good’s market price even when they actually have little use or desire for the item (Weaver & Frederick, 2012). This of course, does not mean that the disparity is absent, but that its cause may in these cases differ from what has often been suggested.⁵

There has also been a series of “reviews” of the credibility of much of the evidence of behavioral findings, including those purporting to show large differences in people’s valuations of gains and losses, and of the efficacy of their use in policy design, or whatever. While likely not reflecting a random sampling of opinions on the matter, some tend strongly toward seeing much more modest gains from behavioral insights, or more faith that standard theory will continue to provide adequate guidance – or at least useful defenses against criticism. The flavor of such views is reflected in a couple of these reviews of the behavioral and policy literature. In one that looked at fairly general areas, including benefit–cost analysis, the judgment was, “Overall then, this review of advances in behavioral economics suggests that the two leading approaches for expanding revealed preference analysis to incorporate behavioral insights are not ready for use in policy analysis”

⁵ It may also not alter how the disparity might be treated in policy design and the like, as people’s valuations are still reflected in the numbers regardless of this being a major source of them.

(Smith & Moore, 2010). A still more recent review focusing more on energy issues, concludes, “We view the current state of the behavioral welfare economics literature as an important foundation for future research, but the existing theoretical work appears to be far from ready for use in practical policy analysis” (Gillingham & Palmer, 2014). Although, again, continued testing and scrutiny and replication are both useful and necessary, some may also feel that in light of the accumulated evidence such continuing overarching negative conclusions, such as these two, may also be a bit curious.

A few implications

To the extent that decisions are at all guided by explicit or even implicit estimates or impressions of the monetary assessment of changes in the well-being, or welfare, of individuals affected by implementation of a project, change in policy, or other such change, the current practice of commonly substituting the use of WTP measures when the WTA measure is more appropriate is, on present evidence, likely to seriously distort assessments and bias outcomes. Given the typical magnitude of the disparity between valuations of the welfare changes associated with gains and losses and reductions of losses, the social costs of this current practice are likely to result more from serious understatements of the value of imposing, or failing to mitigate, losses – from, for example, allocations or interventions that do too little to control pollution, prevent and treat injuries and diseases, and reduce errors and crimes.

An oft used argument for the knowingly inappropriate use of the WTP measure to assess the importance of losses and reductions or elimination of losses, is that such a measurement will provide a conservative estimate of the value at issue – a rationalization presumably based on the belief that only an income or wealth effect could be the source of any difference, and that the resulting disparity would therefore be a relatively modest one. This use of the WTP measure can be justified in cases in which a value meeting some minimal threshold level⁶ is sufficient to determine a choice or decision, and this would be the case regardless of the magnitude of any disparity. However, given the magnitudes of the disparities commonly observed – those reviewed by Horowitz and McConnell (2002) had a mean WTA to WTP ratio of 6.7 (median of 2.6) and those by Tuncel and Hammitt (2014) a geometric mean of 3.28 – and the infrequency of cases where such threshold

⁶ This might be the case, for example, when an initiative is viewed favorably if the output or outcome at least exceeds the costs of bringing it about.

values can be so decisive, a descriptive term similar to *wrong* might seem to generally provide a more accurate characterization than *conservative*.

Aside from issues of misallocations that likely result from the use of inappropriate measures to assess changes, the disparity between the measures can also undermine various other kinds of analyses. For example, estimates of price elasticities may well differ depending on whether or not any account is taken of the real possibility of differing reactions to price increases and price decreases. In a surprisingly rare instance in which this was done, Putler (1992) found that when he separated the price/quantity observations used to make such estimates into those following an increase and those following a decrease in the retail price of eggs, the result was an estimate of -1.10 for price increases and -0.45 for price decreases. The far greater sensitivity of purchase decisions to price increases than to decreases is presumably due to the increase being taken as a more important loss relative to the less important gain accruing from a price decrease. While marketing studies have noted this difference in terms of a factor to take into account in store promotions and the like (Somervuori & Ravaja, 2013), there appears to be, again, surprisingly little attention to this in a world in which estimates of price elasticities are both common and important.

In the well-known study of controlling unwanted behavior by implementing a price, Gneezy and Rustichini (2000) recorded how imposing a fine for parents coming late to pick up their small children from a day-care center failed to accomplish the anticipated improvement. People instead took the sanction as a price to be allowed to come late, and this, together with the fine also fairly well destroying the moral sanction parents had felt when they did not arrive in time, resulted in more, not less, lateness. What seems less appreciated is that the level of the fine was set at a relatively modest level, which parents apparently took to be a clear signal that their tardiness was actually not all that serious a problem, and this too prompted them to act accordingly. That there was no obvious relationship of the level of fine to the costs that were imposed appears to be a common failing of nearly all implementations of Pigouvian taxes, and likely contributes much to their unpopularity and lack of effectiveness. A closer and more transparent tie of sanction and the value of the imposition resulting from the activity may reduce both their unpopularity and their ineffectiveness. But, here again, it seems important that the estimates of the costs accurately (and transparently) reflect the losses that the actions impose. Here too, the WTA measure seems called for.

In all, the benefits from a better sorting out of the choice of measure issue appear highly likely to well exceed the costs.

References

- Fryer, Roland G., Levitt, Steva D., List, John & Sadoff, Sally (2012). *Enhancing the Efficacy of Teacher Incentives through Loss Aversion: a Field Experiment*. Working Paper 18237. Cambridge, USA: National Bureau of Economic Research.
- Gillingham, Kenneth & Palmer, Karen (2014). Bridging the Energy Efficiency Gap: Policy Insights from Economic Theory and Empirical Evidence. *Review of Environmental Economics and Policy*, 8(1), 18–38.
- Gneezy, Uri & Rustichini, Aldo (2000). Pay Enough, or Don't Pay at All. *The Quarterly Journal of Economics*, 115, 791–810.
- Hammack, J. & Brown, Gardner (1974). *Waterfowl and Wetlands: Toward Bio-economic Analysis*. Baltimore, USA: Johns Hopkins Press.
- Horowitz, John & McConnell, Kenneth. (2002). A Review of WTA/WTP Studies. *Journal of Environmental Economics and Management*, 44, 426–447.
- Henderson, A. M. (1941). Consumer's Surplus and the Compensation Variation. *Review of Economic Studies*, 8, 117.
- Kahneman, Daniel & Tversky, Amos (1979). Prospect Theory: An Analysis of Decisions Under Risk. *Econometrica*, 47, 203–235.
- Knetsch, Jack L., Riyanto, Yohanes E. & Zong, Jichuan (2012). Gain and Loss Domains and the Choice of Welfare Measure of Positive and Negative Changes. *Journal of Benefit-Cost Analysis*, 3(4), 1–18.
- Knetsch, Jack L. & Wong, Wei-Kang (2009). The Endowment Effect and the Reference State: Evidence and Manipulations. *Journal of Economic Behavior and Organization*, 71, 407–413.
- Koszegi, Botond & Rabin, Matthew (2006). A Model of Reference-Dependent Preferences. *The Quarterly Journal of Economics*, 121, 1133–1165.
- Pope, Devin G. & Schweitzer, Maurice E. (2011). Is Tiger Woods Loss Averse? Persistent Bias in the Face of Experience, Competition, and High Stakes. *The American Economic Review*, 97, 129–157.
- Putler, Daniel S. (1992). Incorporating Reference Price Effects into a Theory of Consumer Choice. *Marketing Science*, 11, 287–309.
- Smith, V. Kerry & Moore, Eric M. (2010). Behavioral Economics and Benefit Cost Analysis. *Environmental and Resource Economics*, 46, 217–234.
- Somervuori, Outi & Ravaja, Niklas (2013). Purchase Behavior and Psychophysiological Responses to Different Price Levels. *Psychology and Marketing*, 30(6), 479–489.
- Thaler, Richard H. & Benartzi, S. (2004). Save More Tomorrow: Using Behavioral Economics to Increase Employee Saving. *Journal of Political Economy*, 112, S164–S182.
- Tuncel, T. & Hammitt, James K. (2014). A New Meta-Analysis on the WTP/WTA Disparity. *Journal of Environmental Economics and Management*, 68, 175–187.
- Weaver, Ray & Frederick, Shane (2012). A Reference Price Theory of the Endowment Effect. *Journal of Marketing Research*, 49, 696–707.
- Zerbe, Richard O. (2001). *Economic Efficiency in Law and Economics*. Cheltenham, UK: Edward Elgar Publishing.