Reversal Costs

Tomasz Mielniczuk* and Daniel Pi†

Abstract

We introduce a hitherto neglected but important concept which we call “reversal costs.” Reversal costs are the cost of undoing a decision. They figure in the background of all decision-making, and their analysis can explain many common phenomena. We present a simple formal model and illustrate the concept using several examples.

JEL Codes: D11
Keywords: reversal costs, rational choice theory

In this paper, we introduce a simple, general, and ubiquitous concept which we call “reversal costs.” Reversal costs are the losses associated with undoing a decision. Obvious examples include cancellation fees, damages for breach of contract, the political and procedural burdens associated with the repeal of laws, and the pain associated with the removal of tattoos.

*Erasmus University Rotterdam, Rotterdam Institute for Law and Economics, University of Bologna, Department of Economics; and University of Hamburg, Institute of Law and Economics. E-mail: tomasz.mielniczuk@edle-phd.eu.
†University of Bologna, Department of Economics; University of Hamburg, Institute of Law and Economics; and Erasmus University Rotterdam, Rotterdam Institute for Law and Economics. E-mail: daniel.pi@edle-phd.eu. We thank Francesco Parisi, Luigi Franzoni, Michael Faure, Alan Miller, and Roy Partain for their helpful and encouraging comments. Tomasz Mielniczuk also extends his thanks to Roger Van Den Bergh.
Before proceeding, it behooves us to make very clear at the outset what we believe the novelty of our contribution to be. To be clear: we do not seek to explicate a specific phenomenon. Rather, we seek to describe a general principle, for which there exist many known specific examples.

In some sciences, “progress” is identified with the discovery of laws and principles of ever greater generality. Physics is perhaps the best example of such a field, although the motivation for generality may also be observed in parts of chemistry, genetics, linguistics, and of course economics. What sort of contribution is generalization? To illustrate, it is well understood that $H_2O$ is solid at temperatures below $0^\circ$ C and liquid at temperatures above; and that $C_2H_6O$ is solid at temperatures below $-114^\circ$ C and liquid at temperatures above; and that $Hg$ is solid at temperatures below $-39^\circ$ C and liquid at temperatures above. To demonstrate that all three properties are specific instances of a more general property (i.e., freezing and melting, or more generally still, phase transitions) should not be regarded as redundant with those prior observations. The bare derivation of a general property does not, of course, describe any new specific phenomenon. However, it uncovers instead a more fundamental principle, of which the specific instances partake. Once understood as such, it may be applied to an infinitude of further specific cases. It may of course also be edifying in itself.

These preliminary remarks are not meant to raise expectations for a grand theory. They are meant to indicate the direction—not the magnitude—of our research. We consider reversal costs to be a simple concept. Indeed, we regard its simplicity to be a large part of its appeal. Nevertheless, given the potential for misinterpretation, we have judged it prudent to forestall potential confusions about our motivation and purpose.
Our discussion is structured in three sections. In Section ??, we present an abstract description of reversal costs. In Section ??, we illustrate the role of reversal costs using several hypothetical examples. We conclude in Section ?? with some brief remarks about the significance and implications of reversal costs for theoretical economics.

1 Model and Taxonomy

Let us $s_0, s_1, \ldots$ represent possible states of the world, and let $S$ be the set of all possible states of the world $s_0, s_1, \ldots \in S$. For any given state, $s \in S$, let $U_i(s)$ be the utility of an individual, $i$, in that state of the world. Now, suppose that some individual, $i$, is faced with a decision. He may select among several choices $c_0, c_1, \ldots$ in a choice set $C = \{c_0, c_1, \ldots\}$.

For any given choice, $c_x$, let us define the “naïve” expected utility of selecting that choice to be the sum of the utilities of possible outcomes weighted by the probability that they will occur. In other words, if $S_x \subset S$ is the set of possible states of the world that result from choosing $c_x$, then:\footnote{Define $S_x$ as a function of $c_x$.}

$$ NEU_i(c_x) = \sum_{s_j \in S_x} p(s_j, c_x) U_i(s_j). $$

We call this expected utility function “naïve,” because it ignores the possibility of reversion.

We say that an individual “reverses” his decision when he undoes the effect of his choice. This means that he either (1) effects a return to the position he was in
prior to his choice, or (2) effects a change in his position so that it is as if he had
selected a different choice than the one he did chose.

We will now define an expected utility function sensitive to the possibility of
reversion, from any chosen state \( s_y \) to an alternative state \( s_z \), at cost \( r_y \). Let the
reversal cost \( r_y \) be a function \( r_y = g(s_y, s_z) \in \mathbb{R}^+ \). And define \( s_z \) as the next
best known alternative to whatever \( i \) chooses. By default, we assume \( s_z \) is the
position that individual \( i \) was in prior to making the decision, but it may also
be some other state of affairs where the utility is known. We take the ex ante
position of individual \( i \) to be representative, because it will be a state of affairs
where individual \( i \) knows \( U_i(s) \), having experienced it prior to making his choice.

Next, we partition the set of possible outcomes, \( S_x \). Let \( S_x^+ \subset S_x \) be the
set of outcomes, for which a person would not reverse (because the payoffs of
the outcomes in this set are sufficient not to warrant reversal), and let \( S_x^- \subset S_x \)
be the set of outcomes, for which a person would reverse (i.e., for which either
the individual’s original position or the expected consequence of some alternative
choice is preferable to the obtained outcome). Formally:

\[
\forall s_j \in S_x^+(U_i(s_j) \geq U_i(s_z) - r_j)
\]

\[
\forall s_k \in S_x^-(U_i(s_k) < U_i(s_z) - r_k)
\]

With this definition, we may define the expected utility function (informed by the
possibility of reversion) to be:

\[
EU(c_x) = \sum_{s_j \in S_x^+} p(s_j, c_x)U_i(s_j) + \sum_{s_k \in S_x^-} p(s_k, c_x)(U_i(s_z) - r_k)
\]
What is the intuition here? The idea is that in some cases, there can exist a “lower bound” to the undesirable consequences of choices. In situations where the option to “undo” a decision exists, the worst case scenario is to be put back in the position one was in prior to making the choice, less the cost of exercising the option to “undo.”

In other words, the worst case consequence of any choice is to be down the cost of reversing the decision. This is a fully general claim (i.e., it applies to all decisions made by all individuals), because in cases where reversal is impossible, we can simply set the reversal cost to be infinite, and the model remains valid. Several results follow directly from these definitions.

**Theorem 1.1.** If $\forall s_j \in S_x(r_j \geq U_i(s_z) - U_i(s_j))$, then $EU(c_x) = NEU(c_x)$.

**Proof.** This follows directly from the definitions. If $r_j \geq U_i(s_z) - U_i(s_j)$ for all $s_j$, then $S^-_x = \emptyset$ (by Formula ??), and $S^+_x = S_x$ (by Formula ??). Therefore, $EU(c_x) = \sum_{s_j \in S^+_x} p(s_j, c_x) U_i(s_j) = \sum_{s_j \in S_x} p(s_j, c_x) U_i(s_j) = NEU_i(c_x).$  

**Theorem 1.2.** As the cost of reversal decreases, the expected value of a choice increases.

**Proof.** Again, the proof follows trivially from the definitions. Observe that in Formula ??, the reversal costs are negative terms. Thus, $EU(c_x)$ will tend to increase as $\sum r_k$ decreases.

**Corollary 1.2.1.** As the cost of reversal increases, the expected value of a choice decreases.
2 Examples

2.1 Simple Example

Let us begin with a simple case: Why is it that some retailers offer full refunds for goods (within some period), whereas others charge a restocking fee, and still others sell their products “as is”? Traditionally, economists have explained these differences as instances of signaling, or more extravagantly, the exploitation of the “endowment effect.” We think these explanations surely play a role, however we think that reversal costs are also a factor worth considering.

Consider a hypothetical retailer, Globotron, Inc., offering to sell widgets for $100. And consider an individual, Bob, who would be willing to pay at most $90 for a widget. Ceteris paribus, Bob will elect not to purchase widgets from Globotron.

Let us speculate further about Bob’s willingness to pay $90 for the widget. Suppose that Bob foresees two possible consequences of purchasing a widget. In scenario A, Bob derives great pleasure from the widget, and values it at $150, but in scenario B, he discovers it to be a disappointment, and values it at only $30. For the sake of simplicity with no loss of generality, suppose these are the only two possible outcomes. Say that Bob’s best estimate of probabilities is that there exists a 50% chance that scenario A will arise, and a 50% chance that scenario B will arise. This gives us the expected value of the widget for Bob being $90.

\[ 0.5 \times 150 + 0.5 \times 30 = 90. \]

---


Let us now suppose that Globotron offers the following deal: for any customers dissatisfied with their widgets after one month, Globotron will allow customers to return the widget for a full refund. This creates a “no-lose” situation for Bob. Suppose that Bob purchases the widget for $100. Now, if he finds that he loves the widget (scenario A), then Bob will keep it and enjoy the consumer surplus of $50. If, on the other hand, he finds it a letdown, then Bob will simply return the widget, for which he will receive the $100 he paid for it. Thus, his expected payoff is now $0.5 \times 150 + 0.5 \times 100 = $125. And now, because $125 > 100$, Bob buys the widget.

Of course, these are the extremum cases (offering no refund, and offering a full refund). There exists a spectrum of possible reversal costs that the seller could elect to impose. For instance, Globotron could charge a restocking fee. By Theorem 2, we know that for any restocking fee up to $50, Bob will buy the widget ($0.5 \times 150 + 0.5 \times (100 - 50) = $100$). Of course, the indifference point for other consumers may of course differ, and determination of what the optimal restocking fee should be will be a function of the demand curve (which itself is a function of consumer surplus, which is a function of reversal costs).

### 2.2 Asymmetrical Reversal Costs

Reversal costs become more interesting in cases where different choices are associated with reversal costs of differing magnitudes. For example, imagine a man, Archie, who is considering two possible romantic interests: Betty and Veronica. Archie must choose to pursue either Betty or Veronica—he cannot attempt to date

---

both at the same time. However, he is uncertain who would be a better mate.

Let us add the conceit that Veronica would not tolerate being Archie’s second choice, whereas Betty is indifferent as to Archie’s prior romantic history. Thus, the reversal cost for choosing Veronica is zero, whereas the reversal cost for choosing Betty is infinite.

Let us further suppose that Archie’s naïve expected utility from dating Betty is greater than his naïve expected utility from dating Veronica. He believes that he will probably be happier with Betty. Specifically, let us say that \( NEU(b) = 10 \) and \( NEU(v) = 7 \). But let us suppose further still that Veronica is temperamental and unpredictable, whereas Betty is a “known quantity.” Archie knows with 100% certainty that he would feel \( U(b) = 10 \) with Betty. However, there is a 20% chance that he would feel \( U(v) = 11 \), and an 80% chance that he would feel \( U(v) = 6 \) with Veronica, which gives the \( NEU(v) = 0.2 \times 11 + 0.8 \times 6 = 7 \).

However, when we figure in reversal costs, the arithmetic changes. If Archie chooses Betty, then because the reversal costs are infinite, he effectively forfeits the option to pursue Veronica later, and thus he will always enjoy \( EU(b) = 10 \). But if he chooses Veronica, then he may either stay with her if the relationship works out \( U(v) = 11 \), or because the cost of reversing when he chooses Veronica is zero, switch to Betty in case things don’t work out \( U(b) = 10 \). Thus, the payoff of choosing Veronica is \( EU(v) = 0.2 \times 11 + 0.8 \times 10 = 10.2 \). When reversal costs are considered, Archie should choose Veronica.
2.3 Unexpected Incentives

We provide one final example. Imagine the policymaker whose goal is to reduce the rate of tax fraud. The practical limitations of enforcement make it difficult or impossible to improve the detection rate, which is naturally low. However, high penalties ensure a reasonable measure of deterrence.\(^5\) Despite setting an efficient level of deterrence, given the heterogeneity of the population, there remain some people for whom the incentives are insufficient to deter.

Now, let us suppose that if a person misreports his income this year, then he will likely be forced to misreport his income next year in order to cover up inconsistencies that may send red flags to the tax authority. Suppose also that misreporting next year may further force him to misreport the following year, and therefore the year after that, and the year after that. Thus, in order to avoid detection, the one-time tax cheat is bound to remain a tax cheat indefinitely.

The policymaker knows that some tax cheats will subsequently come to regret their decision to misreport. These people, if given a “do over,” would report their income honestly. However, having already gone down the path of misreporting, they are trapped by the threat of detection if they do not continue misreporting. If they try to “go straight,” then their earlier deceptions may be revealed, and they will be penalized. Thus, they are perversely forced to perpetuate the fraud by the very system designed to prevent the fraud.

Recognizing this conundrum, the policymaker seeks to provide such “repenters” a way out. The most obvious solution would be to decrease the severity of sanction. However, this has the effect of reducing general deterrence, and increasing the total

The number of tax cheats. The policymaker therefore devises a more subtle solution: he offers tax cheats the option to “come clean,” by stepping forward to confess their fraud, for which they will simply pay their back taxes and a small fine. The underlying idea is simple. The tax enforcement situation divides citizens into two subsets: law-abiders and lawbreakers. Ideally, we want all citizens to be abiders. However, some will inevitably end up in the lawbreaker class. Moreover, individuals in the lawbreaker class will tend to find themselves stuck there, unable to move into the abider class, even if they should want to mend their ways. What the “come clean” initiative would do—the policymaker supposes—is make more permeable the barrier between lawbreaker and law-abider. Thus, anyone “stuck” in the lawbreaker category would have a less costly way of extricating himself from that class.

This seems sensible, and yet we hasten to point out a potentially fatal countervailing effect. The policymaker’s initiative reduces the reversal cost attached to tax fraud. Theorem ?? tells us that this increases the expected utility of committing tax fraud. Thus, by making it easier for some individuals who are “trapped” in the class of lawbreakers to get out, the policy also makes it more attractive for individuals to enter the class of lawbreakers in the first place. Though the policy does succeed in making the boundary between lawbreaker and law-abider more permeable, that permeability facilitates movement in both directions.

Consequently, we arrive at the surprising result that by making it easier for lawbreakers to rejoin society as law-abiders, the policymaker creates incentives for law-abiders to become lawbreakers. This does not however imply that policymakers should not attempt remedies of this form, but rather that the effect of such “coming clean” programs should be optimized with respect to the countervailing
effect described in Theorem ??.

By varying the fine attached to the “coming clean” alternative, policymakers may be able to find an optimal point between lawbreaking and law-abiding incentives.

It is also worth noting that although we have framed our example in the context of tax fraud, the same principles could just as easily apply to other areas such as gang membership in the criminal realm, terrorist activity in the national security realm, and price-fixing cartels in the antitrust realm.

3 Further Examples

The illustrative examples we have provided represent but a small fraction of situations where a reversal cost analysis might prove edifying. For further examples, consider the limited liability companies, which are devised to allow individuals to fail inexpensively, reducing the cost of “starting over” to encourage entrepreneurs to strike out on risky enterprises. Or consider the creation of “brand ecosystems,” which integrate services, increasing the cost for a consumer moving to a competitor’s products. Or consider witness protection programs, which perversely lower the cost of joining a criminal gang by offering a potential exit option if things go badly. Or consider the cost of divorce, which decreases couples’ incentives to marry by making it more difficult to back out later. Or consider provisions that make a proposed law easier to repeal (for example, by requiring active renewal after a given period of time), which decreases incentives for legislators to oppose it.

Reversal costs are easily modeled and deployable in a wide variety of theoretical contexts. They are also ubiquitous, present in the shadow of all decision-making.
Moreover, their effect on incentives (especially Theorem ??) will in most cases be easily empirically testable. For these reasons, we hope the concept may prove to be of some use.

References


