A comparative approach to protective behavior

Jonathan Baron*

Andrea D. Gurmankin

Howard Kunreuther

*This paper was never published. It is available only at http://www.sas.upenn.edu/~baron.

Abstract

We examined the determinants of protective behaviors (PBs) against health and property risks. Subjects answered questions about how much they engaged in each of several PBs (DoIt), the badness and probability of the event that each PB prevents, and the effectiveness and difficulty of the PB. We compared the expected utility for each PB from these judgments with the ratings of DoIt. Some of the differences may arise from a perceived budget constraint on PBs. This explanation is supported by the result of a framing manipulation: DoIt ratings were higher for a given PB presented as a member of a set of 3 PBs than a set of 7. Comparing preferences for adopting different PBs within an individual permits the discovery of inconsistencies. The paper concludes by suggesting interventions to increase people's adoption of those PBs perceived to have the highest cost-effectiveness.

Key words: protective behavior, risk reduction, risk

Introduction

People can reduce risks by undertaking protective behaviors (PBs), such as washing their hands before eating, wearing a seat belt, performing health screening tests, avoiding second-hand smoke, limiting alcohol consumption, and testing for radon gas in their homes. Governments and other institutions often set up programs to try to encourage people to undertake such PBs.

Considerable research has focused on what factors lead people to adopt PBs and what interventions are likely to increase the adoption of PBs whose benefits exceed costs for the population of interest (Cho, Keller, & Cooper, 1999; Kirscht 1988; Weinstein 2000; Weinstein, Sandman & Roberts 1990; Wurtele 1988). Most of these PBs take little time, cost little, and protect against serious — if sometimes improbable — bad events. In this sense, each of these behaviors is probably cost-effective, if it is the only one under consideration.

On one level, PBs are "rational" if they are undertaken whenever they are cost-effective. But the number of possible PBs is large, and it increases as new technology develops. As a person adopts more and more behaviors, the marginal cost of adding another one becomes greater, in terms of time, effort, and money. In the extreme, when one is spending 90% of one's waking hours taking precautions — usually a sign of psychopathology — there is little time left for anything else. As Hoehn and Randall (1989) contend, "Too many proposals pass the benefit-cost test."

Our thesis is that a given PB may be rational when considered in isolation but not when considered in the context of other PBs. This fact has two implications for the study of protective behavior. First, people might think of each PB in the context of competing PBs. They might have "PB budgets," so that they become less concerned about a given PB as a function of the number of PBs that they consider. Such an effect may be subject to framing. That is, people may be less inclined to do any particular PB when they consider a large number at once than when they consider a smaller number. Thus, the cost-effectiveness of a given PB considered in isolation is only part of the picture in determining whether people do or should undertake the PB.

In the experiment that we report here, we ask whether expressed willingness to undertake a PB is affected by the number of PBs under consideration. We manipulate the number — either 3 or 7 PBs — experimentally. We hypothesize that people will express less willingness to undertake a PB when they are considering more PBs at the same time. This effect is designed to simulate what normally goes on when people think about PBs.

Second, we need to evaluate rationality from the perspective of portfolio decision making. Analysis of a single PB in isolation, the focus of most research, might show that it is cost-effective, yet, in the context of other PBs, its cost might be high enough to exceed its expected benefit. We cannot know all the PBs that people consider, so it is difficult to assess the cost-effectiveness of each in this broader context. Instead, we should look for what might be called comparative cost-effectiveness. That is, it is irrational to perform a PB with a higher expected utility than some other PB that is not undertaken.

The idea of asking for judgments in the context of several other judgments of the same type has been applied to probability judgments. For example, Windschitl (2002) argued, on the basis of his own data and those of others, that probability judgments of a single risk, such as a smoker dying of lung cancer, are inflated and can be reduced when several other causes of death are listed. We extend this approach to a more complete set of judgments about protective behavior. Our approach is also similar to that of the theory of reasoned action (Ajzen, 1996) and such related approaches as the health belief model (Janz & Becker, 1984; again, see Cho, Keller, & Cooper, 1999, for an overview). These models are designed to predict behavior from several measures, such as intention, attitudes, social norms, motivation, efficacy, as well as (usually) probability and utility of relevant outcomes. Our approach differs because it assumes that the rational determinants are those that define the expected utility of behavior.

We illustrate the comparative approach with a study in which subjects evaluated the cost-effectiveness of a set of PBs. For each PB, we asked about the probability and disutility of the bad event, the effectiveness of the PB in reducing the risk, and its difficulty (in terms of money, time, unpleasantness, and self-control). We asked subjects to make judgments of these properties. Using these judgments, we can, in principle, estimate each subject's expected-utility of each PB and correlate this measure with the extent to which the subject adopts each PB.

The use of subjects' judgments is just one way to approach this question. Surely these judgments contain a great deal of error. We might have attempted to get more objective information about the probability of each event and the effectiveness of the behavior, but our interest was in people's perception of the risk and how this affects their adoption of PBs. Furthermore, objective data do not take into account the various factors that affect true individual differences in perceived probability and effectiveness, which may play a large role in what people should do (Nease & Owens, 1994). For example, people truly differ in their risk of heart disease, in the effectiveness of a low-fat diet in reducing that risk, and in their disutility for heart disease as opposed to other bad events. As we shall explain, the comparative approach allows us to analyze both individual and aggregate data and thus to estimate the importance of individual differences in probability and utility for predicting protective behavior, from the subjects' point of view.

Basic models of protective behavior

Consider the following example. You have to determine (implicitly) the effectiveness of buying organic vegetables reduce the chances of contracting cancer from chemical residues on vegetables. In determining whether it is worthwhile to adopt this PB you need to compare the utility of the reduced risk of contracting cancer from this action with the disutility of the additional time and monetary expenditure of buying organic vegetables. In deciding whether to buy organic vegetables you also need to consider a set of other PBs for reducing other risks to your health and safety, such as washing fruit before eating it, getting an examination of the colon for polyps and testing your house for radon gas.

You can evaluate the decision as to which PBs to invest in for each different risk, and the frequency with which they should be undertaken, using a simplified expected utility model. The expected utility is the benefit of the PB minus its cost. The benefit is based on the answer to three questions: the probability of the bad event in question without the PB (Prob); the disutility of that event (Bad, relative to a baseline of no bad event); and the effectiveness of the PB (Effective), where 100% effective means reducing the probability of the bad event to 0. The EU of the PB is the difference between the EU of not doing the PB and the EU of doing it. Without the PB the EU is $\text{Prob} \cdot \text{Bad}$. With the PB, the EU is $\text{Prob}(1 - \text{Effective}) \cdot \text{Bad}$. The difference is thus $\text{Effective} \cdot \text{Prob} \cdot \text{Bad}$, and this is the expected benefit of the PB.[1]

The cost of the PB is the difficulty of performing it. We ask subjects to compare this difficulty to the disutility of the bad event, where 100% difficulty means "just as bad as" the bad event in question.[2] Thus, without performing the PB, the EU of the cost is 0, and

[1] Here, and throughout, we use terms like Prob and Bad to refer to variables that we actually measure or compute from our data.

[2] A rating of 100 is implausible for most PBs, which are nowhere near as odious as the events they prevent.

with the PB the EU is Difficult · Bad, where "Difficult" is the response to the question about difficulty. Hence, the complete definition of EU is:

$$EU = \text{Effective} \cdot \text{Prob} \cdot \text{Bad} \; - \text{Difficult} \cdot \text{Bad}$$

We also asked about the extent to which each subject reports actually doing the behavior, a variable we label DoIt. We can then ask about the relation between DoIt and EU. Ideally, a person should always do a PB when EU is positive and never do it when EU is negative.

Experiment

We consider the following questions:

1. Are decisions about action (DoIt) affected by the number of actions under consideration?

2. Do people differ in the rationality of their PBs, according to their own judgments of DoIt, Prob, Bad, Effective, and Difficult?

3. Do people appear more rational when allowance is made for individual differences in the threshold for action?

The experiment used 42 PBs, in 6 sets of 7. Each set of PBs focused on a different issue, such as avoiding food risks, house safety, etc. Half the subjects saw all seven PBs in

---

Still, 13% of the responses were 100. Subjects varied enormously in how they used the scale. 13% of the subjects had mean badness responses of over 50 (out of 100), 16% had mean responses less than 10. It is plausible that some of the very high responses resulted from subjects starting too high and then trying to be consistent with their earlier responses, which they could not revise.

each set for sets 1, 3, and 5, and they saw 3 PBs from each of sets 2, 4, and 6, selected randomly from the seven PBs for each subject. To counterbalance the assignment of set sizes to sets, the other half of the subjects saw 7 PBs for sets 2, 4, and 6, and they saw 3 PBs each for sets 1, 3, and 5. The six sets were presented in the same order to all subjects.

The experiment was conducted on the World Wide Web. See `http://www.psych.upenn.edu/~baron/qs.html` for details. People discover this site because it is linked from many other sites (including those that list experiments on the Internet and how to earn money on the Internet). Use of the Web for research has several advantages over the alternatives for this kind of research (usually students): the subjects are much more varied than those from other convenience samples; expenses connected with data entry and checking are reduced; and, because it is easy to check answers as the subject enters them, fewer responses need to be discarded because they are nonsensical (Baron & Siepmann, 2000). Moreover, the general quality of the data is at least as high as that of data from paper questionnaires, and, in general, substantive results do not differ from those of comparable methods (Birnbaum, 1999, 2000; McGraw et al., 2000). Because subjects are paid, it is possible to track individual identities and ensure that nobody completes the same study twice.

Method

Two-hundred forty-three subjects completed a questionnaire on the World Wide Web, for $5 each. They ranged in age from 19 to 69 (median 38); 71% were female; and 12% were students. (We discarded data from three other subjects whose high speeds of answering the questions were outliers and from two others who gave the same response to some questions for every PB.) The introduction to the questionnaire began:

Protective behavior

This is about decisions about what to do to protect yourself from harm.

Please answer from your own point of view, as if these decisions were completely up to you and nobody else cared about them.

Some of the situations may be ones that you do not experience. Try to imagine yourself in these situations.

[There were then some instructions about how to use the browser.]

The list of risks was as follows. Each table contains a set of items related to a particular activity, with the activity in the top line. The next line shows the time horizon used for probability judgment (Prob) and the anchor used for the badness judgment (Bad), that is, the outcome to which each bad outcome was compared. Each of the remaining lines shows a PB, followed by a bad event that the behavior protects against.

1. preparations for two-week car trip away from a house	
Time: during the trip	*Anchor:* total destruction of your house
get your car inspected	car breakdown
unplug unnecessary appliances (TV, etc.)	fire
stop delivery of the newspaper	break in
check to see that all windows are locked	break in
turn off the water supply	water damage
put on lights or light timers to make it appear that someone is home	break in
make arrangements to have plants cared for	death of plants

2. periodic medical and dental checkups	
Time: in your lifetime	*Anchor:* your life expectancy cut in half
have annual blood tests for fats and cholesterol	heart disease
do monthly self-examination of breasts or testicles	cancer
do monthly self-examination for moles	skin cancer
have an annual flu shot	flu
floss your teeth after every meal	gum decay
have twice-a-year teeth cleaning and dental exam	cavities
have an eye examination every two years	serious eye problems

3. food	
Time: in your lifetime	*Anchor:* death within the next year
wash hands before eating	infection from food
eat only organic vegetables	cancer from pesticides
avoid genetically modified food	cancer from food
avoid using un-washed dishes or utensils	infection from food
avoid unclean restaurants	food poisoning
avoid buying food past the 'sell by' date	infection from food
avoid food touched by other people	infection from food

4. house safety	
Time: in your lifetime	*Anchor:* total destruction of your house
test your house for radon gas	lung cancer from radon
keep an up-to-date fire extinguisher	large fires
keep a burgular alarm	theft
check your roof once a year	leaks
use a dead-bolt lock	break in
use shatter-proof glass in first-floor windows	break in
keep smoke detectors and change their batteries yearly	large fires

5. personal computer security	
Time: in a year	*Anchor:* total loss of your computer and everything on it
use a password for your personal computer	cracking
check weekly for security alerts for programs that you use	cracking
use virus protection software	computer damage from viruses
change your internet password monthly	cracking
back up important files daily	loss of work from power outage or disk crash
use an un-interruptable power supply	loss of work from power outage
use a surge protector on the power cord	damage to computer from power surge

6. financial security	
Time: in your lifetime	*Anchor:* loss of all your savings
leave your credit card at home when you go out	theft of credit card
pay your bills at least a week before they are due	financial penalties for lateness
check your bank statements	losses from bank errors
avoid using a credit card on the internet	theft of credit card number
keep important documents in a fireproof safe or bank vault	loss in a fire
refuse to lend money to friends	inability to collect
avoid writing down your bank PIN and carrying it with you	ATM theft

To illustrate, here are the questions asked about the PBs at the beginning of each set, illustrated for a set of 3 PBs. The answers to these questions are called DoIt. These 3 PBs were presented together on one page. Then a series of pages asked several questions about each PB.

> Here are some things you might do concerning preparations for a two-week car trip away from your house. Indicate how often you would do each one if you were in this situation and the decision were entirely yours. ('Always' means 'whenever you have the opportunity').

unplug unnecessary appliances (TV, etc.), to protect against fire

Never 1/4 of the time Half the time 3/4 of the time Always

[printed on buttons]

get your car inspected, to protect against car breakdown

Never 1/4 of the time Half the time 3/4 of the time Always

check to see that all windows are locked, to protect against break in

Never 1/4 of the time Half the time 3/4 of the time Always[3]

The following illustrates the questions asked about each PB and the bad event it protects against, for a set with three PBs. Each page asked about a separate PB, but the set was repeated at the top of the page for each PB. The words in brackets indicate the variable names used in this article to refer to the subjects' responses:

Here are the protective behaviors concerning preparations for two-week car trip away from a house, again:

unplug unnecessary appliances (TV, etc.)
get your car inspected
check to see that all windows are locked

Now consider just: unplug unnecessary appliances (TV, etc.), to avoid fire.

[Bad] How bad is the typical example of 'fire' that you protect yourself against when you unplug unnecessary appliances (TV, etc.), where 100 is 'as bad as total destruction of your house' and 0 is 'not bad at all'? (Pick the closest.)

[3] In addition, we asked, "Have you actually been in this situation, or are you just imagining what you would do if you were in it?" The answer to this question did not affect any results, so we do not consider it further.

0 .02 .05 1 2 4 8 16 32 64 100 [printed on buttons]

[Prob] If you do not unplug unnecessary appliances (TV, etc.) at all, what is the percent probability that you will experience 'fire' in the trip?

0% .02% .05% 1% 2% 4% 8% 16% 32% 64% 100%

[Effective] How much is the probability of fire reduced when you unplug unnecessary appliances (TV, etc.)?

Not at all 1/4 Half way 3/4 Completely

[Difficult] How difficult is it for you to unplug unnecessary appliances (TV, etc.), on a scale where 0 is 'absolutely no difficulty' and 100 is 'so difficult that doing it on every occasion is just as bad as the typical case of fire that it would prevent'. (Pick the closest.)[4]

0 .02 .05 1 2 4 8 16 32 64 100

Results

We first tested the hypothesis that there is a subjective PB budget, so that DoIt is higher in sets of 3 than in sets of 7. Then we discuss the relationship between EU and DoIt. Finally, we consider individual differences, and the possibility that people appear more rational when PB budget constraints are taken into account.

[4]We asked two additional questions about worry: "How much would you worry about fire if you do not unplug unnecessary appliances (TV, etc.) at all?" and "How much would your worry decrease if you unplug unnecessary appliances (TV, etc.)?" The first question was badly worded, because it implies that the worry is about one's own negligence rather than about the bad event. We thus ignore these questions here, except to point out that the first item was very highly correlated with DoIt (.50 mean within-subject correlation, as compared to .34 for the second item).

Framing budgets: effects of set size

The experiment included a within-subject framing manipulation in which the number of PBs presented together in a set was either 3 or 7. As we explained, the assignment of PB sets to set sizes was counterbalanced, and within the 3 PB condition the PBs were selected randomly for each subject. To reduce the error associated with the random selection of items for the smaller set, we used normalized scores to analyze set-size effects. We normalized by subtracting from each rating (e.g., the DoIt response to a particular PB for a particular subject) the mean of that rating for the same PB calculated across both set sizes (3 and 7) and across all subjects. (The mean of the normalized ratings is thus zero.) After normalizing, we computed means for each set size for each measure for each subject, and tests were across subjects.

DoIt was higher for set size 3 than set size 7, as hypothesized. The difference was 3%, on the 0–1 DoIt scale, an effect that is small but statistically significant ($t_{40} = 3.10$ for the within-subject comparison, $p = .0021$). This difference was not the result of any differences in EU: the mean EU for set size 3 was 0.1% less than that for set size 7, a non-significant difference. EU was calculated for each item for each subject as $EU = \text{Effective} \cdot \text{Prob} \cdot \text{Bad} \; - \text{Difficult} \cdot \text{Bad}$, before being normalized.

Prediction of DoIt

We first ask whether the EU model makes sense as an approximate predictor of what people do. In particular, we ask whether all the components of EU are relevant to prediction of investment of time and money in specific PBs.

Exhibit 1 shows the mean ratings on each variable for each of the 42 PBs (abbreviated, but in the order listed above), with the scales transformed to go from 0 to 1, except for EU (expected utility). On the average, PBs were closely related to EU, with a correlation of

.82 across the 42 PBs ($p = .0000$).

— Insert Exhibit 1. —

The Bad ratings used different standards for different sets, and the Prob ratings used different time periods. But the time periods were chosen to be realistic.[5] For all cases, people should undertake the PB when its EU is positive, regardless of the scales.

To examine prediction of DoIt from the components of EU, we approached the data analysis in two ways, which we call common and idiosyncratic. The common approach involved averaging the data across all subjects and then computing the correlation across the 42 PBs, as in the just-reported correlation of .82 between DoIt and EU. This kind of analysis tells us what predictors are common to the subjects.

The idiosyncratic approach is concerned with predictors of each subject's DoIt response that were not explained by factors common to all subjects. For this analysis, we used the normalized scores described above. We computed correlations between DoIt and each predictor within each subject across the PBs, and we tested the mean of each correlation by a t-test across subjects. By subtracting from each score the mean of that score on the same PB, we removed any error variance associated with the PB itself. (Note that, if we computed within-subject correlations without normalizing and then asked whether the mean of these within-subject correlations differed from zero, we could find that some correlation is positive for most subjects — hence significantly positive when tested across subjects — because every subject is affected by some component of error that is common to all subjects. For example, one PB, for idiosyncratic reasons, might have high scores on two variables.)

All predictors played a role in predicting DoIt. For the common analysis, DoIt was

[5]Exploratory analyses suggested that we would not gain much additional predictive power from including dummy variables for the different standards used.

correlated positively with Prob ($r = .46$, $p = .0020$), Bad (.48, $p = .0014$), and Effective (.37, $p = .0173$) and negatively with Difficult ($-.78$, $p = .0000$). For the idiosyncratic analysis, DoIt was correlated positively with Prob (mean $r = .22$, average of within-subject correlations), Bad (.19), and Effective (.21) and negatively with Difficult ($-.23$). DoIt was also positively correlated with EU (normalized after being computed for each PB, mean $r = .18$). All of these means (based on the idiosyncratic analysis) were significantly different from zero ($t_{240} > 13$ across subjects); they are smaller than the common correlations because much of the error variance remains after normalizing. In sum, it seems that subjects were, on the average, using all relevant variables in deciding whether to undertake PBs.

Rationality

A primary purpose of the study was to illustrate the measurement of the rationality of protective behavior for each subject. In general, protective behavior is rational if DoIt is 1 (always do it) when EU is positive and 0 (never do it) when EU is negative.

Exhibit 2 presents an overview of the relation between EU and DoIt (converted to a 0–1 scale). Each point is one PB and one subject. (Cases in which EU is 0, the modal value, are excluded so that the graph would scale properly. Many other cases had EU close to 0.) The height of each bar thus represents the total number of observations at each value of EU for a given DoIt response (with 1 being certain to do it an 0 certain not to do it). It is apparent that the EUs for many PBs are negative, and subjects still express some willingness to undertake these PBs. But the largest group is in the upper right corner, where EU (of the PB considered in isolation) is positive and action is certain.

— Insert Exhibit 2. —

We measured rationality in two ways. One was to compute the Goodman-Kruskal γ

(gamma) correlation between EU and DoIt, within each subject. The γ statistic is the ratio of the number of concordant pairs of PBs minus the number of discordant pairs to the sum of these two numbers. A concordant pair (1,2) is one in which DoIt[1] is higher than DoIt[2] and EU[1] is higher than EU[2]. A discordant pair is one in which DoIt[1] is higher than DoIt[2] and EU[1] is lower than EU[2]. A discordant pair implies irrationality: EU[2] is higher, but the subject does PB1 and not PB2. (Ties in either measure are ignored, since they cannot indicate clear irrationality.) This measure is inherently comparative. The γ statistic is also useful for subjects who are nearly perfectly rational, in the sense that they would almost always undertake PBs with positive EU and almost never undertake PBs with negative EU. Because DoIt was essentially a step function of EU for these subjects, γ would be near 1 for them, but other measures of correlation, such as r, would not be so high. The γ statistic had a mean of .41 (median .46), with 86% of the subjects showing positive γ. [6]

The second way of assessing rationality was to compute the overall expected utility of the PBs that each subject would undertake, as inferred from DoIt. Perfect rationality would consist of performing every PB with positive EU 100% of the time and performing every PB with negative EU 0% of the time (assuming that EU is constant over occasions to do a PB). For each PB, we thus computed the product of DoIt (scaled from 0 to 1, according to the probability of undertaking the PB in question) and EU. The mean of

[6] We also predicted DoIt with a linear regression on Prob, Bad, Effective, and Difficult, for each subject. The coefficients for Prob, Bad, and Effective were constrained to be non-negative, and that of Difficult was non-positive. Although this model had four free parameters (the constant being irrelevant), the mean γ between DoIt and this model's prediction was .41. Thus, the multiplicative EU model fit as well as an additive model with more free parameters. Further exploratory analysis suggested that the responses scales were helpful in allowing subjects to express very low values of Prob and Bad. In previous studies, using more evenly spaced scales, we had found that many subjects were reluctant to give very low probabilities, and this fact led to very poor fits for the EU model.

these products was 0.038, with a range across subjects from −0.064 to 0.382 (on a scale on which 0 is no expected benefit and 1 is perfect protection against events with disutility 1 and probability 1). To assess rationality, we compared this mean to its maximum value assuming that the subject always undertakes a PB with positive utility, which was simply the mean of all the positive EUs. This maximum had a mean of 0.051, and a range from 0 to 0.437. The difference between these two means is our second measure of irrationality.[7] On the average, then subjects got 75% of the EU that they could get from their PBs (0.038/0.051).

We asked whether both rationality measures could be predicted from age and gender, in a regression. The γ measure was significantly higher in older subjects ($t_{236} = 2.76$, $p = .0062$) and non-significantly higher in females ($t = 1.06$; the overall regression was significant $p = .0137$). The second irrationality measure showed effects in the same direction (older subjects and females more rational), but they were not significant. Again, these results are merely illustrative of the kind of analysis that can be done and are difficult to interpret.

The value of individual-level analysis

Subjects differed consistently across PBs in DoIt, after controlling for both EU and item category (car trip, etc.) (analysis of variance: $F_{240,6983} = 3.71$, $p = .0000$; with subjects included, R^2 was .20, without subjects it was .10).

Such differences could lead to two different discrepancies from behavior that would be rational if we viewed PBs in isolation. First, people could fail to undertake PBs that seem

[7]We used the difference of the two means, rather than their ratio, because sometimes the maximum was zero: costs were always higher than benefits for some subjects, yet even these subjects said they would sometimes undertake PBs.

to be cost-effective for them. Such failure could result from a higher cost than is apparent from the analysis of the PB in isolation, such as the cost that results from a PB budget constraint. Note that “failure” could mean partial failure, so that DoIt is less than 1.0, even though EU is positive. Second, PBs could have negative EU and people could undertake them anyway. Thus, we examined cases in which EU was positive and those in which it was negative, separately. We defined TooLittle as 1−Doit (where DoIt ranged from 0 to 1), for those cases with positive EU, and TooMuch as Doit, for those cases with negative EU. Individual differences were significant (and of the same magnitude) for both measures. However, TooLittle declined with age ($r = -.18$, $p = .0066$) and TooMuch was not related to age ($r = -.07$); older people seem less inclined to omit PBs with positive EU.

To look at individual differences another way, we asked whether the second rationality measure, the one based on the overall EU of the subject’s choices, could be increased if we added, for each subject, a constant to DoIt. We constrained DoIt to lie between 0 and 1. (Thus, if the constant were .25, all DoIt responses of 0 would be replaced with .25, .25 would be replaced with .50, but 1.0 would not change.) The constant represents the extent to which subjects do too much (if the constant is negative) or too little (if it is positive). The constant itself could range from −1 to 1.

The mean constant was positive (0.12, $t_{240} = 3.15$, $p = .0019$), supporting our hypothesis that people have a PB budget constraint. When the constant was considered, the mean irrationality measure was reduced from .39 to .25. For subjects with positive constants (those who did too little by this measure), the irrationality measure changed from .33 to .22. For those with negative constants, from .49 to .28. In sum, subjects seem more rational if we consider the possibility that they are setting their action threshold in the wrong place, and, on the whole, subjects seem to do less than implied by EU based on their own judgments.

Discussion

We have demonstrated a new kind of methodology for studying protective behavior. It relies on comparisons among different behaviors that compete for time and effort. We have shown here that individuals differ systematically (across PBs) in the relationship between the EU of a PB, as determined from their own judgments, and their tendency to perform the PB.

These individual differences may arise for several reasons. First, people may differ in the way they respond to questions, including questions about Prob, Bad, Effective, Difficult and DoIt. Second, people may truly differ in their tendency to take protective action, holding EU constant.

We have suggested one reason for such differences. Namely, the true cost of a PB may increase as a function of the number of PBs already being taken when the opportunity arises. Our questions about ratings did not allow people to consider such factors. But we manipulated the number of PBs being considered together in a set. The size of the set did affect the DoIt ratings, while EU was unaffected. This result provides evidence for a budget constraint on PBs.

The budget constraint provides further support for the general strategy of assessing the rationality of PBs comparatively. An alternative approach is to attempt to measure the effects of the budget constraint directly and incorporate it into the Difficult measure. One way to do this might be to ask subjects to estimate the total time and money that they spend on PBs in general before asking about the cost of adding a new one.

Applications

The method we used here could be used as an outcome measure for interventions directed at inducing people to undertake a PB with true positive EU, such as getting a flu shot (the item with the lowest residual when we predicted DoIt from EU). When some people engage in this target PB less often than they engage in other PBs that are less cost-effective, a desirable intervention (such as an argument in a pamphlet) would increase the PB in question, and/or decrease the other PBs. This approach to evaluation of program outcomes would meet one objection to existing interventions, which is that they might increase the frequency of one behavior only at the expense of some other behavior that is more cost-effective for the person in question.

The comparative method could also be used as the basis of programs to help people assign priorities to PBs. This could be accomplished by showing people a list of the PBs they undertake and do not undertake, with their corresponding utilities. A statement of the form "If you take the trouble to do X, then you certainly ought to be doing Y (which has a higher EU)," or the converse, could be helpful in increasing consistency and in inducing people to adopt more cost-effective behavior.

The method could also be used as a research tool. Our sample of PBs was broad, but the same method could be used to study experimental manipulations of specific types of PBs, perhaps even pairs that are matched in some attributes and different in others.

A major limitation of our approach is that the prediction of rational behavior from individual judgments is at best an approximation. On the one hand, people really differ in risk factors, and they have some awareness of these differences. On the other hand, they may have false beliefs, and they may differ in their use of probability scales. In some cases, it might be better to use expert judgments rather than judgments made by those affected. These are matters for further research.

References

Ajzen, I. (1996). The social psychology of decision making. In T. Higgins & A. W. Kruglanski (Eds.), *Social psychology handbook of basic principles*, pp. 297–325. New York: Guilford Press.

Baron, J. (1992). The effect of normative beliefs on anticipated emotions. *Journal of Personality and Social Psychology, 63*, 320–330.

Baron, J. (1997). Biases in the quantitative measurement of values for public decisions. *Psychological Bulletin, 122*, 72–88.

Baron, J., Hershey, J.C., & Kunreuther, H. (2000). Determinants of priority for risk reduction: the role of worry. *Risk Analysis, 20*, 413–427.

Birnbaum, M. H. (1999). Testing critical properties of decision making on the Internet. *Psychological Science, 10*, 399–407.

Birnbaum, M. H. (Ed.) (2000). *Psychological Experiments on the Internet.* New York: Academic Press.

Diefenbach, M. A., Miller, S. M., & Daly, M. B. (1999). Specific worry about breast cancer predicts mammography use in women at risk for breast and ovarian cancer. *Health Psychology, 18*, 532–536.

Drotzz-Sjoberg B. M. & Sjoberg L. 1990. Risk perception and worries after the Chernobyl accident. *Journal of Environmental Psychology, 10*, 135-49.

Hoehn, J. P., & Randall, A. (1989). Too many proposals pass the benefit-cost test. *American Economic Review, 79*, 544–551.

Janz, N. K., Becker, M. H. (1984). The Health Belief Model: A decade later. *Health Education Quarterly, 11*, 1–47.

Kahneman, D., & Ritov, I. (1994). Determinants of stated willingness to pay for public goods: A study of the headline method. *Journal of Risk and Uncertainty, 9*, 5–38.

Kahneman, D., Ritov, I. & Schkade, D. (1999). Economic preferences or attitude expressions?: An analysis of dollar responses to public issues. *Journal of Risk and Uncertainty, 19*, 203–235.

Kirscht JP. 1988. The health belief model and predictions of health actions. In D.S. Gochman (Ed.), *Health Behavior: Emerging Research Perspectives* (pp. 27-41). New York: Plenum.

McGraw, K. O., Tew, M. D., Williams, J. E. (2000). The integrity of Web-delivered experiments: Can you trust the data? *Psychological Science, 11*, 502–506.

Mellers, B. A., Chang, S., Birnbaum, M. H., Ordóñez, L. D. (1992). Preferences, prices, and ratings in risky decision making. *Journal of Experimental Psychology: Human Perception and Performance, 18*, 347–361.

Myers JR, Henderson-King DH, & Henderson-King EI. 1997. Facing technological risks: the importance of individual differences. *Journal of Research in Personality, 31*, 1-20.

Nease, R. F., Jr., & Owens, D. K. (1994). A method for estimating the cost-effectiveness of incorporating patient preferences into practice guidelines. *Medical Decision Making, 14*, 382–92.

Sjoberg L. 1998. Worry and risk perception. *Risk Analysis, 18*, 85-93.

Weinstein N. (2000). Perceived probability, perceived severity, and health-protective behavior. *Health Psychology, 19*, 65–74.

Weinstein, N., Sandman, P., & Roberts, N. (1990). Determinants of self-protective behavior: home radon testing. *Journal of Applied Social Psychology, 20*, 783–801.

Windschitl, P. D. (2002). Judging the accuracy of a likelihood judgment: The case of smoking risk. *Journal of Behavioral Decision Making, 15*, 19–35.

Wurtele SK. 1988. Increasing women's calcium intake: the role of health beliefs, intentions and health value. *Journal of Applied Social Psychology, 18*, 627-639.

Exhibit 1: Mean ratings on the 42 PBs, scaled from 0 to 1 except for EU, which is inferred from other ratings.

	DoIt	EU	Prob	Bad	Effective	Difficult
Inspect car	0.63	−0.00	0.13	0.22	0.54	0.14
Unplug appliances	0.50	0.01	0.06	0.49	0.53	0.05
Stop newspaper	0.69	0.01	0.10	0.23	0.41	0.04
Windows locked	0.87	0.04	0.15	0.34	0.52	0.02
Shut water	0.29	0.01	0.06	0.24	0.70	0.11
Light timers	0.59	−0.01	0.09	0.27	0.46	0.08
Plant care	0.55	0.02	0.32	0.08	0.63	0.16
Blood test	0.52	0.03	0.27	0.38	0.47	0.12
Breast self-exam	0.52	0.02	0.17	0.42	0.39	0.06
Mole self-exams	0.39	0.00	0.10	0.28	0.44	0.08
Flu shot	0.33	0.02	0.42	0.13	0.47	0.16
Floss teeth	0.43	0.02	0.31	0.13	0.54	0.18
Dental exam	0.59	0.02	0.47	0.16	0.56	0.19
Eye exam	0.71	0.02	0.19	0.23	0.53	0.11
Wash hands	0.81	0.08	0.29	0.25	0.60	0.04
Organic veg.	0.25	−0.06	0.06	0.19	0.41	0.31
GM food	0.45	−0.02	0.08	0.19	0.37	0.21
Dishes washed	0.92	0.12	0.41	0.29	0.62	0.04
Clean restaurant	0.86	0.09	0.40	0.33	0.60	0.07
Sell-by date	0.90	0.08	0.32	0.27	0.57	0.03
Touched food	0.53	−0.02	0.20	0.19	0.46	0.20
Radon test	0.31	−0.04	0.07	0.27	0.50	0.21
Extinguisher	0.68	0.01	0.13	0.47	0.47	0.07
Burglar alarm	0.39	−0.02	0.15	0.25	0.54	0.23
Check roof	0.47	0.01	0.29	0.21	0.55	0.22
Dead-bolt lock	0.82	0.03	0.19	0.32	0.50	0.06
Shatter-proof glass	0.32	−0.04	0.09	0.18	0.39	0.27
Smoke detectors	0.86	0.08	0.20	0.57	0.61	0.04
Password for PC	0.59	0.03	0.14	0.26	0.49	0.06
Security alerts	0.49	0.01	0.12	0.25	0.50	0.12
Virus protection	0.86	0.15	0.42	0.53	0.64	0.06
Change password	0.25	−0.00	0.09	0.22	0.47	0.13
Backup files	0.48	0.01	0.21	0.31	0.66	0.18
UPS (power)	0.45	0.03	0.23	0.26	0.65	0.22
Surge protector	0.91	0.13	0.28	0.53	0.65	0.03
Credit card home	0.28	−0.00	0.13	0.23	0.53	0.24
Pay bills early	0.70	0.07	0.39	0.18	0.76	0.08
Bank statement	0.88	0.08	0.27	0.23	0.74	0.04
Internet payment	0.32	−0.00	0.17	0.30	0.57	0.25
Documents in safe	0.54	0.02	0.12	0.30	0.67	0.12
Don't lend	0.51	0.05	0.33	0.20	0.79	0.23
Memorize PIN	0.89	0.12	0.25	0.47	0.67	0.04

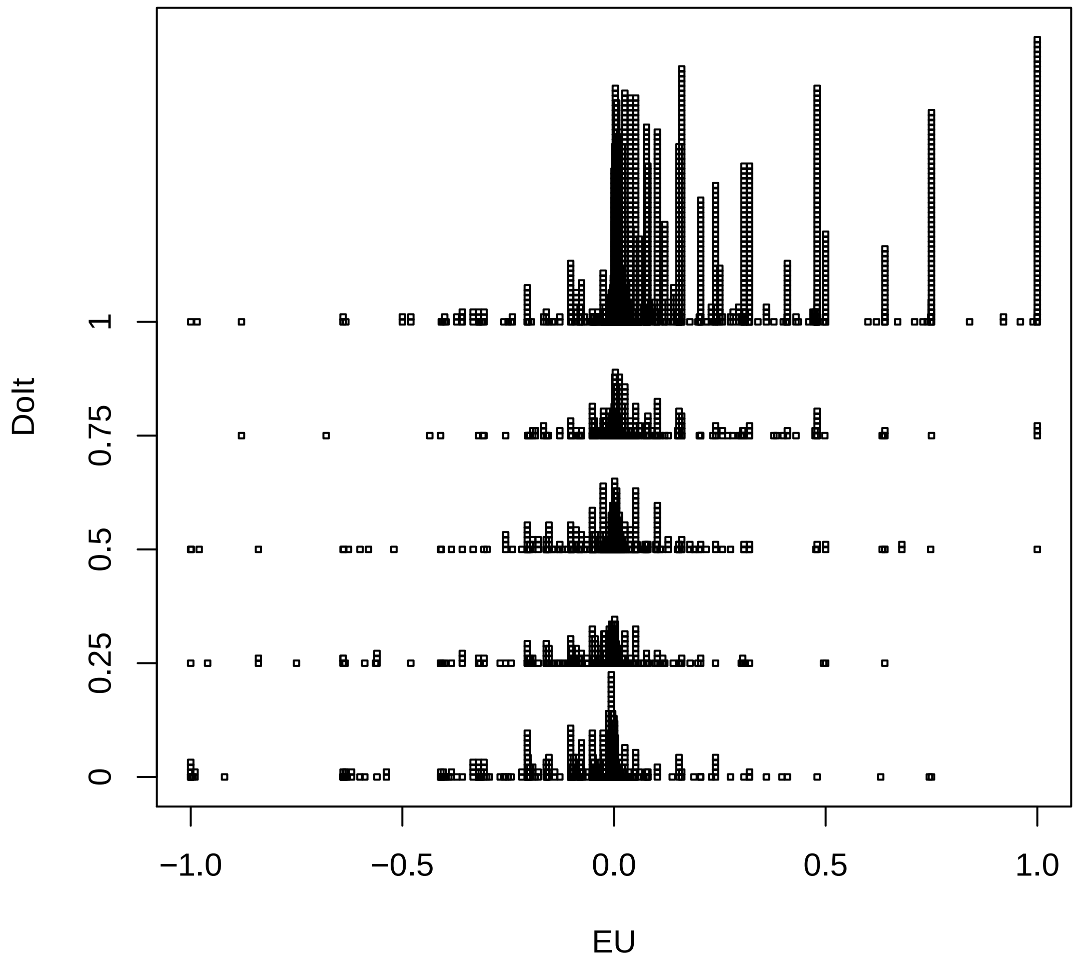

Exhibit 2. DoIt responses as a function of EU (excluding cases with EU of 0).

Biographical sketches

Jonathan Baron is Professor of Psychology, University of Pennsylvania. His interests are in the relations among decision making, moral judgment, and public policy.

Andrea Gurmankin is pursuing a Ph.D. in Psychology and a masters in medical ethics at the University of Pennsylvania. Her interests include medical ethics, medical decision making, and physician-patient interactions.

Howard Kunreuther is the Cecelia Yen Koo Professor of Decision Sciences and Business and Public Policy, and Co-Director of the Risk Management and Decision Processes Center at the Wharton School, University of Pennsylvania. His research is about decision processes, insurance, and low-probability risk.