De-biasing by calculation, elaboration, and reduction

Jonathan Baron
Department of Psychology
University of Pennsylvania

December 19, 2023

Abstract

We applied three methods of debiasing to four non-consequentialist biases. One method was to require calculation of the outcomes, so that people would attend to consequences. A second method was to expand a scenario with additional information, e.g., about other people than those described in the scenario. The third method was to provide a minimal description in terms of consequences alone. The four biases were omission bias, zero-harm bias, preference for ex-ante equality, and preference for group equality (even when these made consequences worse). The minimal method reduced biases in two experiments, and subjects tended to accept the minimal redescription as a fair summary. The expansion method reduced biases in one experiment but not in another. The calculation method was tested only for the equality bias, which it reduced.

A non-consequentialist bias is a judgment about what to do in some scenario that departs systematically from the judgment corresponding to the best outcome in that scenario. Arguably, when people act on such judgments, their actions yield non-optimal consequences (Baron, 1994, 1998). These biases are this of potential concern to those who are affected by them.

Here, I present experiments illustrating three different methods of debiasing allocation biases. De-biasing experiments are useful in part because they tell us something about the malleability of the biases. To the extent to which biases are malleable, we may not want to take them so seriously as guides to public policy. They also tell us about the nature of the biases.

The experiments concern four allocation biases. Omission bias is the judgment that harmful actions are worse than omissions that are even more harmful acts (Ritov & Baron, 1990; Spranca, Minsk, & Baron, 1991; see Ritov & Baron, 1999, for more citations).

What we shall call zero-bias is the bias toward complete solutions, i.e., zero risk or zero harm (Baron, Gowda, & Kunreuther, 1993; Ritov, Baron, & Hershey, 1993). For an extreme example, people would rather reduce a risk from 10% to 0 than from 30% to 10%. This bias is related to the proportionality bias, in which people judge risk reductions by proportion rather than absolute change (Fetherstonhaugh et al., 1997; Baron, 1997); reduction to 0 amounts to complete reduction, even if the risk is small.

The ex-ante bias is the finding that people want to equate ex-ante risk within a population even when the ex-post risk is worse (Ubel, DeKay, Baron, & Asch, 1996a). For example, many people would give a screening test to everyone if the test would prevent 1000 cancer cases rather than give a test to half (picked at random) that would prevent 1200 cancer cases.

The equality bias is the preference for equal treatment of two groups, even when unequal treatment would be better on the whole (Baron, 1993, 1995; Ubel, DeKay, Baron, & Asch, 1996b). For example, people would rather help 50% of each of two equal-sized groups than 80% of one group and 40% of the other.

Two general approaches to debiasing such biases have been used in previous studies. One is to extend the description, adding information that is plausible and consistent with the original scenario. For example, Ubel, Baron, and Asch (2001) reduced the equality bias by describing a third group that received no benefit at all. Subjects seemed to think that this third group removed the possibility of equal treatment for all groups, and, once this possibility was gone, it was better to aim for the greatest benefit overall.

Another study points to a way of expanding the omission bias so as to reduce it. Royzman and Baron (2001) found that omission bias was largely confined to cases in which the harm from action was direct. If the action was understood as releasing some other cause of harm, then harm caused by action was typically judged no worse than harm caused by omission (which must necessarily also involve some other cause). In the present study, we extend a case of action to specify that there is another cause.

The second method of debiasing is to redescribe the scenario in terms of consequences alone, referring simply to the number of people helped or harmed in each option. The implication of this method is that such factors as fairness and the act-omission distinction are irrelevant. If people can see things this way, they should be more willing to make judgments in terms of consequences alone.

1 Making calculations as a de-biasing method

This experiment examined the effect of training about optimal allocation on the equality bias. Some subjects got training in the first half of the experiment and others got training only in the second half. The training consisted of a brief paragraph explaining optimality, plus feedback on each trial about the maximization question, which had to be answered correctly in order to submit answers to the other questions.

The experiment asked about maximization and allocation, it asked what would be best overall. At issue here is whether people take their own preferred allocation to be different from what is best. If so, they knowingly favor non-optimal outcomes. Such an effect was found by Baron and Jurney (1993).

It also examined the effect of two other variables on allocations. The first was the relative size of the two affected groups. Allocation of a given amount of drug to a smaller group would cure a larger proportion of that group, other things being equal. If subjects were sensitive to proportion as well as to total numbers, they would allocate more to the smaller group. Proportionality effects have been found by Stone, Yates and Parker (1994), Baron (1997), and Fetherstonhaugh et al. (1997). The other variable was the seriousness of the disorder, blindness vs. blindness in one eye. Seriousness did not affect the results, so we do not discuss it further.

1.1 Method

Fifty-eight subjects completed a questionnaire on the World Wide Web, for \$3. Their ages ranged from 19 to 58 (median 30), 33% were male, and 22% were students. The questionnaire began:

> **Allocation of drugs for blindness**
>
> This questionnaire is about allocation of scarce medicine to different groups of patients. In these cases, imagine that a new drug has been invented, which will prevent blindness in older people when given in the early stages of their disorder. It does this by curing the condition that causes the blindness. The drug can be used both to prevent complete blindness or blindness in one eye.
>
> Because the drug is new, supplies are limited. The problems concern how the drug should be allocated to different groups, when there is not enough drug to treat everyone. Cost is not an

issue; the drug is left over from a research study and will decay if not used immediately.

In each case, you will see the cure rate for two groups of patients, A and B, with somewhat different conditions. The cure rate is the percent of patients who are cured if they get the drug. (No cures without it.)

The question is how to allocate the drug to the two groups. You are also asked how you would allocate the drug if you wanted to cure the most people, and how you would allocate it if you just want the best outcome on the whole. (The best outcome is however you define it. There are no right answers to that question.)

Bear in mind the following:

- The research has already been done. We know how much the drug will help each group. So there is absolutely no need to test the drug further. The only issue is proper allocation (until more of the drug becomes available).
- Nobody can get any information about how effective the drug will be for each patient, other than knowing what group the patient is in. The groups are the same in terms of age, sex, income, etc.
- There is not enough drug for either group. Even if you gave all the drug to one group, there would not be enough for all the patients in that group.
- There are 64 cases, three questions about each.

Each item was of the following form, with alternatives in brackets:

Condition: completely blind [blind in one eye]

Group	Size of group	Cure rate
A	1000 [3000]	60% [100%]
B	1000 [3000]	80%

What percent of the drug would you allocate to Group A? ☐

What percent of the drug allocated to Group A *would produce* *the highest overall cure rate*? $\boxed{}$

would produce the best outcome overall? $\boxed{}$

On the training trials, the order of the first two questions was reversed, so that the maximization question came first. On these trials, an incorrect answer to this question caused all the answers to be erased, and an alert to appear saying that the question had been answered incorrectly. The training trials were either the first 32 or the second 32. At the beginning of the training trials, the subject saw an alert saying:

> IMPORTANT INSTRUCTIONS:
> When a drug is more effective at treating one group, you can get the highest cure rate by allocating ALL of the drug to that group. Each person in that group is more likely to benefit from the drug than each person in the other group, so the chance of curing someone is highest if you give the next dose to someone in the group that is helped most. Make sure you understand this paragraph, because you will have to answer test questions about it each time.

On half of the trials, the numbers assigned to Groups A and B were reversed (so that A had a cure rate of 80%). The group sizes were varied orthogonally.

1.2 Results

Table 1 shows the mean percent responses for the two orders. The Allocate question asked how the subject would allocate the drug, with responses transformed to that the optimal allocation is always 100%. (The 1.5% of responses below 50 were raised to 50.)

Training increased the response to Best (on a scale where 100 is optimal) across both orders ($t_{57} = 2.64$, $p = 0.0106$, two tailed). Training did not, however, affect Allocate significantly ($t_{57} = 1.47$). Perhaps this failure is the result of training effects carrying over to the second half when training was given first. The measure of training effects in the first half included even the early trials, when training might not have had its effect. This explanation is supported by the fact that Allocate in the second half was significantly higher when training came first (95.15%) than when training was second (91.59%, $t_{56} = 2.35$, $p = .0225$). Of course, training also affected

Table 1: Experiment 2 mean responses in percent, as a function of the question, the experimental conditions (with numbers in percent), and the order of training vs. no training; the optimal allocation is always 100.

	Allocate	Best	Cure
No-train first	86.81	89.29	91.29
Train second	91.59	94.76	100.00
Train first	93.28	93.50	100.00
No-train second	95.15	93.97	94.03

the response to Cure because subjects were required to give the optimum in the training condition.

In the training condition, 47% of the subjects gave allocation responses less than 100% on at least some trials (even though they could not make a final response to the Allocate question until they had correctly answered the Cure question). In the no-training condition, however, answers to Allocate and Cure were not significantly different combining both orders ($t_{57} = 1.49$). Nor were Allocate and Best significantly different overall. In general, subjects thought that their Allocations were best overall.

Best was greater than Allocate when the subject's mean Cure was greater than the subject's mean Allocate response (88.24% vs. 79.27%,$t_{20} = 2.86$, $p = 0.0049$ one tailed, for the no-training condition; 87.82% vs. 83.25%, $t_{26} = 2.01$, $p = 0.0274$ one tailed, for the training condition, where the test involved only cases where Allocate was less than 100%). In other words, when subjects believe that their own Allocations to the more effective group are less than optimal, they also believe that the best overall allocation is higher than their own response. They thus knowingly recommend options that fail to bring about the best consequences.

All responses favored the larger group, on the average (Allocate, $t_{57} = -2.81$, $p = 0.0068$; Best, $t_{57} = -2.97$, $p = 0.0043$; Cure, $t_{57} = -1.91$, $p = 0.0612$). Most likely, this is the result of some subjects forgetting the instructions that a fixed amount of drug was available. As group size was counterbalanced, this should not affect the other results.

2 Minimalism and expansion: First study

Another pair of studies explored two other methods of de-biasing. The "minimal" method involved stripping away information and focusing on consequences alone, a minimal description. I tried this method on four different biases: omission bias, zero-harm bias, preference for ex-ante equality, and preference for group equality (even when these made consequences worse). A second method involves expansion: provision of additional information that might change the bias. In the first study I shall simply summarize the positive results and not dwell on what didn't work.

2.1 Method, first study of minimalism and expansion

Ninety-five subjects completed a questionnaire on the World Wide Web. Their ages ranged from 18 to 62 (median 38); 34% were male; 15% were students. The questionnaire began:

> **Health insurance**
>
> This questionnaire concerns decisions made by health insurance companies about which treatments to cover.
>
> On each screen, you will see some information about two treatments for serious conditions. The two treatments have the same cost, which is high. An insurer cannot afford to cover both treatments, so it chooses one.
>
> All conditions are chronic, making for a low quality of life and usually a shorter life too. Examples of such chronic conditions are severe arthritis, senility, emphysema, Parkinson's condition, and heart disease.
>
> The questions just talk about "conditions" without specifying which conditions. Imagine that the questions refer to serious conditions, that are all **equally** serious.
>
> Sometimes the treatment leads to a different condition as a side effect. This different condition is just as serious as what the treatment cures: no more, no less.
>
> You will also see information about an insurance company's decision about whether or not to cover the treatment.
>
> Each screen has two or three questions. Treat each screen as a separate case, as if this is all you know about it. There are 32 screens, each with two or three questions.

The 32 screens were presented in a different random order for each subject. They constituted a 4x4x2 design: type of bias (omission, zero, ex-ante, equality), type of de-biasing (control condition, minimal, expanded, minimal-test), and action of the company (two levels, choosing the option with the best consequences, the "better" option, or the other option). As an example, the omission bias case was [with comments in brackets]:

> Treatment A cures 50 people out of 100 who come in with condition X each week, and it leads to no other conditions.
>
> Treatment B cures 80 of the people with condition X, but it leads to condition Y (randomly) in 20 of the 100 patients. X and Y are equally serious.
>
> [added for expanded de-biasing]
>
> Treatment B does not cause condition Y directly. It reduces people's resistance to other causes of condition Y.
>
> [added for minimal de-biasing]
>
> In other words, treatment A leads to 50 people with condition X and nobody with any other condition, and
> treatment B leads to 20 people with condition X and 20 people with condition Y (which is equally serious).
>
> Which treatment should the company choose?
> | Certainly A | Probably A | Probably B | Certainly B |
>
> The company chose treatment A [B on half the trials].
>
> Would this choice make your more likely or less likely to choose this company as your insurer?
> | More likely | Probably more | Probably less | Less likely |
>
> [In minimal-test, the following was added.]
>
> A critic of the company argues against the company's decision by pointing out that the consequences were worse. The critic says that the decision amounts to a very simple choice:
> treatment A leads to 50 people with condition X and nobody with any other condition, and
> treatment B leads to 20 people with condition X and 20 people with condition Y (which is equally serious).
>
> [The wording is identical to minimal de-biasing]
>
> Is this a fair summary of what the decision is about?
> | Yes, completely | Basically yes | Not really | Not at all |

The basic form of the other three biases was as follows:

Zero. X and Y are two forms of a condition.

Treatment A can be given to 100 people with form X who come in each week, and it cures 60 of them.

Treatment B can be given to 50 people with form Y, and it cures all 50 of them.

Equality. A condition has two forms. Each week, 100 people come in with form X, and 100 with form Y.

Treatment A cures 80 of the 100 with form X and 40 of the 100 with form Y.

Treatment B cures 50 of the 100 with form X and 50 of the 100 with form Y.

Ex-ante. Treatment A can be given to 100 patients with a condition each week, and it cures 30.

Treatment B is in short supply, so it can be given only to 50 patients picked at random each week. It cures 40 of these 50.

2.2 Results

2.2.1 Biases

The items were designed so that one answer was optimal. Choice of the other answer could result from random error (misreading, or making an unintended response) as well as bias. However, we can at least compare the four different biases. Table 2 shows the frequency of each of the response options, with "worst" and "best" replacing A and B, according to consequences. It is apparent that the omission and zero-risk biases were stronger than the other two, where the "worst" response was rarely chosen. The table includes responses from all conditions, however, including the de-biasing conditions.

2.2.2 De-Biasing effects

In general, the minimal de-biasing manipulation reduced bias, but the expanded de-biasing manipulation did not. Table 3 shows the mean ratings on a scale on which 0 is neutrality and each step is 1. We compared each de-biasing manipulation to the combined results of the control condition and the minimal-test condition, which was identical to the control condition up to the point of the argument question. The minimal-test and control

Table 2: Frequency of responses, in percent, for question about what to do in the first experiment ("worst" and "best" indicate consequences).

	Certainly worst	Probably worst	Probably best	Certainly best
Choice question				
Omission	15.7	37.5	42.5	4.3
Zero	14.3	30.0	37.5	18.2
Ex-ante	2.5	14.9	51.4	31.2
Equality	2.4	10.1	61.7	25.8
Trust question				
Omission	16.6	34.6	41.7	7.1
Zero	13.6	28.9	39.5	18.0
Ex-ante	5.0	13.2	55.3	26.6
Equality	2.9	12.1	59.3	25.7

conditions did not differ significantly. Also, we combined the trust question and the choice question; these did not differ either.

The ratings for the minimal manipulation were significantly higher than the (combined) control condition ($t_{94} = 2.80$, $p = 0.0062$, two tailed). For individual biases, only the effects on zero and ex-ante were significant ($t_{94} = 3.87$, $p = 0.0002$, and $t_{94} = 3.22$, $p = 0.0018$, respectively). The result for the ex-ante bias suggests that this bias is in fact present, even though ratings were generally high.

The ratings for the expanded manipulation were almost significantly lower than the combined control condition, an effect opposite to the hypothesis ($t_{94} = -1.91$, $p = 0.0587$). These negative results were limited to the equality bias ($t_{94} = -2.64$, $p = 0.0096$) and the zero bias ($t_{94} = -2.81$, $p = 0.0060$).

The expanded manipulation may have failed for the equality bias because the difference between 80% cure rate and 20% may have stood out as a larger difference than that between 80% and 40% (see Appendix). If subjects judged inequality by comparing the maximum and minimum, then we would expect this effect. Likewise, the expanded form of the zero bias changed it into a proportionality bias (Fetherstonhaugh et al., 1997; Baron, 1997), which may have been stronger: although neither group was cured completely group B had a greater proportion. In sum, these reversed effects can be explained, after the fact, in terms of other biases. The fact remains that the expanded manipulations for equality and omission were ineffective, and these

Table 3: Mean rating (-1.5 to 1.5, with 0 indicating neutrality) as a function of type of bias and de-biasing condition. Positive numbers favor the optimal choice.

	Control	Minimal-test	Minimal	Expanded
Choice question				
Omission	-0.12	-0.16	-0.23	-0.08
Zero	0.08	0.11	0.30	-0.11
Ex-ante	0.56	0.58	0.71	0.60
Equality	0.60	0.66	0.66	0.52
Trust question				
Omission	-0.12	-0.12	-0.14	-0.05
Zero	0.12	0.12	0.28	-0.04
Ex-ante	0.52	0.46	0.63	0.53
Equality	0.60	0.62	0.57	0.52

Table 4: Frequency of responses, in percent, for question about whether the minimal condition was a "fair summary."

	Yes, completely	Basically yes	Not really	Not at all
Omission	20.0	50.0	25.3	4.7
Zero	30.5	38.9	25.8	4.7
Ex-ante	35.8	48.9	14.7	0.5
Equality	30.5	52.1	13.7	3.7

were exactly the two manipulations that were based on previous results.

2.2.3 Acceptance of arguments

The effectiveness of the minimal manipulation suggests that at least people saw the summary argument in terms of consequences as fair. Table 4 shows the ratings for the four bias conditions. A majority of subjects seemed to think that the summary was fair, although fewer thought so in the omission and zero conditions, where the biases were strongest.

2.2.4 Individual differences in the minimal argument

In general, subjects who thought the argument in the minimal condition was more fair were those who were less biased overall ($r = .59$ across all biases, combining choice and trust, $p = .0000$). This correlation was large and highly significant for each of the biases except for omission ($r = .16$).

These correlation could result from effects of prior beliefs on the evaluation of the minimal argument. Or it could result from the de-biasing effect of the argument on the biases in the minimal condition (and perhaps transferring elsewhere), or both. To assess the effect of the argument on the bias, we regressed the bias in the minimal condition on the bias in the (combined) control conditions and the evaluation of the argument, combining all biases and both choice and trust questions. Although the effect of the control condition was highly significant ($t_{92} = 12.25$, $p = .000$), the effect of argument evaluation was not significant ($t = 0.42$). In sum, we have no evidence that individual differences in the perception of argument fairness lead to individual differences in the effect of the minimal de-biasing manipulation. Rather, the evaluation of the argument seems to be affected by prior bias. Apparently, the minimal argument is equally effective for subjects with different prior opinions.

3 Minimalism and expansion: Second study

A second study used a different set of scenarios, designed to remove the problems with the expansion manipulation. They were also more like the original problems for which this manipulation had been effective. The experiment also included a question about whether the redescription should increase or decrease the bias.

3.1 Method

One-hundred seven subjects completed a questionnaire on the World Wide Web. Their ages ranged from 16 to 74 (median 35); 24% were male; 16% were students. The questionnaire began:

> **HMO**
>
> This questionnaire concerns decisions made by a health maintenance organization (HMO) about which treatments to provide for its members. A HMO is an organization that provides health care to individuals and families for an annual fee.
>
> On each screen, you will see some information about two treatments or means of preventing serious health conditions. The two options have the same cost, which is high. The HMO cannot afford to provide both for its members.

You will sometimes see information about the HMO's decision about whether or not to cover the treatment.

Each screen has one, three, or four questions. Treat each screen as a separate case, as if this is all you know about it. There are 16 screens.

The following example shows the questions asked, using equality bias as the example. (The other biases are in the Appendix.)

A HMO must make the following choice:

Each year, 100 patients in group I are having kidney dialysis and waiting for kidney transplants. 100 patients in group II are also on dialysis and waiting for transplants. Group I has a success rate of 80%, so 80 out of 100 will be able to give up dialysis. Group II has a success rate of 50%.

The HMO has two potential sources of kidneys. Each source will provide 100 kidneys, and one source must be chosen. Source A requires that all the kidneys go to group I. Source B requires that 50 kidneys go to each group.

[Added in the expanded condition:]

A third group of 100 patients is also on dialysis and waiting for transplants. The success rate for this group is 20%. Neither source will provide any kidneys for this third group.

[Added in the minimal condition:]

In other words, Source A will lead to 80 successes. Source B will lead to 65 successes (40 from group I, 25 from group II).)

Which option should the HMO choose?

Certainly A	Probably A	Probably B	Certainly B

[Added in all the de-biasing conditions:]

Does the above statement in **bold** conflict with the description that preceded it?

Not at all	Possibly	Yes

Should the above statement in **bold** make one option a better choice (or an even better choice)?

Definitely make A better	Possibly A	No change B	Possibly B
Definitely make B better			

[Added in the minimal condition:]

The HMO chose option A [B].

The HMO argues for it's decision by pointing out that the consequences were better.

[A critic of the HMO argues against the HMO's decision by pointing out that the consequences were worse.]

The HMO [critic] says that the decision amounts to a very simple choice: **Source A will lead to 80 successes. Source B will lead to 65 successes (40 from group I, 25 from group II).)**

Is this a fair summary of what the decision is about?

Yes, completely	Basically yes	Not really	Not at all

There were four conditions for each of the four biases: control (no de-biasing), expanded, minimal with the HMO choosing the biased option, and minimal with the HMO choosing the optimal option. The 16 screens were presented in an order randomized for each subject.

3.2 Results

Table 5 shows the mean bias scores, with 0 indicating neutrality and positive numbers indicating a bias. The unit is one step on the response scale (so that maximum is 1.5 for the choice question and 2 for the Change question). The change question is the one about whether the new information made one choice better. The two forms of the minimal condition did not differ, so they were combined.

Overall, the expanded condition reduced the biases, compared to the control condition ($t_{106} = 2.26$, $p = 0.0259$), and the effect of the minimal condition was nearly significant ($t_{106} = 1.92$, $p = 0.0575$, two tailed).

The answers to the question about whether the minimal argument was a fair summary did not differ across the biases. The distribution was as follows: Yes, completely — 28%; Basically yes — 47%; Not really — 21%; Not at all — 3%.

The answers to the conflict question did not differ, and were distributed: Not at all — 56%; Possibly — 31%; Yes — 13%. The effect of de-biasing was

Table 5: Mean ratings (–0 indicating neutrality, and each unit one step on the response scale) as a function of type of bias and de-biasing condition, second experiment. Positive numbers favor the optimal choice.

	Control	Expanded	Minimal
Choice question			
Omission	0.04	0.10	-0.02
Zero	0.32	0.59	0.40
Ex-ante	0.18	0.19	0.24
Equality	0.36	0.43	0.58
Change question			
Omission		0.20	-0.02
Zero		1.00	0.69
Ex-ante		0.25	0.50
Equality		0.33	0.91

significant for the cases on which the subject said "not at all" (compared to the respective control cases) for both extended de-biasing ($t_{84} = 2.12$, $p = 0.0371$) and minimal de-biasing ($t_{83} = 2.63$, $p = 0.0101$).

It is apparent in Table 4 that the various de-biasing manipulations worked differently for the different biases. The minimal manipulation was totally ineffective (going the wrong way) for omission bias. It is not clear why this is, but it is possible that the numerical equivalence of the minimal description was not apparent to some subjects.

4 Experiment 5: Effect of de-biasing manipulations

Baron (1992) found that anticipated emotion was associated with biases against the option that yields the best consequences. For example, when people chose an omission over a more harmful act, they tended to anticipate greater regret from the act than the omission. Baron found that de-biasing, by calling attention to the consequences and their effect on others, changed the anticipated emotion as well as the choice.

The present experiment asks whether cognitive de-biasing also affects current emotions associated with biases. By "cognitive," we mean that the de-biasing concerns the perception and interpretation of the situation, not the emotions. By contrast, emotional de-biasing might involve something like de-sensitization or mood induction.

We took the de-biasing methods, which we illustrate shortly, from Baron (2004). One is based on reduction of the case to its essentials in terms of consequences (after the original presentation). Baron (2004) found that subjects usually did not regard such reduction as a misrepresentation, and it did cause them to attend more to consequences. The other de-biasing method is based on elaboration, adding some detail that is consistent with the original statement but which reduces the salience of the argument for biased responding.

We used four biases: omission bias, zero risk, ex-ante risk, and equality. Omission bias is the judgment that harmful actions are worse than omissions that are even more harmful acts (Ritov & Baron, 1990; see Baron & Ritov, 2004, for recent discussion).

What we shall call zero-bias is the bias toward complete solutions, i.e., zero risk or zero harm (Baron, Gowda, & Kunreuther, 1993; Ritov, Baron, & Hershey, 1993). For an extreme example, people would rather reduce a risk from 10% to 0 than from 30% to 10%. This bias is related to the proportionality bias, in which people judge risk reductions by proportion rather than absolute change (Fetherstonhaugh et al., 1997; Baron, 1997); reduction to 0 amounts to complete reduction, even if the risk is small.

The ex-ante bias is the finding that people want to equate ex-ante risk within a population even when the ex-post risk is worse (Ubel, DeKay, Baron, & Asch, 1996a). For example, many people would give a screening test to everyone if the test would prevent 1000 cancer cases rather than give a test to half (picked at random) that would prevent 1200 cancer cases.

The equality bias is the preference for equal treatment of two groups, even when unequal treatment would be better on the whole (Baron, 1993a, 1995; Ubel, DeKay, Baron, & Asch, 1996b). For example, people would rather help 50% of each of two equal-sized groups than 80% of one group and 40% of the other.

4.1 Method

Seventy-six subjects completed the study. (Ten others were eliminated, 6 of which were much faster than others, and 6 of which answered the "Best" question incorrectly more often than not in the control condition, described below.) Their ages ranged from 20 to 62 (median 38), and 25% were male.

The questionnaire began with a short introduction explaining that the items concerned the policies of a Health Maintenance Organization (HMO). The omission bias case read as follows, with the reduction and elaboration arguments in brackets:

100 people are expected to get flu X this year.

Vaccine A prevents flu X in 50 of these people, and it leads to no other conditions.

Vaccine B prevents flu X in 80 of these people, but it leads to flu Y in 20 patients who would not have gotten flu without vaccine B.

Flu X and flu Y are both very serious, and equally serious. Both can cause death.

[*Reduction:* In other words, vaccine A leads to 50 people with flu, and vaccine B leads to 40 people with flu, 20 with flu X and 20 with flu Y (which is equally serious).]

[*Elaboration:* Vaccine B does not cause flu Y directly. It reduces people's resistance to catching flu Y from other people.]

How would you feel toward the HMO if it chose A?
Not angry at all   Somewhat angry   Very angry
As angry as I could be about a choice like this

How would you feel toward the HMO if it chose B?
Not angry at all   Somewhat angry   Very angry
As angry as I could be about a choice like this

Which option should the HMO choose?
Certainly A   Probably A   Probably B   Certainly B

Which option leads to the best outcome on the whole, for everyone?
Certainly A   Probably A   Probably B   Certainly B

The remaining items, each presented with the same four questions, were:

Zero

X and Y are two kinds of cancer, equally serious. Each year, 100 people get cancer X and 50 get cancer Y.

Treatment A is given to the 100 people with cancer X. It cures 60 of them.

Treatment B is given to the 50 people with cancer Y, It cures all 50 of them.

[*Reduction:* In other words, treatment A cures 60 cases of cancer, and treatment B cures 50 cases of cancer.]

[*Elaboration:* The total number of cancer cases of all types, including X and Y, is 1000 each year. Treatment A thus cures 60 out of 1000 and treatment B cures 50 out of 1000.]

Ex-ante

200,000 patients are covered by the HMO's "premium plan." There are two tests that screen for several types of cancer in the early stages.

Screening test A can be given to all 200,000, and it will prevent 50 deaths per year from cancer.

Screening test B is in short supply. It can be given only to 100,000 of the 200,000 patients in the premium plan. It will prevent 60 deaths per year from cancer.

[*Reduction:* In other words, test A prevents 50 deaths, and test B prevents 60 deaths.]

[*Elaboration:* There is not enough of test A or test B to give it to the 400,000 patients in the HMO's "standard plan" (the only other plan). Thus, test A can be given to 200,000 of the 600,000 patients covered by the HMO, and test B can be given to 100,000 of the 600,000 patients. Test A prevents 50 deaths and B prevents 60 deaths.]

Equality

Each year, 100 patients in group I are having kidney dialysis and waiting for kidney transplants. 100 patients in group II are also on dialysis and waiting for transplants. Group I has a success rate of 80%, so 80 out of 100 will be able to give up dialysis. Group II has a success rate of 50%.

The HMO has two potential sources of kidneys. Each source will provide 100 kidneys, and one source must be chosen. Source A requires that all the kidneys go to group I. Source B requires that 50 kidneys go to each group.

[*Reduction:* In other words, Source A will lead to 80 successes. Source B will lead to 65 successes (40 from group I, 25 from group II).]

[*Elaboration:* A third group of 100 patients is also on dialysis and waiting for transplants. The success rate for this group is 20%. Neither source will provide any kidneys for this third group.]

Each of the four biases was presented once with no de-biasing, once with reduction, and once with elaboration. In addition, we made up a control condition to go with each bias, in which the bias was absent but the difference in numbers was otherwise the same. For example, the control condition for ex-ante was:

15. There are two tests that screen for several types of cancer in the early stages.

Test A prevents 50 deaths per year from cancer.

Test B prevents 60 deaths per year from cancer.

The control condition for zero contained an error and the data from this item were not used. Performance in the control conditions was similar, however, and we replace the missing value with the mean response to the other three for each subject for each question.

4.2 Results

We formed a measure of Anger by taking the difference between the two questions about anger. Then we converted the direction of each measure — Anger, Choice, and Best (the last question) — so that positive numbers favored the option with the best consequences, and we transformed the scales linearly so that 0 represented neutrality between the two options, 1 represented the most extreme pro-consequence response (or combination, for Anger) and -1 represented the most extreme anti-consequence response. Table 6 shows the means for the conditions and questions.

In the Choice condition, all biases were present, in the sense of being lower (less consequence-oriented) than the control condition. Each difference between the first column (standard condition, no de-biasing) and the last (control) was highly significant (across subjects, all $p < .001$). Combining all four biases, both de-biasing manipulations were effective in making subjects more concerned with consequences ($t_{75} = 3.17$, $p = 0.0022$, for Elaboration; $t_{75} = 4.43$, $p = 0.0000$, for Reduction).

The Best condition behaved essentially like the Choice condition.

Some of the de-biasing manipulations worked better than others. For example, the Elaboration manipulation for Ex-ante did not work. This manipulation was taken from Ubel et al. (2001), where it had a small but significant effect. The differences among conditions are surely the result of the way items are worded and many other factors and are of no particular interest here. (Likewise, we cannot compare the two de-biasing methods in general, because the effectiveness of each, especially elaboration, depends on the details of its implementation.) We thus focus on the effects across all biases.

The main result of interest is that the effects of the de-biasing manipulations on Choice were also present for Anger (and the significant comparisons

Bias	Standard	Elaboration	Reduction	Control
		Anger		
Omission	−0.12	−0.08	−0.10	0.24
Zero	−0.05	0.07	0.08	0.29
Ex-ante	−0.03	−0.01	0.04	0.31
Equality	0.03	0.07	0.09	0.31
		Choice		
Omission	−0.23	−0.12	−0.13	0.56
Zero	0.05	0.26	0.22	0.66
Ex-ante	−0.08	−0.11	0.06	0.63
Equality	0.19	0.26	0.46	0.79
		Best		
Omission	−0.21	−0.11	−0.06	0.54
Zero	0.07	0.23	0.26	0.66
Ex-ante	−0.03	−0.01	0.19	0.66
Equality	0.20	0.32	0.46	0.79

Table 6: Means for conditions and questions, Experiment 5.

were all significant there too). This result is consistent with that of Baron (1992). It suggests that the emotional response to the two options is affected by their cognitive evaluation. When this evaluation changes as a result of a de-biasing manipulation, the emotional response does too. The de-biasing manipulations were directed at the subjects' perception, not at their emotional response, so they could not affect the emotion directly.

Biases in choice were also found in emotion. The difference between the standard condition (no de-biasing) and the control condition served as a measure of bias. The bias for choice was correlated .63 with the bias for Anger.

5 Conclusion

The success of each method suggests something about the nature of the bias. If the bias is reduced by asking subjects to calculate so as to understand the best outcome, this suggests that they did not pay much attention to the numbers initially. An alternative interpretation — not tested here — is that the calculation is a form of experimenter bias. That is, the request to calculate makes the subjects think that the experimenter has a “right answer" in mind for the remaining questions.

If bias is reduced by elaboration, this suggests that the bias iteself depends on a particular framing of the situation. Likewise, if bias is reduced by the minimal condition, it suggests that people do not spontaneously think of this re-framing themselves. Again, though, experimenter bias could be involved.

For practical purposes, the important thing is that the biases are labile, whatever the cause, and are subject to framing effects — subject to change with redescription of the same situation. They should thus not be taken as people's last, most considered, view on the issues.

De-biasing experiments can also tell us something about how to manipulate the biases. This information is part of the general psychology of persuasion. Persuasive methods can be used for any purpose by anyone. Very likely, persuasive methods can be devised to increase biases. Thus, although information about persuasion is useful, it is not in itself a solution to any problem.

References

Baron, J. (1993). Heuristics and biases in equity judgments: a utilitarian approach. In B. A. Mellers and J. Baron (Eds.), *Psychological perspectives on justice: Theory and applications*, pp. 109–137. New York: Cambridge University Press.

Baron, J. (1995). Blind justice: Fairness to groups and the do-no-harm principle. *Journal of Behavioral Decision Making, 8*, 71–83.

Baron, J. (1997). Confusion of relative and absolute risk in valuation. *Journal of Risk and Uncertainty, 14*, 301–309.

Baron, J., Gowda, R., & Kunreuther, H. (1993). Attitudes toward managing hazardous waste: What should be cleaned up and who should pay for it? *Risk Analysis, 13*, 183–192.

Baron, J. & Ritov, I. (2004). Omission bias, individual differences, and normality. *Organizational Behavior and Human Decision Processes, 94*, 74–85.

Fetherstonhaugh, D., Slovic, P., Johnson, S., & Friedrich, J. (1997). Insensitivity to the value of human life: A study of psychophysical numbing. *Journal of Risk and Uncertainty, 14*, 283–300.

Ritov, I., & Baron, J. (1990). Reluctance to vaccinate: omission bias and ambiguity. *Journal of Behavioral Decision Making, 3*, 263–277.

Ritov, I., Baron, J., & Hershey, J. C. (1993). Framing effects in the evaluation of multiple risk reduction. *Journal of Risk and Uncertainty, 6*, 145–159.

Ubel, P. A., De Kay, M. L., Baron, J., & Asch, D. A. (1996a). Cost effectiveness analysis in a setting of budget constraints: Is it equitable? *New England Journal of Medicine, 334*, 1174–1177.

Ubel, P. A., De Kay, M. L., Baron, J., & Asch, D. A. (1996b). Public preferences for efficiency and racial equity in kidney transplant allocation decisions. *Transplantation Proceedings, 28*, 2997–3002.