Decisions by coin toss: Inappropriate but fair

Gideon Keren*　　　　Karl H. Teigen

Tilburg University　　University of Oslo

Abstract

In many situations of indeterminacy, where people agree that no decisive arguments favor one alternative to another, they are still strongly opposed to resolving the dilemma by a coin toss. The robustness of this judgment-decision discrepancy is demonstrated in several experiments, where factors like the importance of consequences, similarity of alternatives, conflicts of opinion, outcome certainty, type of randomizer, and fairness considerations are systematically explored. Coin toss is particularly inappropriate in cases of life and death, even when participants agree that the protagonists should have the same chance of being saved. Using a randomizer may seem to conflict with traditional ideas about argument-based rationality and personal responsibility of the decision maker. Moreover, a concrete randomizer like a coin appears more repulsive than the abstract principle of using a random device. Concrete randomizers may, however, be admissible to counteract potential partiality. Implications of the aversion to use randomizers, even under circumstances in which there are compelling reasons to do so, are briefly discussed.

Keywords: coin toss, randomizers, equipoise, decision aversion.

1 Introduction

It has been said that life is a lottery and not a chess game. Even major events in life (like a person's sex, health, marriage, children, and career) are heavily influenced, if not wholly determined, by random factors (Bandura, 1982, 1998). Yet, rational decision makers prefer the chess game metaphor, where moves are restricted by rules and are selected according to how well they serve initially specified goals. Good decisions should be justifiable, that is, they should be determined by explicit, well defined, rational and ethical considerations, where strong and consistent reasons are given more weight than weak and capricious ones. Thus one would not place much credence on decisions that are simply a result of a random mechanism like a lottery or a coin toss. A doctor who prescribes medicine by simply opening his drug manual at random would quickly lose his patients (and eventually his license).

Even a rational decision maker may occasionally meet choices that cannot be resolved by reasons and arguments alone. Typical situations are choices between two equally valuable options (should I go right or left), or settling a dispute between two actors who both want an advantage (who should play the white pawns). In some such situations randomizers are accepted, and even institutionalized

as tie breakers. For instance, an old standing practice in soccer games is to toss a coin before the start of the game in order to determine which team will play the first half on which side and who is going to kick the ball first. There are several reasons justifying the use of a coin in this case: It gives both sides an equal chance and thus it is fair, it is efficient and fast and, most important, it probably has almost no effect on the final outcome of the game — its importance is negligible. In contrast when the outcome is important, a coin toss becomes less acceptable. Thus, in the 1968 European football championship, the semi-final between Italy and the Soviet Union was decided (after extra time) by a coin flip (Italy won and became the European champion). The fact that such an important game was randomly determined was so aversive that it was decided from then on to replace the coin toss by penalty kicks.[1]

People's reluctance to use a randomizer as a decision device, has been demonstrated in field studies. Oberholzer-Gee, Bohnet and Frey (1997) studied the social acceptability of different decision procedures, including a lottery, to determine the sitting of nuclear waste facilities in Switzerland. Using a large representative sample, they asked participants to rate alternative decision

*The authors would like to thank Talia Keren, who initiated our interest in this topic. Address: Gideon Keren, Tilburg Institute for Behavioral Economics Research (TIBER), Department of Social Psychology, Tilburg University, P.O. Box 90153, 5000 LE Tilburg, The Netherlands. Email: G.Keren@uvt.nl.

[1]There are strong arguments for asserting that penalty kicks are in many respects a different type of randomizer. However, even if one accepts that claim, penalties apparently are perceived as dependent on skill (which supposedly soccer is all about), and the random component which is by far the most determinant factor, is concealed. More often than not, people are unaware of the overwhelming role of chance encounters in their life (Bandura, 1982) because the underlying randomizer is not sufficiently transparent.

procedures to determine the place for locating future nuclear waste repositories. Two important results emerged from their study. First, using a lottery was located low in the list — only 26% respondents rated this procedure as acceptable. Second, the more acceptable methods, such as negotiations, the current method used which ultimately requires the federal parliament approval, and handing the decision to foreign experts, all are assumed implicitly to eventually be reason-based.

Randomness is in particular repudiated in the legal system, which, like rational decision making, is believed to rest entirely on reason and reason alone. The use of an explicit randomizer, such as the coin, is utterly forbidden. Thus, the Michigan Supreme court publicly reprimanded a Wayne county judge for flipping a coin to decide where two girls in a custody dispute would spend Christmas (Aschenfelter, 2003). In its decision, the court claimed that "The public's trust and confidence in the judiciary was damaged." In another case, a judge in a N.Y. City court flipped a coin to determine whether to sentence an individual to 20 or 30 days in jail. Although the coin flip produced an outcome in an inexpensive and prompt fashion, the judge was censured. His critics did not complain that he had reached the wrong decision. Rather, they complained about process. Specifically, the critics claimed that "the coin flip offended this society's commitment to rationality." Evidently, whether or not a judge's mental processes actually amount to anything more than a mental coin flip, the community wishes judicial rulings to appear to be the product of contemplative, deliberative, cognitive processes.

There could be several reasons why people dislike the use of a coin. First, the use of a randomizer implies relinquish of control, a property that people rarely want to give up (Skinner, 1996). Second, there are contexts in which the use of randomizers, coin being a *par excellence* example, is associated with negative connotations such as gambling. Third, and perhaps more important, it is expected that people could justify their choice by adequate arguments and reasons (Shafir, Simonson & Tversky, 1993). Fourth, they should be accountable for their decisions (Tetlock, 1992), in particular when the consequences of the decision are of great importance. Coins, and other randomizers, cannot deliver appropriate reasons and cannot be regarded as responsible agents, thus they may be objectionable both on rational and moral grounds.

However, there are cases where the use of a randomizer is rationally and ethically defensible, and could even be recommended. These are situations where the decision maker is unable to attach more importance to the arguments favoring one alternative than to the arguments favoring the other, in short in situations of indifference. Although rational choice theory (e.g., classical utility theory) is rather silent about situations of indifference and

never prescribes explicitly a decision by lot, it is compatible with the idea that a decision-maker who is unable to make up her mind should use a random device to determine her choice.

Indifference, however, may under some circumstances be a misleading term because it seems to imply lack of concern and lack of feelings. One may indeed be indifferent between having fish or steak for dinner, but a physician who has to allocate a kidney to one rather than another patient can hardly be characterized as being indifferent. Elster (1989) proposed to distinguish between situations in which the choice consequences are insignificant and hence the decision-maker is truly *indifferent* and *indeterminacy* in which the consequences are highly important but the arguments in both directions can be judged as equally strong.

Also, from an ethical point of view the use of randomizers is sometimes recommended, especially in situations where it is important to ensure an unbiased procedure. Lotteries or lottery-like procedures have been used not only for deciding who should win the million-dollar prize, but also for less rewarding outcomes like drafts for military service, car controls, or in assigning patients blindly to the control or experimental condition in tests of a novel medical treatment. Such procedures have the advantage of being perceived as *fair,* by giving everybody *an equal chance* of being selected (Broome, 1984).[2] Under such circumstances, the decision *to use a randomizer* may be seen justifiable and based on moral and rational considerations, even if the final outcome of the decision is due to chance.

1.1 The present studies

In this article we present a series of studies designed to examine the extent to which the use of a randomizer, such as a coin flip, is considered as an acceptable decision device. There are at least two general conditions under which a decision maker may consider the use of a coin. First, and most important, a coin may be used in situations of indecision, in particular when the two alternatives are equally attractive (unattractive) and the relevant attributes are non-compensatory and can not be directly compared (e.g., Luce, Payne, & Bettman, 1999). Sec-

---

[2]Despite the uncompromising cling of the legal system to a reason-based approach, it has not rejected the use of a randomizer under all circumstances. In a well known case (U.S. versus Holms; see Broome, 1984), the captain of a sinking ship was accused of being involved in throwing people overboard from an overcrowded lifeboat, in 1841. Judge Baldwin found the captain guilty not because he did not prevent the slaughter of people, but rather because he had not done it by lot: "In no other than this or some like way are those having equal rights put upon equal footing …" (Broome, 1984, p. 38). In this case, the ethical imperatives of equal chance (for the shipwrecked) and impartiality (on the side of the decision maker) seem to have overruled the need for reason-based and justifiable choice.

ond, a coin may be contemplated under circumstances of what Beattie, Baron, Hershey, and Spranca (1994) termed "decision aversion", situations in which people prefer to avoid making decisions. Such situations usually arise when the consequences of the decisions are of tremendous importance, as for example determining who among two soldiers will be send on a mission in which the chance of survival is small. Third, a coin or a similar lottery procedure might be advocated in situations where it is important to avoid a biased decision, for instance by a partial or prejudiced decision maker. Our study investigates whether, and under what conditions, people are willing to accept the use of a randomizer in such situations.

Consider the following scenario which, supposedly, most people would find difficult to cope with, and will probably lead to decision aversion. A young man and an old lady were involved in a serious car accident. The police have determined, unequivocally, that it was the young driver's fault: He did not comply with a stop sign. The two victims are brought to hospital in a critical condition. The physician on duty determines that to save their lives, both have to be placed immediately in the intensive care unit. Unfortunately, only one unit is currently available and the physician has to decide which of the two should be saved. Presumably, the arguments in favor of saving each of the two victims are equally strong. On one hand, the accident has been caused by negligence of the young man, suggesting that the old lady should be saved. On the other hand, the life expectancy of the old lady is much shorter, suggesting that the young man should get the priority. Which of the two arguments should prevail?

The first experiment tested the extent to which people were willing to accept the use of a randomizer, more specifically a coin, in a situation of indeterminacy, using the accident scenario described above. There were two main conditions. In the *judgment* condition, participants judged whether there were stronger reasons for one of the two options (i.e., saving the young man or saving the old lady) or whether the reasons were equally strong. In the *choice* condition, participants were asked to choose between the two options or alternatively use a coin to determine their choice. It was hypothesized that in the judgment condition most participants would acknowledge the indeterminacy of the situation (i.e., admit the existence of equally strong reasons supporting one or the other option). Nonetheless, it was predicted that because of people's reluctance to use a randomizer in important decisions (involving life and death), few participants would accept the use of a coin. In other words, we expect in these situations a discrepancy between people's judgments and their actual choice decisions.

In the subsequent experiments, we tested several variations over the same theme. To make sure that the re-

sults of Experiment 1 were not due to peculiarities of the young-man/old-lady scenario, a kidney allocation scenario, which perhaps could claim a higher degree of realism, was introduced in Experiment 3. In successive experiments, we tested the following factors that could possibly affect people's willingness to use a coin in situations of indeterminacy, namely (1) seriousness of outcome; (2) focus on indeterminacy; (3) focus on fairness; and (4) the nature of the randomizer.

(1) People may find it especially problematic to use a coin flip to decide serious outcomes, like matters of life or death, even in situations they admit to be indeterminate, resulting in high *judgment-decision discrepancies.* In situations of less importance, we expect people to be more willing to accept random procedures as a decision aid. In such situations we predict the judgment-decision discrepancy to be reduced or even to disappear. This prediction is tested in Experiment 4. Using a similar logic, one might think it would be easier to use a randomizer when the outcome is not categorical (life or death), but probabilistic (a smaller or greater chance of life or death). Probabilistic alternatives are, in a sense, less severe than deterministic ones; after all, the "losing" patient in the accident scenario is not condemned to die, but merely to a reduced chance of survival. This possibility is explored in Experiment 5.

(2) People may decline to use a randomizer in the hope of eventually finding a more rational solution to the decision problem, even in cases where they do not see the advantage of one option over the other and have a hard time making up their mind. Elaborate descriptions that underscore the indeterminacy of the situation could remove this possibility, and might accordingly make people more willing to accept a randomizer as a way out of the impasse. This could be achieved by casting the decision as a conflict between individual opinions, or letting them explicitly acknowledge the lack of other solutions. The effects of such prompts are studied in Experiment 6. Furthermore, an increase in similarity between the alternatives might make people realize that there are no compelling reasons for favoring one or the other option and thus be more amenable to use a randomizer. Thus in Experiment 7 participants are given a choice between saving two middle-aged ladies rather than the young-man/old-lady dilemma in our original scenario. With more similar options, the likelihood of finding a correct rational solution is reduced, and perhaps also the decision maker will feel that the outcome makes less of a difference.

(3) A chance device is not only helpful for the decision maker but also assures other implicated parties against partial or biased decision making. People might be more favourably disposed towards chance procedures when focus is directed toward their fairness. This can be done simply by asking people to indicate not just what are pre-

ferred or acceptable ways of making decisions in case of indeterminacy, but what is a *fair* way of doing it. In the present studies we ask which of three alternatives in the accident scenario is more fair (Experiment 2), and instruct participants to rate decisions made by a randomizer for fairness as well as for acceptability (Experiment 7). In addition, we describe in Experiment 10 a situation where the decision maker may be biased (or suspected of being biased) towards one of the alternatives. The question we raise is whether the use of a coin, which may eliminate the suspected bias, is now acceptable.

(4) Randomness is an intricate concept. Although every one "knows" what randomness is, it somehow eludes satisfactory definition. As noted by several authors, randomness lacks both an adequate and precise definition as well as a satisfactory and decisive test that will determine its presence (e.g., Falk & Konold, 1997; Pashley, 1993). Moreover, a large body of research has shown that people's intuitions of randomness are not always compatible with formal models (e.g., Bar-Hillel & Wagenaar, 1993; Falk, 1991; Neuringer, 1986). It is thus not obvious what people are willing to accept as an adequate random procedure. Such procedures can differ on a concrete level; one can, for instance draw lots, toss a coin, or pass the decision to another agent that is supposed not to be biased in any particular direction. Will such procedures differ in how random they are perceived and in turn how acceptable, and how fair, they are judged to be? This is an issue explored in Experiment 8.

Random procedures, however, can also be described at different levels of abstraction. Vallacher and Wegner (1987) argued in their action identification theory that an action described on a specific level ("I am reading an article") have different connotations from the same action described at a more general level ("I am studying"). Problematic procedures are typically described on a more detailed level than smoothly running, routine activities. Analogously, suggestions about "tossing a coin" might evoke other, and perhaps more objectionable associations than "using a randomizer". In construal level theory, Trope and Liberman (2003) have discussed how events that are psychologically (for instance temporally) close tend to be described on a more concrete level than events that are perceived as more distant. When people focus on far and abstract tasks they typically focus more on advantages and reasons *why* they should be done, whereas immediate prospects draw attention to procedural details, with a primary concern for *how* they should be performed. As a result, a task or a procedure described on low levels of construal will often result in more negative evaluations than the same task construed at a higher, more general level. Experiment 9 was designed to test the acceptability of random procedures described at different levels of construal, to test whether the acceptability

of chance in *principle* is similar or different from the acceptability of using a concrete randomizer like the coin.

2 Experiment 1

2.1 Method

Participants. A total of 119 undergraduate students were recruited at the campus of Eindhoven University of Technology and were asked to participate in a task that lasted for a few minutes. They received a chocolate bar for their participation.

Design and procedure. All participants were presented with the following scenario:

> Dr. Freedman and Dr. Peterson were both on duty in the emergency room when the ambulance arrived with two critical injures, an old lady aged 69 and a young man aged 32. They were injured in an accident, caused by a collision between their two cars. The police has determined, unequivocally, that it was the young man's fault — evidently, the lady had the right of way. The two physicians agree that, in order to save their lives, both the old lady and the young man have to be placed immediately in an intensive care unit. Unfortunately, the hospital is a small one and currently there is only one intensive care unit free. The two physicians are faced with the problem of whom should they place in the intensive care unit: The old lady or the young man? They both realize that this difficult decision needs to be made very quickly.

> Dr. Freedman believes that the old lady should get priority because the accident was caused by the young man's fault. Dr. Peterson believes that the young man should get priority because his life expectancy is obviously much larger. They are uncertain as to what to do, and quickly consult a third physician who happened to be near by.

Participants In the judgment condition were further told: "This physician holds that both arguments are equally valid and equally strong."

Participants in the choice condition were further told: "This physician holds that both arguments are equally valid and equally strong. Hence, he believes that the best way to solve the dilemma is simply by using a coin flip."

The experiment consisted of two conditions. Participants in the judgment condition were asked to indicate with whom of the three physicians they most agreed. Their task was to judge whether the arguments were more valid and weighed heavier in favor of (a) saving the old

Table 1: Response distributions in Experiment 1, 2 and 3 as a function of judgment or choice.

Experiment 1: Accident scenario

	(a) Old Lady	(b) Young man	(c) Equality*	*N*
Judgment	10 (16.4%)	15 (24.6%)	36 (59.0%)	61
Choice	17 (29.3%)	31 (53.4%)	10 (17.2%)	58

Experiment 2: Accident scenario (Fairness)

	(a) Old Lady	(b) Young man	(c) Equality	*N*
Judgment	4 (8.9%)	7 (15.6%)	34 (75.6%)	45
Choice	10 (22.2%)	19 (42.2%)	16 (35.6%)	45

Experiment 3: Kidney transplantation scenario

	(a) Robert	(b) John	(c) Equality	*N*
Judgment	23 (30.3%)	13 (17.1%)	40 (52.6%)	76
Choice	34 (44.7%)	22 (28.9%)	20 (26.3%)	76

* Equality implies flip of a coin in the choice condition and judging the arguments being equally persuasive in the judgment condition.

lady, (b) saving the young man, or (c) both arguments were equally valid and of equal weight. Participants in the choice condition were asked to decide which of the following would be the most appropriate action: (a) save the old lady's life, (b) save the young man's life, or (c) make the decision by the flip of a coin. The scenario was constructed such that the arguments for options (a) and (b) would be assessed as equally appealing, and hence it was expected that a large number of participants in the judgment condition would opt for option (c). Following the above discussion, however, it was predicted that in the choice condition, the corresponding option (c), namely the flip of a coin, would not be endorsed by most participants.

2.2 Results and discussion

In both conditions of this experiment, as in the various conditions of the following experiments, the order of arguments for both choice options, (e.g., in the present scenario, the arguments for saving either the old lady or the young man), were counterbalanced. In all experiments, no systematic differences due to the order of presenting the arguments were found. Consequently, in all the analyses below the data were combined across order presentations.

The results are portrayed in the upper part of Table 1. The frequencies for options (a) and (b) were combined and compared to the frequency of option (c). As predicted, a majority of participants in the judgment condition (36 of 61, or 59.0%) judged the arguments for both options to be equally compelling, i.e., carrying equally

strong weights. In contrast, only a minority in the choice condition (10 out of 58, or 17.2%) considered the corresponding option of using the coin to be the most appropriate action. The difference between the two proportions is highly significant ($z=4.68$; $p<.001$). Evidently, although the two options were judged to be equally (un)attractive, it does not justify the use of a coin. One possible explanation for this *judgment-choice discrepancy* is that choice, unlike judgment, implies commitment (e.g., Ganzach 1995), and participants refuse to make such a commitment based on a coin flip.

3 Experiment 2

Coin tosses and random draws have been used in situations of indeterminacy not only to break ties but also to make sure that the decisions are unbiased and not (consciously or unconsciously) influenced by irrelevant sympathies or prejudices on the part of the decision maker. Random procedures can be considered fair in the sense that none of the options are given an advantage over the other. Experiment 2 was conducted to investigate whether a focus on fairness would make participants more favourably disposed towards decision by coin toss. The same scenario as in Experiment 1 was employed, except that this time participants were asked to make their choice and judgment decisions on the basis of fairness (rather than on the basis of what is more appropriate).

3.1 Method

Participants. A total of 90 undergraduate students were recruited at the campus of the Pabo College in Eindhoven, and were asked to participate in a task which lasted a few minutes. They were randomly assigned to the choice (*N*=45) or the judgment (*N*=45) task.

Design and procedure. The experiment was identical to Experiment 1, except that participants receiving the judgment task were asked to judge whether the arguments for saving the old lady were most fair, the arguments for saving the young man were most fair, or that the arguments for the two protagonists were equally fair. Similarly, participants in the choice task, were asked to decide which of the three choice decisions — saving the old lady, saving the young man or, flipping a coin — was the most fair one.

3.2 Results and discussion

As shown in the middle panel of Table 1, a majority (75.6%) of the participants in the judgment task considered the arguments to be equally fair. In contrast, a significantly smaller proportion (35.6%) of participants in the choice task considered it most fair to use a randomizer. Apparently, although most participants in the judgment condition tend to agree that the arguments in support of each of the two protagonists are equally fair, only a minority in the choice condition accepts the coin flip as a fair solution. Evidently, the aversion to using random devices for resolving a conflict is deeply rooted. The pattern of results of Experiment 2 is highly similar to that obtained in Experiment 1. Comparing the results of Experiment 1 and 2 using a Breslow-Day homogeneity of odds ratios test (Agresti, 1996) yielded a non-significant result (χ^2 = .109; *p* = .74).

Fairness can be thought of in two distinct ways. One, usually referred to as *distribution (allocation) fairness,* involves equal sharing. Thus dividing $100 fairly among two persons would imply that each one gets $50. In the above scenario, it would entail that since the arguments for saving either the old lady or the young man are perceived as equally strong, both should get equal opportunities (chances) of survival. A second interpretation of fairness (which does not contrast the previous one), concerns the process by which the decision has been made. For instance, consider two sisters who inherited an old family piano. One may believe that deciding who of the two sisters should get the piano by asking a neutral judge is a more fair procedure than tossing a coin. Fairness in the present context seems to be more related to process fairness, a point on which we further comment in the discussion of Experiment 9.

4 Experiment 3

The main characteristics of the accident scenario, employed in Experiment 1 and 2, are the importance of the outcomes (life or death), the irreversibility of the decision, the lack of any compelling normative considerations that could tilt the decision in one or the other direction, and the urgency to reach a decision fast required by the situation (procrastination would lead to the worse outcome namely both victims will die). To obtain further generality, the main results were tested using a different scenario with similar characteristics, which may be perceived as being perhaps more realistic than the accident scenario. Specifically, a scenario was employed in which a decision had to be made as to who, among two patients, will receive the only kidney available. The situation of indeterminacy was achieved by the fact that one patient was younger and thus had a longer life expectancy while the other had, according to the physicians, a higher chance of a successful transplantation.

4.1 Method

Participants. A total of 152 undergraduate students, recruited at the campuses of the university of Nijmegen participated in the experiment. They performed the task on a computer (laptop), along with several unrelated judgmental tasks, and were paid €5.50 (approximately $7.00 at the time the experiment was conducted) for their participation.

Design and procedure. All participants were exposed to the following scenario:

> John and Robert are two patients waiting urgently for kidney transplantation. The hospital just received a kidney that matches the requirements of the two patients. The physicians responsible for the transplantation are facing a difficult decision, namely to whom they should allocate the kidney. On one hand John, aged 50, is younger than Robert, aged 57. On the other hand, Robert's condition is better and the three physicians believe that he has an 85% chance that the transplantation will be successful compared with John whose chance for a successful transplantation they assess as 75%. As you may note, John has a slightly higher life expectancy. On the other hand, Robert has a slightly higher chance for a successful operation.
>
> - Physician A thinks that the kidney should be given to John because, given that the transplantation is successful, he has a longer life expectancy compared to Robert.

- Physician B believes that the kidney should be given to Robert because he has a higher chance for a successful transplantation than John.

- Physician C believes that the arguments to allot the kidney to John are equally strong and equally convincing as the arguments to allot the Kidney to Robert.

Participants in the *judgment* condition were further asked:

If you had to judge, which of the following options would you most agree with:

- The arguments to allot the kidney to John, who has a longer life expectancy, are more compelling.

- The arguments to allot the kidney to Robert, whose chance for a successful transplantation is higher, are more compelling.

- The arguments to allot the kidney to one or the other patient are equally strong and equally compelling.

Participants in the *choice* condition were further asked:

If you had to make the choice, which of the following options would you choose.

- Allot the kidney to John, who has a longer life expectancy.

- Allot the kidney to Robert, whose chance for a successful transplantation is higher.

- Toss a coin to decide to whom to allot the kidney because the arguments in favor of the two patients are equally compelling.

In both conditions, the equality (coin toss) option was in the third place, and the first two options (regarding Robert and John) were counterbalanced.

4.2 Results

The results are portrayed in the lower part of Table 1. As can be seen, there was a slight preference, in both the judgment and choice conditions, in favor of Robert (the patient who had a higher chance for a successful operation). Hence, we did not achieve complete indeterminacy. Notwithstanding, and more important, there was as expected a substantial and statistically significant difference ($z=3.30$; $p < .001$) between the proportion of participants who judged the reasons for allocating the kidney to one or the other patient as being equally convincing (53%), and the corresponding number of participants in the choice condition who opted for the coin toss (26%).

These results further demonstrate the deep rooted aversion that people posses regarding the use of a coin as an acceptable decision device for solving complex and important decision problems.

5   Experiment 4

Do the above results generalize also to situations in which the consequences are less imperative, and the question is not one of life or death? Experiment 4 was designed to test Elster's (1989) hypothesis concerning the distinction between *indifference* and *indeterminacy*. Specifically, as in Experiments 1–3, participants were allocated to either the judgment or the choice conditions. Two scenarios (between subjects) were used; one with "medium" important consequences (first authorship of an article) and one with "low" important consequences (whether to attend a theater play or a concert). It was predicted that, when decision consequences are of minor importance, participants would be less reluctant to use the coin.

5.1   Method

Participants. A total of 286 undergraduate students, recruited at the campuses of Nijmegen University and the Free University of Amsterdam, participated in this and other, unrelated decision-making tasks, for which they were paid an equivalent of approximately $5.00.

Design and procedure. Participants from Nijmegen University (*N*=127) were exposed to the "first author" scenario (medium consequences), and were randomly assigned to the choice or the judgment task. They were presented with the following cover story.

John and Peter have written a short article for a computer science journal. The article was praised by the editor as original and well written and was accepted for publication. Before sending the final version for print, the two authors have an argument regarding the order of the authors' names. John claims that he was the one that came up with the original idea and therefore his name should appear first. Peter claims that he was the one that has actually written the article and has done most of the work, and thus his name should appear first. They are uncertain as to what to do, and quickly consult a friend who happened to be nearby.

Participants in the *judgment* condition were further told: "The friend thinks that both arguments are equally valid and equally strong."

Participants in the *choice* condition were further told: "The friend thinks that both arguments are equally valid

Table 2: Response distribution in Experiment 4 as a function of judgment or choice.

1. First author scenario			
	(a) John	(b) Peter	(c) Equality	*N*
Judgment	11 (17.2%)	26 (40.6%)	27 (42.2%)	64
Choice	16 (25.4%)	28 (44.4%)	19 (30.2%)	63
2. Entertainment scenario			
	(a) Theater	(b) Concert	(c) Equality	*N*
Judgment	7 (9.0%)	38 (48.7%)	33 (42.3%)	78
Choice	8 (9.9%)	31 (38.3%)	42 (51.9%)	81

and equally strong. Hence, he believes that the best way to solve the dilemma is simply by using a coin flip."

Participants from the University of Amsterdam (*N*=159) were exposed to the "choice of entertainment" scenario, and were randomly assigned to the choice or judgment task. They were presented with the following cover story.

> Both Richard and Brad have just obtained their driving license. They want to celebrate the event with an evening of entertainment. On the evening that suits them both, a play is performed at the theater and a jazz concert is given at the music hall. The two friends find it difficult to choose. The play will be performed by a very good theater company, and they are both fond of the music that will be played at the concert.
>
> Richard suggests to go to the theater. The theater is much easier to reach than the music hall. Also, he has not seen the auditorium since its recent renovation. Brad suggests to go to the concert. The music hall is situated far out of the city centre, but next to it is a nice restaurant where they could go and have dinner before the show. Besides, he loves the music hall's cozy atmosphere. They are uncertain as to what to do and quickly consult a friend who happens to be nearby.

Participants in the *judgment* condition were further told: "Their friend thinks that both arguments are equally valid and equally strong."

Participants in the *choice* condition were further told: "Their friend thinks that both arguments are equally valid and equally strong. Hence, he believes that the best way to solve the dilemma is simply by using a coin flip."

As in the previous experiments, participants in the judgment task were asked to judge for which option (John or Peter in the first scenario; theater or concert in the second scenario) there were more compelling reasons or, as

third alternative, whether the reasons for the two options were equally compelling. Participants in the choice condition had to choose one of the two options or opt for a coin flip.

5.2   Results and discussion

The results are portrayed in Table 2. In the authorship scenario (mild consequences), the proportion of participants in the judgment task that considered the arguments to be equally strong (27 out of 64, or 42.2%), was larger but did not differ significantly from the proportion of participants in the choice task (19 out of 63, or 30.2%) that considered the use of a randomizer to be appropriate. In the choice of entertainment scenario (unimportant consequences), there seemed to be a clear preference for the concert over the theater. More important, however, the proportion of participants that judged the arguments for the two events to be equally forceful (33 of 78, or 42.3%) did not differ significantly from the corresponding proportion in the choice task that considered the use of a randomizer appropriate (51.9%). Further, we compared the proportion of participants who accepted the coin in the choice conditions of the three different scenarios of Experiments 1 and 3. Using the Cochran Armitage test (Agresti, 1996) indicated a significant trend (*z*=4.28; *p* <.001) in which the proportion of participants accepting the use of a coin was the smallest in Experiment 1, larger for the first author scenario, and even larger for the choice of entertainment scenario.

In sum, the pattern of responses to the choice and judgment conditions differed significantly only in the case of severe consequence (Experiments 1–3), as opposed to situations where the consequences were relatively unimportant (Experiment 4). Participants were less reluctant to accept a randomizer in cases where they were indifferent, whereas they preferred to make a choice in a situation of indeterminacy.

Table 3: Response distributions of choices in the two conditions of Experiment 5.

	(a) Old lady	(b) Young man	(c) Coin toss	N
Deterministic choice	13 (29%)	18 (40%)	14 (31%)	45
Probabilistic choice	13 (29%)	19 (42%)	13 (29%)	45

6 Experiment 5

Experiment 5 asked whether a concrete random procedure (such as coin toss) is more acceptable in a situation where the outcome itself is uncertain. In the original version of the old-lady/young-man scenario, the coin toss literally determined life and death for the patients involved. It may be considered less "callous" to let the coin decide upon treatment priorities, condemning none of the victims to a sure death, but use it for allocating them to two treatment facilities with different chances of survival. In the present experiment the coin gave both patients a chance of being saved, but with different probabilities.

6.1 Method

Participants. A total of 90 undergraduate students were recruited at the campus of the Pabo College in Eindhoven. They performed the task on a computer (laptop), along with several unrelated judgmental tasks, and were paid €4.00 (approximately $5.00) for their participation.

Design and procedure. All participants were exposed to the basic scenario employed in Experiment 1. They were randomly allocated to one of two conditions. One condition (deterministic choice) was an exact replication of the choice condition of Experiment 1. The other condition was very similar except that participants in this condition were told that the hospital had two (rather than one) intensive care units. The two units, however, were not identical: One was a modern newly purchased unit while the other one was an old model, purchased 25 years ago, that lacked several of the new model features. The physicians estimated that whoever (of the two injured) is going to be placed in the new unit, has an 80% survival chance. They further estimated that whoever (of the two injured) is going to be placed in the old unit, has a 40% chance to survive.

As in the choice condition of Experiment 1, participants in both conditions had to choose one of three alternatives (save the young man, save the old lady, or toss a coin).

6.2 Results and discussion

The results are portrayed in Table 3. The percentage of participants who chose the coin toss is almost the same in both conditions, and is not different from the comparable percentage in the choice condition of Experiment 1. The results serve as yet another replication of the reluctance to use the coin. Further, the phenomenon is robust and is equally strong also under probabilistic conditions, namely when the decision does not unequivocally determine life and death. Thus, even if less important decisions allow for the use of a coin, as demonstrated in the previous experiment, the present attempt to reduce the severity of the accident scenario (by varying probabilities rather than outcomes) was not successful in making a randomizer more acceptable.

7 Experiment 6

Reluctance to use a randomizer could be due to a hope of finding a better and more rational solution. In other words, participants may believe that the situation is not completely indeterminate after all, and that a reason-based solution may be found out of the impasse. The following two experiments examine the effects of emphasizing the indeterminacy of the situation, and in this way attempt to weaken one source of resistance to the coin. In Experiment 6, this is done in one condition by making it clear that none of the physicians is going to be swayed by the other's argument, an in another condition by explicitly pointing out the lack of an alternative procedure.

7.1 Method

Participants. A total of 314 undergraduate students, recruited at the campuses of the universities of Tilburg, Nijmegen, and Utrecht participated in the experiment. They performed the task on a computer (laptop), along with several unrelated judgmental tasks, and were paid €5.50 (approximately $6.50 at the time the experiment was conducted) for their participation.

Design and procedure. Participants in this experiment were presented with a very similar version of the cover story employed in Experiment 1. They were allocated to one of three conditions. In each condition, half of the participants were assigned to the judgment and the other half to the choice task. The *original* condition was an exact replication of Experiment 1 except that the old lady's age was changed from 69 to 59 years, because preceding results had shown that among participants who were not

Table 4: Response distributions in Experiment 6 as a function of judgment or choice.

	(a) Old Lady	(b) Young man	(c) Equality	N
I. Control Condition				
Judgment	12 (19.7%)	12 (19.7%)	37 (60.6%)	61
Choice	20 (35.7%)	23 (41.1%)	13 (23.2%)	56
2. Conflict Condition				
Judgment	5 (9.4%)	10 (18.8%)	38 (71.7%)	53
Choice	13 (27.1%)	12 (25.0%)	23 (47.9%)	48
3. No better option				
Judgment	11 (22.5%)	10 (20.4%)	28 (57.1%)	49
Choice	16 (34.0%)	12 (25.5%)	19 (40.5%)	47

indeterminate (in either the choice or the judgment condition), there was a slight tendency to prefer saving the young man. Participants in the *conflict* condition read the same cover story with the addition that the disagreement among the two physicians was emphasized by noting that the two were very strong-minded about their own preference (implicitly suggesting a conflict that leads to an impasse). In the *no better option* condition, the scenario was again the same except that the third physician drew the attention of the other two physicians to the question whether there were any better choice alternatives to the use of a randomizer. The experiment thus consisted of a 3 (Original, Conflict, and No better option conditions) X 2 (judgment vs. choice) between-subjects design.

7.2   Results and discussion

Table 4 shows that overall, participants who chose either the old lady or the young man were about equally distributed over the two options, indicating that participants, as a group, did not have a clear preference for either one of the two options. As before, the focus of interest is on the distribution of responses over the randomizer versus choosing one of the options, i.e., the old lady or the young man, data were collapsed over the latter two for further analysis.

Results in the *original* condition replicated earlier findings. A majority (61%) of the participants in the judgment condition considered the arguments for saving the old lady or the young man as equally strong. In contrast, less than one fourth (23%) of the participants in the *choice* condition considered the use of a coin to make a decision as appropriate. As in Experiment 1, the proportion of participants in the judgment condition assessing the arguments to be equally strong was significantly larger than the corresponding proportion in the choice condition who assessed the coin to be acceptable ($z =$

4.09; $p < .0001$). These results again show people's reluctance to use a coin in a situation of indeterminacy.

In the *conflict* condition, again a majority (72%) of participants in the judgment condition considered the arguments for saving the old lady or the young man as equally strong. About half (48%) of the participants in the choice condition were willing to use a coin to make a decision. The proportion of participants accepting the coin in the judgment condition was significantly larger than the corresponding proportion in the choice condition ($z = 2.44$; $p < .001$), though the effect size was smaller than in the original condition.

Finally, in the *no better option* condition, a small majority (57%) of participants in the judgment condition considered the arguments for saving the old lady or the young man as equally strong, yet less than (40%) of the participants in the choice condition were willing to use a coin to make a decision. The difference was in the same direction as in the other two condition, yet (close to but) not statistically significant ($z = 1.63$; $p = .052$).

Examining only participants who had to make an explicit choice, we tested whether the proportion who accepted the coin as an appropriate choice procedure, was larger in the *conflict* and *best option* condition compared with the original (base line) condition. Indeed, the proportion of participants who accepted the coin was much larger in the conflict condition than in the original condition (47.9% vs. 23.2%). This difference was statistically significant ($z = 2.64$, $p < .005$). The proportion of participants who accepted the coin in the best option condition (40.4%) was also significantly larger ($z = 1.88$; $p < .05$) than in the original condition.

Evidently, highlighting the conflict and suggesting that perhaps there are no other alternatives to the use of a randomizer, elevated the number of participants who accepted the coin as a suitable choice procedure. Notwith-

standing, in both of these latter conditions, there was a clear gap between the proportion who judged the arguments to carry equal weights and the corresponding proportion of participants who accepted the coin as a proper decision device. Further, although the objection to the use of a coin was mitigated in the two conditions, nonetheless more than 50% still refused to endorse the coin.

In the present experiment an attempt was made to reduce the resistance to the use of the coin by pointing out the lack of a better alternative. Yet, participants may still assume, even if erroneously, that by investing adequate effort one may find sufficient reasons to justify one or the other option. What would happen, however, if the two alternative options were made so similar that no sensible reasons would exist that can justify the choice of one option over the other. For instance, would a coin be more acceptable if the accident victims were two middle aged women (so that they are, as far as the physician is concerned, equal on all relevant dimensions). The following experiment was designed to test this possibility.

8 Experiment 7

8.1 Method

Participants. A total of 144 undergraduate students, recruited at the campus of Tilburg University participated in the experiment. They performed the task on a computer (laptop), along with several unrelated judgmental tasks, and were paid €5.00 (approximately $6.50) for their participation.

Design and procedure. Participants were exposed to two versions of the cover story employed in Experiment 1, presented separately, one after the other, on two different screens. One was the same cover story as employed in Experiment 1 except that the third physician (who was supposed to break the tie) was omitted. The other cover story was the same except that the young man and the old lady were replaced by two middle aged women, and no mention was made as to who was responsible for the accident. Hence, in this condition, it was impossible to come up with reasons that would favor one of the victims more than the other. At the end of each cover story, participants were told that the two physicians find it extremely difficult to decide which of the two victims to save and, due to the time pressure, decide to use a coin. Half of the participants were asked, (after each story separately) to judge on a 0–10 scale to what extent they thought the use of the coin was an *acceptable* decision procedure (0 completely unacceptable; 10 completely acceptable). The other half had to rate, on a similar scale, the *fairness* of the use of a coin (0 completely unfair; 10 completely fair). Within each of these two conditions (judgment of acceptability or fairness), the order of presenting the two

scenarios was counter-balanced: Half of the participants in each condition were first exposed to the original (old-lady/young-man) cover story and after rating it were exposed to the two middle-aged women scenario. The other half received the two scenarios in a reversed order.

It is important to note that, regardless of whether participants judged the coin acceptability or fairness, the first and second ratings were essentially different. Specifically, the ratings of the scenario presented first (either the old-lady/young-man or the two middle aged ladies) constitutes what Hsee (1996; Hsee, Loewenstein, Blount, & Bazerman, 1999) has termed *separate* ratings. Participants in this case are not yet familiar with the following scenario, and thus are unable to compare the two scenarios. When exposed to the second scenario, they are already familiar with the previous scenario which, in some respects, can serve as a reference point. Hence, the ratings on the second scenario can be conducted comparatively to the former ratings, a condition termed by Hsee as *joint* ratings.

8.2 Results and discussion

The results are portrayed in Table 5. The acceptability and fairness judgments were analyzed apart, and within each condition the first (separate) and second (joint) ratings, were also analyzed apart. Each analysis consists of a simple comparison based on an independent t-test.

Separate ratings. There was no reliable difference in rating coin acceptability for the old lady/ young man ($M = 3.31$) and the two middle aged ladies ($M = 3.03$) scenarios ($t = .367$; $p > .70$). Similarly, for the coin fairness ratings there was no reliable difference between the old-lady/young-man ($M = 2.84$) and the two middle aged ladies ($M = 3.51$) scenarios ($t = .894$; $p > .37$). It thus seems that participants did not distinguish between these two scenarios — use of the coin received a rather low rating on both the acceptability and the fairness scale for both scenarios.

Joint ratings. Coin acceptability was rated much higher in the two middle aged ladies ($M = 5.51$) compared with the old lady/young man ($M = 1.70$) scenario, a difference that was highly significant ($t = 5.56$; $p < .0001$). A similar difference was also observed for the coin fairness judgments: Fairness ratings of the middle aged ladies scenario ($M = 5.32$) was much larger than the comparable rating for the old-lady/young-man scenario ($M = 1.74$), a difference that was statistically reliable ($t = 4.91$; $p < .0001$).

Two main conclusions can be drawn from these results. First, in the separate presentation mode, when viewed in isolation (supposedly, the more likely situation), the coin is equally rejected (and judged unfair) for both scenarios. The fact that in the two middle aged ladies scenario, given

Table 5: Mean judgments (standard deviations) of acceptance and fairness of coin tossing in Experiment 7; 0 = entirely unacceptable (unfair); 10 = entirely acceptable (fair).

	Acceptance		Fairness	
	n	Mean (S.D)	*n*	Mean (S.D.)
First rating (separate)				
1. Young man vs. old lady	35	3.31 (3.34)	37	2.84 (3.23)
2. Two middle aged ladies	37	3.03 (3.30)	35	3.51 (3.19)
Second rating (joint)				
1. Young man vs. old lady	37	1.70 (1.87)	35	1.74 (2.16)
2. Two middle aged ladies	35	5.51 (3.71)	37	5.32 (3.76)

the information, there are no possible reasons to prefer one over the other did not reduce the aversion to using a coin. Second, there is a large difference in acceptability of the coin in the case of the second ratings which, supposedly are compared with the first ratings. Under such a comparative condition, the first rating provides a reference point to which the second can be compared. Participants' attention is thus directed to the difference between the two scenarios. Indeed, under these conditions the aversion to using the coin in the two middle age ladies scenario is drastically reduced. Nonetheless, it should be noticed that even in this case, the acceptability and fairness of the coin do not significantly exceed the middle (5.5) scale value. Evidently, the dislike of using a coin for vital decisions is deeply rooted, even when no reason exists for preferring one option to another.

9   Experiment 8

In all the preceding experiments, the coin has been used to represent a prototypical chance device. However, there may be features of this specific randomizer that makes it an unwanted tie breaker, at least in a medical scenario which supposedly is about saving lives and not about gambling. The present experiment was designed to investigate whether the manner in which the randomizer is described will influence its degree of acceptability. A coin toss belongs to a more general category of randomizers and can alternatively be called "a random device" or a procedure that offers "equal chances" to the parties involved. Such more general or more abstract descriptors may distract people's attention away from particular unattractive properties of the coin and ask for their endorsement of the principles involved rather than their attitudes toward one particular instantiation of these principles.

9.1   Method

Participants. A total of 152 students, recruited from a pool of participants (from different Dutch universities) who registered for taking part in ongoing experiments via e-mail, participated in the experiment. They were paid for this as well as other experiment at the rate of €1 per experiment.

Design and procedure. All participants were exposed to the basic scenario employed in Experiment 1. The scenario ended by informing participants that one of the physicians thought the old lady should be saved because the accident was not caused by her whereas the other physician thought the young man should be saved because he had a longer life expectancy. Participants were further assigned to four different conditions, three that involved a choice decision and one that involved judgment. Participants in the *Coin* condition were asked the following: Suppose you had to make a choice. Would you (i) save the old lady, (ii) save the young man, and (iii) toss a coin to determine who should be saved. This condition is in fact a replication of the corresponding condition in Experiment 1. Participants in the *Random device* condition were exposed to the same choice decision except that the third option was formulated as "use a random device to determine who should be saved." The purpose of this manipulation was to test whether people reject any random procedure or whether their objection is more directed toward the use of the specific random device, namely the coin. For participants in the *Equal Chance* condition, the third choice option was "give an equal chance for saving either of the two victims." The term "equal chance" is even more abstract than a random device; moreover, it directs attention towards the purpose of the selection procedure, rather than to the actual process or procedure being used. Finally, participants in the *Judgment* condition had to judge whether there were stronger reasons to save (i) the old lady, (ii) the young man, or whether (iii) the rea-

Table 6: Response distributions in Experiment 8 as a function of decision task.

	N	Old Lady	Young man	Equality
Decision — Coin	40	10	18	12 (30%)
Decision — Random device	36	7	5	24 (67%)
Decision — Same chance	37	4	4	29 (78%)
Judgment — Reasons	39	3	10	26 (67%)

sons to save each of the two were equally strong. This is a replication of the corresponding condition in Experiment 1. In each condition, the order of the first two options (i.e., old lady and young man) was counterbalanced.

9.2   Results and discussion

The results are shown in Table 6. As can be seen, the coin condition is the only one in which the majority is opposed to the use of a randomizer, namely the coin. Evidently, only 30% find the coin procedure acceptable. In contrast, when the procedure is described in more abstract terms, namely using a random device or granting the two victims equal chance, the majority (67% and 78% in the two conditions, respectively) expresses their approval. Finally, 67% of the participants in the judgment condition believe that the reasons to save each of the two victims are equally strong.

The results in the coin and the judgment conditions replicate those of Experiment 1. Indeed, the proportion of participants who thought that there were equally strong reasons for saving each of the two victims was significantly larger than those accepting the coin in the coin condition ($p < .001$ by a Fisher exact test). Further, the majority of the participants accepted the use of a randomizer when it was framed in abstract terms as using a randomizer or giving the two victims an equal chance, yet it was rejected by most participants when the randomizer was explicitly named as a coin. The proportion accepting the coin was significantly smaller than either the abstract randomizer formulation ($p = .0014$ by a Fisher exact test) or the equal chance formulation ($p = .00002$ by a Fisher exact test). This suggests that participants thought that a random procedure, that would give equal chance, was acceptable. They detest, however, the concrete procedure of a coin.

One possible explanation for this finding may be in terms of construal theory (Trope & Liberman, 2003) according to which we tend to overemphasize abstract, high-level goals and undermine the concrete, low-level steps needed to reach them. Thus, participants find the higher-level, abstract inspiration to obtain fairness appealing and thus endorse the statements that postulate (in

an unspecified way) "equal chance" and the "use of a random device". However, when it comes to the concrete achievement of fairness by throwing a coin, the potential drawbacks of that action are now at the center of attention. The shift from the abstract intention to ensure fairness to actually throw the coin is typical of the difference between thinking about the distant future and thinking about the near future. Whereas in the distant future we mainly consider the outcome attractiveness, in the near future, we think about the feasibility and immediate consequences of our decisions and actions.

10   Experiment 9

In Experiment 8 the (concrete) coin procedure was generally rejected, whereas two more abstract formulations were judged to be quite acceptable. However, the experiment does not allow us to conclude whether the coin is rejected because of its concreteness, or because of other potential defects. In Experiment 9 the coin procedure was compared to three other concrete procedures, the question being to which extent a coin toss is regarded as a fair and representative random procedure.

10.1   Method

Participants. The experiment was conducted with 242 participants. Ninety-three responded by e-mail and were recruited from the same pool of participants of Experiment 3 (only participants who did not take part in the previous experiment were recruited). The remaining 149 participants were recruited at the campus of the University of Nijmegen who performed the task on a computer (laptop), along with several unrelated judgmental tasks, and were paid €5.00 (approximately $ 6.50) for their participation.

Design and procedure. All participants were exposed to the basic scenario employed in Experiment 1. By the end of the scenario, participants were told that since the two physicians disagreed as to whom to save, and because the decision could not be deferred, they decided to use a random procedure. There were 4 possible methods: (i) Toss of a coin; (ii) A lottery in which a nurse will blindly

Table 7: Number of times each randomization procedure was ranked as first or second most preferred choice procedure, fairest procedure, and most random procedure, Experiment 9.

	1st rank	2nd rank
Preferred choice procedure		
Coin	28 (35%)	34 (42%)
Lottery	11 (14%)	24 (30%)
Room No.	18 (22%)	17 (21%)
Nurse	24 (29%)	6 (7%)
Fairest procedure		
Coin	32 (39%)	20 (25%)
Lottery	13 (16%)	36 (45%)
Room No.	15 (19%)	23 (28%)
Nurse	21 (26%)	2 (2%)
Most random procedure		
Coin	41 (51%)	26 (32%)
Lottery	19 (23.%)	36 (45%)
Room No.	13 (16%)	17 (21%)
Nurse	8 (10%)	2 (2%)

choose one of the names; (iii) Choosing randomly the room number in which the victim to be saved is located, and (iv) Asking the nurse to make the decision (which, indirectly, also constitutes a random procedure).

Participants were assigned to one of three conditions. Participants in the *choice* condition had to indicate their first and second preferred choice procedure. Participants in the *fairness* condition had to indicate which was the most and second most fair procedure. Participants in the *randomness* condition had to judge which procedure was in their opinion the most and second most random.

10.2   Results and discussion

The numbers of participants in each condition (choice, fairness, randomness) who ranked each of the four alternatives in the first and in the second place are portrayed in Table 7. A separate analysis was conducted for each of the three experimental conditions. Specifically, for each condition we applied a-priori contrasts to Cochran's Q test (Agresti, 1996) comparing the proportion of times the coin flip was ranked first with the other three procedures, and similarly to the proportion of times the coin flip was ranked first and second (combined) compared with the

other three procedures.[3]

In the preferred choice procedure condition, the coin toss was ranked more often as first than in the other procedures ($z = 1.817$; $p = .035$), and as first and second combined ($z = 4.70$; $p < .0001$). In the most fair procedure condition, the coin toss was ranked more often than the other procedures as first ($z = 3.015$; $p = .0013$), and as first and second combined ($z = 3.614$; $p < .001$). Finally, in the most random procedure condition, the coin toss was ranked more often than the other procedures as first ($z = 5.324$; $p < .001$), and as first and second combined ($z = 8.33$; $p < .0001$). In short, compared with other random procedures, the coin was clearly perceived as most fair and most random and, supposedly, was therefore also considered as the best procedure. It thus seems that most participants judge the coin as being the favorite procedure on all the three dimensions. In other words, conditional on making the choice random, the coin was judged as constituting the most preferred procedure.

The results of this experiment confirm that coin toss is regarded as a fair procedure, which is accepted as a better tie breaker than other, less transparent procedures. It is important to realize, as we already noted in the introduction, that fairness here relates to distributional fairness, that is equal allocation of chances. The reluctance to use a coin, which has been established by the preceding studies, is thus not attributable to deficiencies in distributional fairness but rather to a general resistance against any concrete randomization procedure. Apparently, randomizers do not satisfy the requirements of procedural fairness, even though participants admit, on a more abstract level (as demonstrated in Experiment 8), that the parties should have the same chance.

11   Experiment 10

Are there any conditions, under which the use of a concrete randomizer, specifically a coin, would be acceptable and thus considered adequate? As noted in the introduction, and further demonstrated by several studies in different domains of decision making, the perception of fairness is playing a central role in the decision process at both the individual and group level. Fairness, regardless of how exactly it is defined, is a prerequisite to an acceptable decision process. Because fairness and evenhandedness constitute the most prominent attribute of the

---

[3]Given that p_1 = proportion of times that the coin procedure was ranked first, and p_2, p_3, p_4, the proportion of times that the lottery, room number, and nurse, respectively, we test $H_0 : 3p_1 - p_2 - p_3 - p_4 = 0$ against $H_1: 3p_1 - p_2 - p_3 - p_4 > 0$. A similar test was conducted for the combined proportion of times a procedure was ranked first and second, where p_1, p_2, p_3, p_4, stand for the combined proportion that the coin, lottery, room number, and nurse, respectively, were ranked in the first and the second place.

Table 8: Mean acceptability ratings (standard deviations) of decisions in Experiment 10 depending on whether the physician was not familiar or familiar with one of the patients (0 = entirely unacceptable; 10 = entirely acceptable).

	N	Robert	John	Coin Toss
Not familiar	26	6.38 (1.96)	6.50 (1.68)	3.73 (2.98)
Familiar with:				
Robert	27	6.15 (1.56)	5.59 (1.93)	6.11 (3.11)
John	28	6.79 (1.85)	6.04 (1.91)	4.97 (3.15)

coin, it is surprising that in several of the experiments that employed the accident scenario, participants did not consider the coin as a fair solution (see especially Experiment 2 and the fairness ratings reported in some of other experiments). However, it is possible that the coin may become desirable after all when fairness is explicitly and visibly threatened. Experiment 10 was designed to test this conjecture.

11.1 Method

Participants. Eighty-one participants were recruited from the same pool of participants of Experiment 3 and 9 (only participants who did not take part in these previous experiments were recruited).

Design and procedure. Participants were exposed to the same kidney transplant cover story as in Experiment 3. Because in the previous experiment Robert (who was older but was said to have a higher chance of a successful transplantation) was slightly favored over John, a minor change was introduced, namely Robert's age was increased from 57 to 59 in order to make him somewhat less attractive compared with John. Participants were told that the physician on duty had three possible choices (allot the kidney to John, allot the kidney to Robert, or flip a coin) and were asked to rate the acceptability of each of the three options on a scale from 0 (completely unacceptable) to 10 (completely acceptable). One group (base-line) was exposed to the above condition that was in fact a replication of the choice condition in Experiment 3, except that participants had to rate the three different choicer alternatives, instead of choosing one of them. Participants in the other two (experimental) groups received the treatment as the base-line group except that the latter were given the following additional information: "In addition, as it turns out, the physician realizes that he knows John (Robert) who is a distant family member, in contrast to Robert (John) who is a complete stranger."

The only difference between the last two groups was that for one the physician was acquainted with John whereas for the other he was acquainted with Robert.

11.2 Results and discussion

The results are portrayed in Table 8. As expected, and compatible with the results of Experiment 3, in the baseline condition, ratings of allocating the kidney to either of the two patients (mean ratings for Robert and John were 6.38 and 6.50, respectively) was significantly higher ($p < .001$) compared with the coin toss option (mean rating of 3.73). In contrast, in the two experimental conditions there was no difference between the ratings of the allocation options (to one or the other patients) and tossing the coin. Most important, mean acceptability of the coin toss in the two experimental conditions in which the physician was acquainted with one of the patients (mean acceptability for the two conditions combined was 5.53), was significantly higher ($p < .01$) compared with the base line condition (mean acceptability 3.73). Thus, coin acceptability is largely increased when it is realized that the physician is unlikely to make an unbiased and fair decision.

Experiment 10 deliberately introduced a factor that explicitly violated the prerequisite of fairness. Under such conditions the coin (and probably other randomizers as well) is considered as an acceptable decision device. Note that the alternative option, namely a rational reason-based decision is unviable. Evidently, acquaintance of the physician (the decision maker) with one of the patients introduces a bias that cannot be repaired.[4] It is in this condition, where it is realized that a decision that would be both rational and fair is impossible, that the use of the coin is more appealing. Notwithstanding, it is remarkable that, despite the successful manipulation in this experiment, the endorsement of the coin remains limited and its average rating only slightly exceeds the midpoint acceptability of .5. The entrenched dislike for the coin seems not to be entirely eradicated.

---

[4]The only other solution is to get another physician who is not acquainted with either of the two patients, but this possibility is impractical in the particular scenario given the urgency needed in making a decision.

12   General discussion

The present paper demonstrated a deeply rooted aversion to the use of a coin as a decision device, in particular when the consequences are of high importance carrying a high weight. The dislike of the coin has two facets: First, even when people judge the reasons for the two options to be equally convincing and carry equal weights, they nevertheless refuse to choose the coin as a tie breaker, a finding referred to as judgment-choice inconsistency. Second, the proportions of participants who accepted the coin were low in absolute terms (usually not exceeding 30%). The phenomenon has been replicated in several experiments using two different cover stories. In both narratives, most participants judged the reasons for favoring one or the other protagonist (the old lady or young man in the accident scenario; one or the other patient in the kidney scenario) as equally compelling, yet the majority of participants in the actual choice condition refused to use a coin. The (combined) consistent experimental evidence leaves little doubt about the robustness of our findings.

In the introduction, we sketched four sets of circumstances that might affect people's willingness to use a randomizer. First, outcome seriousness. We found, as expected, less reluctance to use a coin in scenarios with mild to moderate consequences. In Experiment 4 the decision to use a coin for deciding disputes of authorship and choice of entertainment was approved by about the same number of participants as those who found the arguments for both positions equally persuasive. Thus in these cases no judgment-decision discrepancies was observed. However, introducing a probabilistic element in the accident scenario did not make the coin more acceptable (Experiment 5). Issues of life and death involve, even in the probabilistic case, what Baron and Spranca (1987) have termed "protected values". These are values linked to absolute moral obligations, they display tradeoff resistance, making compromises and informed decisions difficult. It is likely that random procedures are especially objectionable in areas where such sacred or protected values are involved — unless, of course, one is willing to consider the outcome of a chance event as a way of expressing the will of God (Elster, 1989; Ekeland, 1991).

Second, we suggested that randomizers might be easier to approve in cases where the indeterminacy of the situation is being emphasized and more rational solutions are ruled out. Experiment 6 showed that, when the dilemma is explicitly acknowledged, more participants are willing to use a coin. In Experiment 7, participants rated a random decision between similar victims more acceptable than a random choice between dissimilar victims, like in the original accident scenario. However, this difference emerged only in the joint condition, where participants were able to compare the two situations. Moreover, in both experiments, the acceptance of the coin remained generally low. Thus indeterminacy seems to be a necessary, but not a sufficient condition for approving decisions by chance.

Third, we explored how *fair* people perceive decisions by coin. To this question the experiments provide us with a seemingly puzzling set of answers. A coin flip, according to participants in Experiment 9, is the most fair of all the random procedures suggested (Table 7). Yet, life and death decisions by coin are absolutely not fair. When participants are explicitly asked what is a most fair decision (Experiment 2), or to rate the fairness of decisions by coin (Experiment 7), the emphasis on fairness does not make the coin more acceptable. Evidently, participants are caught in a conflict between distributional fairness (which the coin clearly satisfies, accounting for its high acceptability in Experiment 9), and procedural fairness which the coin does not fulfill as indicated by the results of the two main scenarios we employed in several experiments. The rejection of the coin in these experiments suggests that its rejection on grounds of procedural fairness outweighs its attractiveness in terms of distributional fairness. Failure of distributional fairness necessarily implies failure of procedural fairness but the reverse does not necessarily hold.

It is worthwhile to note that aversion to the use of a randomizer remains even when issues regarding distributional fairness are omitted. A recent (yet unpublished) experiment further supports the above analysis. In this experiment, a scenario was employed regarding a cancer patient, in which the physician is contemplating between two alternative treatments. After carefully weighing the pros and cons of the two treatments, the physician is unable to decide and consequently employs a coin. Note that in this case there is a single patient and thus distributional fairness is irrelevant. Compatible with our previous experiments, results of this experiment show yet again a profound aversion to the use of a coin, a dislike that cannot be attributed to distributional fairness.

Although procedural fairness is most likely the major factor underlying the aversion to the use of randomizers, distributional fairness may also play a pivotal role under certain circumstances as exemplified in Experiment 10. In contrast to the results of the previous experiments, the coin is suddenly approved as an acceptable solution in the last experiment, which suggested a danger of partiality (i.e., a risk for violation of distributional fairness) on the part of the decision maker. Thus people may think it is *not* fair to use a coin in situations where *proper* arguments for either option can be advanced; but it *is* fair in a situation where it is important to shield the decision from being influenced by *improper* arguments. In other words, the randomizer is not seen as an instrument promoting justice, but preventing injustice to prevail.

The last issue explored in our experiments was whether one randomizer is as good (or as bad) as any other. From a formal point of view, the answer to this question must, by definition, be yes, as long as the outcomes are truly independent and unbiased. However, as mentioned in the introduction, the criteria for randomness may be hard to specify and to test (Bar-Hillel & Wagenaar, 1993). In practice, people might have preferences for procedures where the random element is disguised (as in the penalty kick example), but when directly confronted with the question (in Experiment 9) they preferred the coin toss to the other suggested procedures, in respect to both randomness and fairness. Even more telling are perhaps the results of Experiment 8, showing that participants are actually prepared to embrace random procedures as long as they are abstractly described. Thus they seem to approve the principle, but hate the practice. In this respect they are a little like those of us who love humanity but detest people.

The present studies have demonstrated a robust and apparently deeply ingrained reluctance against a chance device like the coin for making an important decision. Yet, people seem to appreciate the usefulness of a random procedure, on an abstract level, to solve a problem of indeterminacy, but experience problems with applying a concrete instantiation of this principle, which seems only tolerable to protect the decision against unwanted partiality. Such ambivalence, in turn, leads to conditions of what Beattie et al. (1994) termed "decision aversion". They report an experiment in which participants had to imagine serving as a trustee for the estate of Mary who had two daughters. Mary's money and possessions were divided equally between the two daughters except for a priceless antique grand piano. Being in the role of the trustee, participants were asked which of three conditions they would have preferred to be in: (A) Someone you know would flip the coin to determine who will get the piano. (B) You decide which daughter gets the piano. (C) It does not matter. (A) and (B) are equally preferable, but were chosen only by a minority. The results showed that 63% of the participants opted for (A) and only 29% for (B), which the authors interpreted as an indication of decision aversion.

While being a viable interpretation, it is not in contradiction with the pattern of results presented here and is compatible with our explanations. There are several fundamental differences between the experiment of Beattie et al. and ours. First, the outcome (who will get the piano) is of intermediate importance and in this respect resembles more the first author scenario employed in Experiment 4. But, most important, the coin flip in Beattie et al.'s experiment is to be carried out by someone else. This implies that the decision maker transfers the responsibility not only to the coin but, at the same time, to an-

other person who flips the coin. In contrast, in both the accident and the kidney scenarios, one is confronted with the option of flipping the coin which implies on one hand surrendering control and responsibility and yet, after the outcome is known, one may still feel responsible (and regret) for flipping the coin that led to a particular outcome (and, legally be accountable for the outcome).[5]

The judgment-decision discrepancy observed in the present studies highlights the fact that judgments and choice decisions carry different implications and may not be based on exactly the same considerations. Judgments can be viewed as a predominantly cognitive process, whereas choices and decisions are actions. As such, they can have more serious consequences and are often (like in the present case) irreversible. They are also associated with a higher degree of commitment, responsibility, and occasional regret (Einhorn & Hogarth, 1981). Because of their behavioral consequences, decisions are more likely than judgments to be censured and sanctioned. For these reasons, it can be expected that people are more careful and strict in their choices than in their judgments (Ganzach, 1995).

To bridge the gap between judgments and decisions, people often rely on guidelines that can be described as explicit judgment-decision links. For instance, in criminal cases the rule is that the defendant should only be convicted if he is judged to be guilty "beyond reasonable doubt" (admittedly, as noted by Saunders and Genser [1999], this is an ambiguous criterion). Action rules typically advise people not to take chances, but to strive for some degree of conviction or certainty before executing their favored plans. Less often, explicit guidelines are offered for what action to take in case of judgments of indeterminacy. As we already proposed, using a randomizer is not incompatible with economic rational choice theory, but it is not explicitly prescribed, as people are assumed to always have preferences, however tiny ones. In contrast, game theory explicitly recommends coin flip in games with more than one equilibrium point (e.g., Rasmusen 1989). In more informal contexts, people find it more natural to respond to indeterminacy with inaction and deferral of choice.

The scenarios explored in the present experiments, describing emergency situations, did not allow for this option, which would have been the natural way out under other, less critical circumstances. They are accordingly faced with a choice between two unattractive strategies:

---

[5]Recently, one of us has submitted an article jointly written with another colleague. To determine the order of authors, a third colleague was approached and asked to toss the coin that will determine the order of authorship. Since this person knew well both authors she was hesitant to perform the task. Her solution was to ask yet another person to determine which author would be assigned to heads and tails, and she herself flipped the coin. In this way she was able to divide the responsibility of the coin's outcome.

Transferring the choice to a random device, or forcing themselves to form a preference for one of two equally justifiable options. The last strategy has the advantage of being rational in the sense of being governed by a continued search for reasons, the disadvantage being that this extended search will have to give some reasons more than their proper weight. The first strategy is more radical, by explicitly acknowledging the limits of reason-based rationality, and by finding both sets of reasons equally valid and equally strong cancelling them all. Decision by use of randomizer might therefore look like a sacrifice of reason, control, and even responsibility. To appear acceptable, higher principles must be invoked, like non-partiality and offering everybody the same chance.

The use of a coin for solving a decision problem can be regarded as an extreme example of assistance by a mechanical decision aid. It is well known that such procedures are viewed with suspicion by decision makers and their clients alike, even in areas where their superiority can be documented (Sieck & Arkes, 2005). In the case of the coin, no such superiority can be claimed. By definition, randomizers perform no better than chance. They cannot even claim consistency, one coin toss being uncorrelated with the next one. Added to this is the problem of responsibility and accountability. If we grant that the decision maker can be made morally responsible for an important choice (like which patient should be given priority), he or she will also be responsible for the decision to use a coin, which is an action that (as we have seen) under some conditions can be justifiable. But who is responsible for the *outcome* of the coin toss? Normally, one only thinks of *persons* as responsible agents. Moreover, people are traditionally only made accountable for consequences they could have controlled and foreseen (Fischer & Ravizza, 1998). This makes responsibility in the case of randomizers a very tricky issue and should be the subject of further studies.

We mentioned already that a major cause for rejecting randomizers is that from early childhood we are raised with the conviction that everything must have an explanation, this is the reason-based approach documented by Shafir et al. (1993). This belief is also represented in the judicial system in the often cited aphorism of Lord Hewart from Rex vs. Sussex: "It is not merely of some importance, but is of fundamental importance, that justice should not only be done, but should manifestly and undoubtedly be seen to be done."[6] Thus, while a coin, or any other randomizer, provides indisputable distribution fairness, it does not offer fairness that can be "manifestly seen". Most probably, it is because of these reasons that the coin is rejected on what we termed procedural grounds. Both the judicial system and lay people are re-

luctant to accept that under some circumstances no rational reasons can be unequivocally advanced in one way or another. It is because of this boundless belief that "God does not play dice" and that, therefore, every decision can and should be backed by some reason, that Elster (1989) concluded that "rather than accept the limits of reason, we prefer the rituals of reason" (p. 37).

Some of the results reported here, specifically the strong and deep rooted repulsion to use of randomizers as decision aids, may seem to some readers as not very surprising.[7] Yet, they have important theoretical and practical implications. One concerns a limitation of choice theory, normative or descriptive, in dealing with situations of indifference and indeterminacy. Despite the ever increasing literature on behavioral choice theory, there are hardly any studies that directly investigate authentic situations of indeterminacy, in particular when the major relevant dimensions of the choice set are non-compensatory. The present paper may offer a starting point for a more rigorous examination of situations of indeterminacy.

Second, the present paper is closely linked to Bandura's (1982) assertion regarding the psychology of chance encounters and life paths. Bandura's pointed out that the more important and meaningful events in our life (e.g., choosing a partner, developing a career) are at the end of the analysis a chance event even though we do not perceive them as such. As he noted, separate chains of events in a chance encounter may have their own causal determinants yet their interaction which lead to the final outcome are determined by a random rather than a pre-designed process. In line with people's denial concerning the weight of the random component in determining life paths, supposedly because they are unaware of hidden random processes, the present experiments demonstrate an active and deliberate (i.e., conscious) attempt to resist the entry of such chance events into important life events.

Finally, our experiments may also carry some practical implications. We believe that in the situations we posed to our participants, specifically the two major scenarios employed in our experiments, the use of a randomizer is the most sensible action to use. In particular, a coin is a fast and efficient decision device, certainly in the particular scenarios we employed in which procrastinations may result in the worst outcome (i.e., both victims or both patients will die). Unlike human decision makers, appropriate random procedures (like a fair coin) are not contaminated (Wilson & Brekke, 1994) by unwarranted biases that may interfere in the decision process in an undesired manner. The coin is not only fair (in terms of distributional fairness) but, not less important, is perceived as such. Finally, a coin (or any other randomizer) may be the most sound choice procedure, in cases where the differ-

---

[6]Rex v Sussex Justices; Ex parte McCarthy [1924] 1 KB 256 at 259.

[7]Even if this is the case, it is highly questionable whether surprise of empirical results should serve as an indication for their importance.

ent dimensions are non-compensatory (e.g., Luce et al., 1999), there are no decisive clear considerations of how the different dimensions should be weighed, and most important, the outcome is indivisible. Unless one adopts Solomon's judgment, the decision is a binary one with no possibilities in between. Put in other words, perhaps we should realize that, under some conditions, we should accept the limits of rational choice theory, and be open for a procedure that is efficient, fair, and defendable if we use a broad definition of what is meant by rationality.

References

Agresti, A. (1996). *An introduction to categorical data analysis.* New York: Wiley.

Ashenfelter, D. (2003, June 21). Judge disciplined for coin flip. *Detroit Free Press.*

Bandura, A. (1982). The psychology of chance encounters and life paths. *American Psychologist, 37,* 747–755.

Bandura, A. (1998). Explorations of fortuitous determinants of life paths. *Psychological Inquiry, 9,* 95–99.

Bar-Hillel, M. & Wagenaar, W. A. (1993). The perception of randomness. In G. Keren & C. Lewis (Eds.), *A handbook for data analysis in the behavioral sciences: Methodological issues* (pp. 369–394). Hillsdale, NJ: Lawrence Erlbaum.

Baron, J. & Spranca, M. (1987). Protected values. *Organizational Behavior and Human Decision Processes, 70,* 1–16.

Beattie, J., Baron, J., Hershey, J. C. & Spranca, M. D. (1994). Psychological determinants of decision attitude. *Journal of Behavioral Decision Making, 7,* 129–144.

Broome, J. R. (1984). Selecting people randomly. *Ethics, 95,* 38–55.

Einhorn, H. J. & Hogarth, R. M. (1981). Behavioral decision theory: Processes of judgment and choice. *Annual Review of Psychology, 32,* 53–88

Ekeland, I. (1991). *The broken dice.* Chicago: The University of Chicago Press.

Elster, J. (1989). *Solomonic judgments: Studies in the limitations of rationality.* Cambridge: Cambridge University Press.

Falk, R. (1991). Randomness — an ill-defined, but much needed concept. *Journal of Behavioral Decision Making, 4,* 215–218.

Falk, R. & Konold, C. (1997). Making sense of randomness: Implicit encoding as a basis for judgment. *Psychological Review, 104,* 301–318.

Fischer, J. M. & Ravizza, M. (1998). *Responsibility and control: A theory of moral responsibility.* Cambridge: Cambridge University Press.

Ganzach, Y. (1995). Attribute scatter and decision outcome: Judgment versus choice. *Organizational Behavior and Human Decision Processes, 62,* 113–122.

Hsee, C. K. (1996). The evaluability hypothesis: an explanation for preference reversals between joint and separate evaluations of alternatives. *Organizational Behavior and Human Decision Processes, 67,* 247–257.

Hsee, C. K., Loewenstein, G. F., Blount, S., & Bazerman, M. H. (1999). Preference reversals between joint and separate evaluations of options: a review and theoretical analysis. *Psychological Bulletin, 125,* 576–590.

Luce, M. F., Payne, J.W. & Bettman, J.R. (1999). Emotional tradeoff difficulty and choice. *Journal of Marketing Research, 36,* 43–159.

Neuringer, A. (1986). Can people behave "randomly"? The role of feedback. *Journal of Experimental Psychology: General, 115,* 62–75.

Oberholzer-Gee, F., Bohnet, I. & Frey, B.S. (1997). Fairness and competence in democratic decisions. *Public Choice, 91,* 89–105.

Pashley, P.J. (1993). On generating random sequences. In G. Keren & C. Lewis (Eds.). *A handbook for data analysis in the behavioral sciences: Methodological issues.* Hillsdale, NJ: Lawrence Erlbaum.

Rasmusen, E. (1989) *Games and Information: An Introduction to Game Theory.* Cambridge: Basil Blackwell.

Resnik, J. (1985). Precluding appeals. *Cornell Law Review, 70,* 1–16.

Saunders, H. D. and Genser, J. G. (1999) Trial and error. *The Sciences,* September/October issue.

Shafir, E., Simonson, I. & Tversky, A. (1993). Reason-based choice. *Cognition, 49,* 11–36.

Sieck, W. R. & Arkes, H. R. (2005). The recalcitrance of overconfidence and its contribution to decision aid neglect. *Journal of Behavioral Decision Making, 18,* 29–53.

Skinner, E. A. (1996). A guide to constructs of control. *Journal of Personality and Social Psychology, 71,* 549–570.

Tetlock, P. E. (1992). The impact of accountability on judgment and choice: Toward a social contingency model. In M.P. Zanna (Ed.), *Advances in Experimental Social Psychology, 24* (pp. 331–376). New York: Academic Press.

Trope, Y. & Liberman, N. (2003). Temporal construal. *Psychological Review, 110,* 403–421.

Vallacher, R. R. & Wegner, D. M. (1987). What do people think they are doing? Action identification and human behavior. *Psychological Review, 94,* 3–15.

Wilson, T. D., & Brekke, N. (1994) Mental contamination and mental correction: Unwanted influences on judgments and evaluations. *Psychological Bulletin, 116,* 117–142.