

A DILEMMAS TASK FOR ELICITING RISK PROPENSITY

Juan Botella, María Narváez, Agustín Martínez-Molina,
Víctor J. Rubio, and José Santacreu
Autónoma University of Madrid, Madrid, Spain

Risk propensity (RP) is a trait characterized by an increased probability of engaging in behaviors that have some potential danger or harm but also provide an opportunity for some benefit. In the present study, a new RP task with several dilemmas was explored. Each dilemma includes the initial set plus successive approximations for estimating the Indifference Value between a Secure and the expected value of an uncertain Game. The scores showed good internal consistency, reasonably test-retest reliability, and good validity. The risk propensity dilemmas (RPD) task is proposed as an addition to an ideal battery of tasks for reflecting a complex trait that should be assessed from varied perspectives, procedures, and scenarios of uncertainty.

The goal of the present research is to explore the feasibility of a new tool for measuring risk propensity (RP). This trait is characterized by a willingness to take risks that ultimately is reflected in an increased probability of engaging in risk-taking behaviors. Risk-taking behaviors are those that involve some potential danger or harm but also provide an opportunity to obtain some benefit (Leigh, 1999). Despite its importance, RP has been traditionally difficult to define and measure as a construct, although some attempts have been made from very different perspectives (Harrison, Young, Butow, Salkeld, & Solomon, 2005). Not only is it important to assess RP in persons who have to make decisions of high responsibility and potential danger, but the factors that have an effect on decision making must also be investigated. One question of interest is why a person in a given context chooses risk-taking behavior. Another is whether some people have a relatively general and enduring disposition that produces a higher probability of their being involved in risky behaviors in varied contexts. In exploring this issue, experimenters must answer the following question: Does a risky behavior reflect a risk propensity for that individual, or is it only a specific response elicited by a situational condition? In other words, do individuals demonstrate consistent risk-taking tendencies across situations, or is risk-taking behavior determined by situational or content-

The present research received financial support from the project AENA-UAM/785001. We are grateful to Jim Juola and two anonymous reviewers for their helpful comments.

Correspondence concerning this article should be addressed to Juan Botella, Facultad de Psicología, Universidad Autónoma de Madrid, Ciudad Universitaria de Cantoblanco, c/ Ivan Pavlov, 6, 28049 Madrid, Spain. E-mail: juan.botella@uam.es

domain factors? Even though the controversy is still alive, empirical research has shown that some personality factors are related to risk-taking behaviors (see, for instance, Skeel, Neudecker, Pilarski, & Pytlak, 2006). This evidence supports the hypothesis that a personality dimension is involved in the probability that the individual chooses risky or conservative alternatives when managing a specific situation; nevertheless, the capacity for predicting risk-taking behavior remains weak (Weber, Blais, & Betz, 2002).

The main motivation for this line of research is a need for a variety of tasks that map the crucial facets of risk taking that should be sampled for a valid measure of the construct. In the present experiment, a new task is proposed for reflecting the RP with specific features not shared with other tasks. Specifically, the task was to estimate for several dilemmas the expected value for an uncertain course of action that would be subjectively assessed as equally attractive as a course of action with a secured result. Sensitivity of the task to factors with well-known effects on risk taking were also studied. However, this task is not proposed as a replacement for other tasks or considered a better measure of risk taking. Instead, the proposal is to add this task to the catalogue of tasks available for the study of RP.

Approaches to the Assessment of Risk Propensity

Several approaches have already been taken for RP assessment. First, measures of several personality traits have been employed that supposedly reflect RP, either because they have shown significant correlations with RP or because risk taking appears as a central element in the definition of the construct. Among the most often employed are sensation seeking (Zuckerman, 1994) and Impulsivity (see Vigil-Colet, 2007). Sometimes, the Eysenck Personality Questionnaire (EPQ; Eysenck & Eysenck, 1998) is employed because it is assumed that sensation seeking forms part of the extraversion-introversion trait and impulsivity forms part of the psychoticism trait. The main limitation of this approach is that the scores necessarily reflect a mixture of tendencies that will be different from RP because the dimension is inferred from measures of more general characteristics.

Second, self-reports of one's own past and present behavior have been employed to assess RP. In the Risk Propensity Scale (Nicholson, Soane, Fenton-O'Creevy, & William, 2004), the subject uses a Likert scale to rate the frequency with which several everyday risky experiences describe them (both now and in the "adult past"). The RP is estimated from the frequency of risky behaviors reported for the situations presented. Its main limitation, common to all self-reports, is that the individual's honest cooperation is assumed. Unfortunately, people are not always truthful, sometimes because the respondents try to hide their tendencies, but also because of simple deficits in the ability for self-monitoring.

Third, tests have been developed in which simulated uncertain situations are proposed to individual participants that are intended to elicit their RP. Some of these situations are related to specific domains, such as medical decisions (e.g., Grol, Whitefield, De Maeseneer, & Mokkink, 1990) or to business and financial domains (e.g., Sitkin & Weingart, 1995). Others combine scenarios from several different domains, perhaps allowing a better measurement of a general RP (e.g., the Choice Dilemma Questionnaire, CDQ, from Kogan & Wallach, 1964; or the Domain-Specific Risk-Attitude scale, DOSPERT, from

Weber, Blais & Betz, 2002). The formats for the responses are also very varied. Some are Likert scales to assess the likelihood of engaging in a behavior; in others, the individual must estimate the probability of success with which they would engage in a risky behavior; and in still others, they must choose between lotteries with different characteristics.

Finally, in some tests an objective assessment of personality is attempted (Rubio, Santacreu, & Hernández, 2004). The individuals are presented with a task in which they must make choices that supposedly reflect their RP. Among them are the Balloon Analogue Risk Task (BART; Lejuez et al., 2002) and the Betting Dice Test (BDT; Arend, Botella, Contreras, Hernández, & Santacreu, 2003; Rubio, Santacreu, & Hernández, 2006; Santacreu, Rubio, & Hernández, 2006).

In the BART (Lejuez et al., 2002), the participant can inflate balloons presented on the screen by clicking the mouse on the pump to introduce air. However, each balloon has a limited and unknown air capacity; when the number of pumps exceeds that limit, the balloon explodes. The individual earns 5 cents for each pump; but if the limit is exceeded and the balloon explodes, all the money accumulated for that balloon is lost. The average number of pumps before stopping reflects the level of RP.

In the BDT (Arend et al., 2003), several bets are offered before the simultaneous toss of a virtual pair of dice (a demo is available at www.uam.es/proyectos/investigacion/psimasd/dadosen/pr05nv10.swf). The alternative bets are of the type “even” (versus odd), “more than 4” (versus 4 or less), “more than 7” (versus 7 or less), or “12” (versus less than 12). The number of points earned when winning with each alternative bet is such that the expected value is the same on each bet. In the “even” (versus odd) bet the player wins only 2 points, because the probability of winning is 1 in 2. However, in the bet of “12” (versus less than 12), the player wins 36 points because the probability of winning is 1 in 36. The preference for a bet as “12” (versus less than 12) is considered more risky than a bet as “even” (versus odd), given that the probability of winning is lower but the amount when winning is higher.

Both the BART and the BDT are natural and valid measures of RP, given that the propensity shows up directly and these tasks don't have some of the limitations of self-reports. For instance, the BART has shown low but significant correlations with Zuckerman's Sensation Seeking Scale and some impulsivity measures (Lejuez et al., 2002), as well as several self-reported risk behaviors such as alcohol and cigarettes consumption (Bechara, 2003; Lejuez et al., 2002), and self-reported rebellious behaviors such as underage drinking (Skeel, Neudecker, Pilarski, & Pytlak, 2006). On the other hand, the BDT has shown satisfactory internal consistency (Cronbach's $\alpha = .827$) and temporal stability (1-year test-retest $r = .430$) (Arend et al., 2003; Rubio et al., 2006). It has also proven its validity in using different criteria, such as some other risk tasks (Arend et al., 2003; Rubio et al., 2006), the risk-taking behavior outcomes of the performance in an *ab initio* air traffic control training course (Santacreu et al., 2006), or the guessing tendency shown in a multiple choice test (Rubio, Hernández, Zaldívar, Márquez, & Santacreu, submitted).

The Risk Propensity Dilemmas (RPD) Task

The task described here has several features that make it a good tool for capturing general RP. It can be classified within the third group of the previous

section, since the participants are presented with a number of dilemmas and they must choose between two courses of action. The most salient and distinctive feature is the format of the responses. It was inspired by other experimental procedures proposed for estimating a judged indifference point, although the adjustment is done on the magnitudes, instead of the probabilities (Lichtenstein & Slovic, 1971). As in the parameter estimation by sequential testing procedure (see Bostic, Herrnstein, & Luce, 1990), several proposals are made and the amounts offered in each one change as a function of the choices made in the previous proposals.

Specifically, several proposals are presented for each dilemma; the goal for each dilemma is to obtain an estimate of what we will call the Indifference Value (IV). The first step in exploring the task was a pilot study with five evaluation dilemmas (which we call here the "basic dilemmas") sharing a common structure designed to estimate the IV. In the version applied here, several variations of those dilemmas have been incorporated. The rationale and structure of the basic dilemmas will be described before the variations of the version employed in the present research. Figure 1 shows the first screen of one dilemma. The participant must choose between A or B (see www.uam.es/proyectosinv/psimasd/dilemasen/ap06nv11.swf for a demo).

You are participating in a TV contest where you can win money answering questions. In any moment you can stop playing and keep the money won. On any question you earn more money giving the correct answer or lose money giving an incorrect answer. In this moment you can stop and keep the **1000** euros won until now or can answer the last question proposed.

You have no idea of the answer, so that in fact you have a probability of .50 of giving the correct answer. If you hit, your final wins total **1500** euros; however, if you miss the question your final wins total **500** euros.

YOU MUST CHOOSE BETWEEN:

(A) Stop and keep the **1000** euros won until now

(B) Answer the question and your final wins could total **1500** euros or **500** euros

A

B

Figure 1. First screen for one of the basic dilemmas. The amounts in bold change as the participant makes his/her choices. Option A corresponds to the Secure (S) and remains unchanged along the item. Option B corresponds to the Game (G), for which the amounts change along the item to estimate the IV (see the text).

In each basic dilemma, a situation is proposed to the individual in which he or she must choose between either a course of action with a Secure result (S; "A" in Figure 1) or a course of action that conveys a situation of uncertainty called the Game (G; "B" in Figure 1). The G has two possible and equally likely results; in the first situation proposed for each dilemma, G's expected value equals S (see the first proposal in the left column of Figure 2).

1st prop.	1st R	2nd prop.	2nd R	3rd prop.	3rd R	Estim IV	
$S[X] \rightarrow 1000$ $G \begin{cases} [Y1] 1500 \\ [Y2] 500 \end{cases} \rightarrow EV = 1000$	S	$S[X] \rightarrow 1000$ $G \begin{cases} [Y1] 1750 \\ [Y2] 1250 \end{cases} \rightarrow EV = 1250$	S	$S[X] \rightarrow 1000$ $G \begin{cases} [Y1] 1875 \\ [Y2] 875 \end{cases} \rightarrow EV = 1375$	S	1437,5	
			G	$S[X] \rightarrow 1000$ $G \begin{cases} [Y1] 1625 \\ [Y2] 625 \end{cases} \rightarrow EV = 1125$	G	1312,5	
		G	$S[X] \rightarrow 1000$ $G \begin{cases} [Y1] 1250 \\ [Y2] 250 \end{cases} \rightarrow EV = 750$	S	$S[X] \rightarrow 1000$ $G \begin{cases} [Y1] 1375 \\ [Y2] 375 \end{cases} \rightarrow EV = 875$	S	1187,5
				G	$S[X] \rightarrow 1000$ $G \begin{cases} [Y1] 1125 \\ [Y2] 125 \end{cases} \rightarrow EV = 625$	G	1062,5
	S	S	937,5				
	G	812,5					
	S	687,5					
	G	562,5					

Figure 2. The 8 possible sequences of responses of a participant in the five basic dilemmas, illustrated with Dilemma 5 (see the text for a detailed explanation). See Figure 1 for the instructions shown in the first screen. The amount conveyed in the Secure is represented by X, whereas the amounts involved in the Game are represented as Y1 and Y2.

In the basic dilemmas the amplitude of the interval that forms the two values involved in G also equals the amount involved in S. That is, the alternatives convey the amounts $S + (S/2)$ and $S - (S/2)$. After the first choice (S or G), the same dilemma is proposed with the same value on S, but with a pair of values in G that have an expected value different from S. Specifically, if the individual chose S in the first proposal, then the pair of amounts associated for G in the second proposal, $[(S + S \cdot (3/4))$ and $[S - S \cdot (1/4)]$, have a higher expected value $[S + S \cdot (1/4)]$, so that the attractiveness of G is increased. If in the second proposal the S is again chosen, in the third proposal the values of G, $[S + S \cdot (7/8)]$ and $[S - S \cdot (1/8)]$, have an expected value even more attractive — $[S + S \cdot (3/8)]$. The same happens when the decision-maker chooses G in the first proposal, although it is achieved through reductions in the expected value. In this way, the Indifference Value (V) is estimated through successive approximations. The IV is defined relative to the expected value for G with which it has the same attractiveness as S. The RP score on each dilemma is obtained by dividing the estimated IV by S. In this way, values lower than 1 reflect a high RP, and values higher than 1 reflect a low RP (see Figure 2). In each dilemma, three proposals are made. In the first one the amounts are the same for all individuals, whereas the amounts of the second are a function of the response the individual gave to the first; and those of the third are a function of the responses given to

the first and second proposals (as a cumulative tree).¹ Figure 2 details how this structure is applied to one of the five basic dilemmas.

The five dilemmas version used in the pilot showed good performance with regard to internal consistency (Cronbach's alpha: .767). Furthermore, it reached a significant, although moderate, correlation of -.29 with another criterion task also assessing RP. This criterion task is a computerized test in which an investment game is proposed, with several virtual participants represented in the screen. Some amount of virtual money is assigned to each individual, which they may use as they wish. They can invest all or part of it in a *variable-rate investment common fund*, where the interest rate depends on the behavior of the virtual partners. They can also keep all or part of the money in their *fixed-rate personal fund*, which guarantees some small profit for the participant. Keeping all or most of the money in the fixed-rate personal fund is considered a conservative behavior. Investing all or most money in the common fund is considered a more risky behavior because of the variations in the interest rate and also because of the uncertainty on the profits, given that the behavior of the virtual participants is unpredictable. The level of risk propensity is assessed by taking as a basis the proportion of money that the participants invest in the common fund.

Even though the results behaved well in the pilot study and the response profiles appear to be useful and sensitive, the experimenters believe that the task can be improved in several ways. Such an improvement is the specific motivation for developing the present version and for making it publicly available. The specific goal of this study is to demonstrate and assess the sensitivity of the task to the factors to be described next. The potential sources of improvement that were explored came from the following changes:

(a) *Variety of domains*. A general RP trait can be better mapped if a large number of (realistic) domains are included in the task designed to elicit it. That is why the present version includes dilemmas from a broader number of different areas than the previous version, although they are mostly within economic decisions. Nevertheless, the structure is easily applied to dilemmas within any context for which a selection process is specifically designed. This same structure can be used in selection processes for medical positions, sports, and others if the contexts of the dilemmas of the version in the present study are replaced by situations more related to the specific context of the selection. For example, the classical health problems of Tversky and Kahneman (1981) can be easily adapted to this format. An example for sports contexts is the F1 drivers' decision on when to stop to change the tires. Changing them immediately has well-known consequences in terms of delay. The consequences of delaying the stop are uncertain, but they can be positive or negative depending on how much the tires degrade in the next laps.

(b) *Framing the dilemmas*. It is well known that the willingness to take risks is differently elicited by the frames of gains or losses (Kahneman & Tversky,

1 Notice that the amplitude of the interval cannot be larger than S. Otherwise, in the final proposal, at the end of the sequence, it would be possible for S to be higher than the upper amount of G. In that situation, it would be nonsensical to make the question, given that it would be irrational choosing G when it is impossible to win more than with S. The same would happen if the lower amount in a proposal for G is higher than S; it would be nonsense to keep with S if any result from the game would be better than S. In the five basic dilemmas, the first proposal conveyed that Amplitude = S. In this way, the range of the individual ratios IV:S in a dilemma is [0.5; 1.5].

1979; Kuhberger, 1998). Other behavioral tasks have shown to be sensitive to the framing (e.g., the BART, see Benjamin & Robbins, 2007). Dilemmas with both types of frames were included here to assess their differential contributions to the general trait under study and to maintain the internal consistency of the scores.

(c) *Amplitude of the intervals.* In all the dilemmas of the pilot study, the amplitude of the interval with the two values of G equals the amplitude in S. The degree to which the dilemma elicits any RP could be a function of the amplitude of the interval employed. That is, if the amount in S is 1000 and the amounts in G are very close (e.g., 999 and 1001 for an extreme case), the individual could be less involved in the task since the courses of action have very similar attractiveness as assessed from an RP perspective. In the present version, two different amplitudes are used (the difference between the upper and the lower alternatives in G could be S or S/2) to evaluate the sensitivity of the responses to this factor.

Method

Participants

The sample was made up of 892 candidates for a training course for a high-level technical position. A degree from a university was the prerequisite to take part in the selection process. All the candidates were university graduates from several academic branches: **engineering, natural sciences, social sciences, and humanities.** Of the total sample, 612 were men (68.6 %) and 280 were women (31.4%). The men's mean age was 29.6 years ($SD = 3.9$) and the women's was 28.4 ($SD = 3.4$). During the selection process, 65% of the sample reported having another job. All participants signed a document in which they acknowledged the informed consent for the research side of the selection process, accepting that their data could be employed for new developments and refinements of the tests. They could also ask that their records be deleted after the selection process, so that their data could be omitted from any research activity.

Apparatus and Materials

Two tests for assessing RP were used: the RPD task and the BDT. The BDT consists of betting on one out of four possibilities (see Figure 3): More than 4 (from 5 to 12, $p = 30/36 = .83$); More than 7 (from 8 to 12, $p = 15/36 = .42$); More than 9 (from 10 to 12, $p = 6/36 = .17$); and a straight bet (number 12, $p = 1/36 = .03$). Individuals were told that each one is associated with different amounts of reward: 1 point, 2 points, 5 points, and 30 points, respectively. They are encouraged to get as many points as they can. In this study, participants had ten 20-second-trials. It was understood that they gathered points from trial to trial. In each one, participants had to make their bets, but they were not informed of the number of trials nor about the results of their bets. Each trial finished with a standard message such as "OK. And now what is your bet?" If someone did not bet in the 20-second period of the trial, the message "*You have bet on no one. You have gathered no points*" appeared. At the end, the message "The task has finished" was shown. As can be verified, the expected values of the four options are the same, so the assumption is that the individuals who bet

on More than 4 are making a more conservative choice than those who choose the straight bet option.

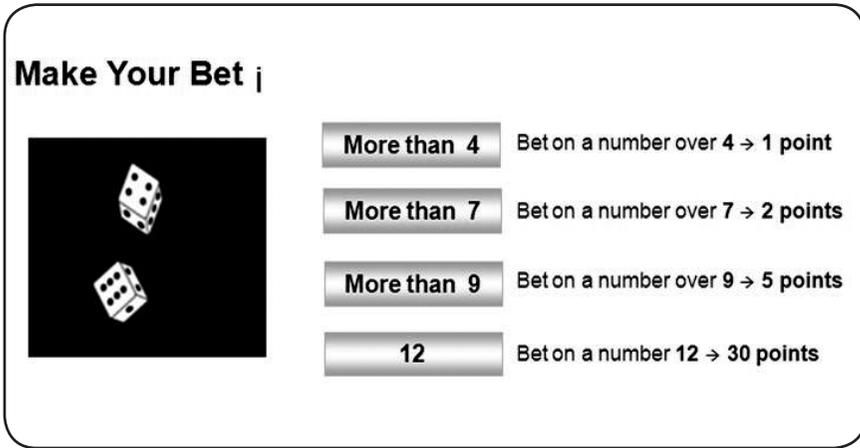


Figure 3. One of the screens employed in the BDT for giving the instructions to the participants (see the text).

As is usual in the study of risk-taking behavior, the risk value of each alternative is defined as the inverse of the probability of getting it correct. In this task, the individual's risk-taking score is calculated as the average of the natural logarithm of the inverse of the probabilities of his or her choices (1.2; 2.4; 6; 36, respectively) as follows:

$$\text{BDT Risk Score} = \frac{\sum_{i=1}^{10} \ln(1/p_i)}{10} \quad [1]$$

Natural logarithms give a closer index of the individual's risk propensity when he or she chooses one alternative out of several (see Sante & Santacreu, 2001). The higher the score, the greater the risk is assumed to be.

The tasks were administered in a large room, with individual workplaces equipped with PC-compatible computers and Philips 107-s, 17-in. monitors.

Procedure

The task was used ostensibly as one of the selection process of candidates for a high-level technical position, together with other tests of personality and cognitive abilities. Each participant responded to 13 dilemmas: 1 learning dilemma and 12 assessment dilemmas. The 12 assessment dilemmas were prepared from 5 basic dilemmas, which were given some variations. These variations included manipulations of the interval's amplitude (S versus S/2) and the decision frame (gains vs. losses). Table 1 show their features, and the Appendix gives the translation from Spanish of the specific instructions for each variation. Each participant responded to 13 dilemmas. The three successive questions allowed an estimate of the IV of each participant for each dilemma. The dilemmas were administered in two different orders. In the first order, the intervals with large amplitude were grouped first, replicating the format employed in the pilot version; in the second order, the dilemmas with large and

short amplitudes alternated, so that any carryover effects could be detected of administering the large-amplitude interval dilemmas over the responses to the short interval. Since comparisons between the results from both versions showed no order differences, the data from all the participants are collapsed for the statistical analyses. As a result, the number of each dilemma in Table 1 does not reflect the order in which it was administered. Details of the instructions given to the candidates and the procedure are also available at the demo (www.uam.es/proyectosinv/psimasd/dilemasen/ap06nv11.swf).

Table 1
Features of the 13 Dilemmas Employed: Context, Interval, and Decision Frame

Context	Interval	Frame	Factor 1	Factor 2
1. Driving to work (learning dilemma)	S	Gains	—	—
2. Sale opportunity (for a seller)	S	Gains	.689	-.218
3. Financial (actions sale)	S	Gains	.819	-.031
4. Salary (change of job)	S	Gains	.461	.822
5. Questions and answers (TV contest)	S	Gains	.828	-.088
6. Promotional premium in a store	S	Gains	.814	-.093
7. Financial (as Dilemma 3)	S	Losses	.647	-.063
8. Sale opportunity (as Dilemma 2)	S/2	Gains	.726	-.199
9. Financial (as Dilemma 3)	S/2	Gains	.792	-.092
10. Salary (as Dilemma 4)	S/2	Gains	.459	.824
11. Contest (as Dilemma 5)	S/2	Gains	.819	-.096
12. Promotional premium (as Dilemma 6)	S/2	Gains	.802	-.069
13. Financial (as Dilemmas 3, 7 and 9)	S/2	Losses	.617	-.065

Note. Interval refers to the relationship between the difference between the two amounts involved and the secure, S. The last two columns show the correlations between the 12 dilemmas of evaluation and the two factors extracted.

Results

In general, the participants responded in a conservative way. In fact, the average of the individual ratios IV:S for the 12 dilemmas are above 1 (all average IVs are higher than S; see Table 2). That is, on the average, in all dilemmas the individuals need an expected value for the Game higher than the Secure for being equally attractive.

For purposes of statistical analysis, the data were transformed into a metric that allows direct comparisons. In the short-amplitude dilemmas the range of the IV:S ratios is [0.75 - 1.25], whereas in the large-amplitude dilemmas it is [0.50 - 1.50]. Moreover, if in a losses dilemma the participant chose first the Secure, then the Game must be more attractive for the second proposal. But this must be done by reducing the absolute value of the expectation (even to being negative). The metric employed is as follows. The sequence of responses "Secure-Secure-Secure" (SSS) is the most conservative, whereas the second more conservative is SSG; finally, and following the same line, the GGG sequence is the most risky. With the value 1 given to the first one, 2 to the second, and so on until the value 8 for the last one, comparisons can be made free from the boundaries set by the specific numbers employed in the proposals. Table 2 shows the averages for the 12 dilemmas on this order score. All the statistical analyses that follow have been done for the order scores.

Table 2
Averages And t-Tests of Means Differences Between Selected Pairs of Dilemmas, both for the IV:S Ratio Scores and for the Order Transformed Data

Context		IV:S Ratio Scores			Order Scores		
		Amplitude		T	Amplitude		t
		Large	Short		Large	Short	
Sale opportunity	(D2,D8)	1.2270	1.0891	34.95	2.6839	3.0751	10.92
Financial	(D3,D9)	1.2112	1.0869	33.72	2.8105	3.1099	8.43
Salary and job	(D4,D10)	1.1334	1.0278	24.40	3.4327	4.0549	17.02
TV contest	(D5,D11)	1.2040	1.0849	37.23	2.8677	3.1413	9.20
Promotion	(D6,D12)	1.1558	1.0551	27.78	3.2534	3.6188	12.99
Financial	(D7,D13)	1.1155	1.0490	13.56	3.5762	3.7164	3.078

Note. The pairs compared shared the same context but employ different interval amplitudes. The last comparison involves losses dilemmas. In all cases the direction of the difference reflects more risky decisions for short (versus large) intervals and for losses (versus gains) frames. Degrees of freedom are 891 for all comparisons; $p < .001$ for all comparisons except for the last one ($p = .002$).

The effects of the two factors manipulated (interval amplitude and framing of the decision) were analyzed first to assess their sensitivity to the task. In the following analyses, specific sets of dilemmas were selected that would allow an answer to be reached for specific questions. The Dilemmas 8 to 12 differ from Dilemmas 2 through 6 only in that the amplitude of the interval employed in the 8-12 dilemmas is half that employed in the 2-6 dilemmas, whereas they share a gains frame. If properly grouped, the responses to those dilemmas allow testing the effects of that pair of factors and of their interaction. An ANOVA within subjects 5×2 (five dilemmas by two amplitudes) on the order scores shows what is obvious in the table. The main effect of both the dilemma and the amplitude is significant: $F(4, 3564) = 158.422$; $p < .001$ for the dilemma; $F(1, 891) = 564.137$, $p < .001$, for the amplitude, but also of the interaction $F(4, 3564) = 18.146$; $p < .001$. The five t -tests show significant differences, as expected, given the large sample employed. However, what is important here is that the direction of the differences is consistent (see Table 2). For each dilemma the shorter amplitude raises significantly lower scores. This is to be expected because of the scoring procedure. Dilemmas 7 and 13 also differ in the amplitude of the interval, although they share a losses frame. Again, significantly more risky responses are raised with the short-amplitude interval dilemma (Table 2).

A four-dilemmas set allows a compact test of the effects of frame and its interaction with amplitude. Specifically, Dilemmas 3 and 7 share a large-amplitude interval, and Dilemmas 9 and 13 share a short amplitude interval. However, whereas Dilemmas 3 and 9 share a gains frame, Dilemmas 7 and 13 share a losses frame. A 2×2 ANOVA of the order scores reflects a significant effect of the interaction $F(1, 891) = 7.456$; $p = .006$ and of the main factors $F(1, 891) = 59.132$, $p < .001$ for the amplitude; and $F(1, 89) = 326.715$, $p < .001$ for the frame. The interaction occurs because the difference is smaller for the short interval than for the large interval even though the losses frame always yields more risky responses.

In short, both factors have an effect on the level of risk of the decisions

made, and both are in the expected direction. Nevertheless, the experimenters are interested in any change in the correlations between the scores as a function of the amplitude and the frame involved. The power of these factors was analyzed in the context of internal consistency.

Table 3 shows the correlations between the order scores on the 10 dilemmas with a gains frame. Internal consistency is .886 (Cronbach's alpha). However, several features of the matrix deserve some attention. The triangle labeled A in Table 3 highlights the correlations between dilemmas that share large amplitude and a gains frame. They are in general high (except for those with Dilemma 4). Triangle B highlights the correlations between dilemmas that share short amplitude and a gains frame. They are in general high (except for those with Dilemma 10; this is the same as Dilemma 4, but with the shortened interval amplitude). Square C highlights the correlations between large-amplitude dilemmas and short-amplitude dilemmas. Again, they are in general high (again, unless Dilemmas 4 and 10 are involved). Rectangle D highlights the correlations between dilemmas that differ only in interval amplitude, although they share the same context; these are the highest correlations in the matrix.

Table 3
Matrix of correlations between the scores on the 10 evaluation dilemmas with a gains frame

	D2	D3	D4	D5	D6	D8	D9	D10	D11	D12
D2										
D3	.534									
D4	.195	.369								
D5	.534	.697	.310							
D6	.481	.616	.286	.662						
D8	.608	.588	.208	.545	.530					
D9	.556	.633	.282	.616	.591	.545				
D10	.171	.306	.792	.288	.285	.205	.291			
D11	.526	.635	.287	.748	.658	.567	.607	.298		
D12	.457	.575	.273	.607	.775	.539	.577	.323	.633	

Comparisons of the correlations within triangle A with those below rectangle C show that the amplitude of the interval makes no difference in the correlations of the scores, even though their averages are different. However, the context does make a difference. In fact, the only three correlations above 0.7 are inside rectangle D. That is, some specificity of the context of the dilemmas is given as to the risk propensity elicited.

Dilemmas 4 and 10 lead to different behavior than the others. Because the difference remains regardless of the amplitude of the interval and the context of the dilemma with which we compare the correlations and the averages, the experimenters believe that the key to the difference is in the context of that pair of dilemmas; its specific context is one of change of job, and the alternatives concern salary. The sample of participants is composed by people who are looking for an opportunity for a better job (for some people it is their first job, but for others it is just a new one).

Furthermore, in these dilemmas the Secure response was associated with change of job (with a better salary), whereas the Game response was associated

with remaining in the present job and waiting for an uncertain promotion within 3 months to a position with a much better salary. So, the S has also some associated uncertainties that perhaps made the participants perceive this alternative as less secure than intended. Nevertheless, what is clear is that the responses for the context of these dilemmas are different from those of the other contexts, although it is not known if this is specific to this sample.

To explore more deeply the different behavior of Dilemmas 4 and 10 from the rest, a factor analysis was applied for the order scores. The result was a solution with two nonrotated factors. The first factor explains 51.5% of the variance; it shows high correlations with all dilemmas except for 4 and 10, for which the correlations are moderate. The second factor explains 12.4% of the variance; it shows small and negative correlations with all dilemmas except for 4 and 10, with which it shows high and positive correlations (see the two columns to the right in Table 1). The conclusion is that Dilemmas 4 and 10, those related with salary, map facets of risk propensity shared with the other dilemmas; but they also map a facet not covered by the others.

Internal consistency for the scores of the 5 original dilemmas (some slightly rephrased from the pilot study and some with a different context) is .794, whereas it is .787 for the scores of the 5 short-interval amplitude dilemmas. That is, the amplitude has no influence on it. When the losses frame dilemma is added (Dilemma 7 to the large-amplitude group and Dilemma 13 to the short-amplitude group), internal consistency improves to .809 and .812, respectively. Of course, internal consistency should improve as the number of items increases; in fact, it reaches .877 when the 10 gains dilemmas are involved and .886 with the complete set of 12 dilemmas. Since Dilemmas 4 and 10 have been shown to reflect slightly different facets of risk propensity tendencies, the internal consistency for the test was obtained without those items; the 10 remaining items show a .901 value for Cronbach's alpha. In short, internal consistency is very good. It improves as the number of items increases. Even when dilemmas slightly different from the majority are included, internal consistency remains close to .90. As a consequence, it is easy to find a combination of dilemmas that optimizes the balance between reliability and length of the task.

A subsample of 275 people participated in both the present selection process and the one in which the pilot version was tested. Despite 2 years of lapse between them, the test-retest correlation was .377 between the pilot version scores and the scores in an application of the original dilemmas (5 dilemmas).

The correlation between the RPD task and the BDT was calculated to assess the validity of the scores (see the description above; Arend et al., 2002), that had been also included in the selection battery. The correlation is -.529 for the 12 dilemmas set and -.545 for the 10 dilemmas set (excluding the two salary dilemmas). The correlation between the RPD task and the scores on the investment game described in the introduction is -.339 for the 12 dilemmas set and -.342 for the 10 dilemmas set (excluding the two salary dilemmas).

Discussion

The RPD task has shown good performance in generating a reliable measure of people's propensities to engage in risk-taking behaviors. It also has shown concurrent validity with another measure of this dimension, such as the BDT. It is based on sequences of three decisions made within a given

decision context, and it was designed to estimate the expected value (subjective utility) of an uncertain situation that is equally appealing as a given secure amount. The contexts were varied, two different amplitudes for the uncertain situation were employed, and both gains and losses framing were involved. Whereas the scores are sensitive to the context of the decision, to the gains/losses frame, and to the amplitude of the interval, they produce significant correlations between them and with the total score. That is, all the dilemmas contribute to capture a complex construct that needs to be approached from several perspectives, with several types of measures, and employing a variety of procedures.

It is important to highlight what happened with dilemmas built on a change of job and salary context. Small differences in the way the dilemmas are designed can have significant consequences on the scores. Each context has to be explored since the responses are sensitive to this. As early as 1962, Slovic pointed out the lack of correspondence between various measures of risk taking. He compared the correlations between several measures: indexed styles of responding to questionnaire tests, self-report questionnaires, and games and lotteries. He concluded that a situational analysis of risk taking was strongly needed. Nevertheless, the lack of convergence between different types of measures is not specific to risk propensity. On the contrary, since apparently there is a generalized problem in the measurement of personality dimensions (Hundleby, 1973; Pervin, 1996; Skinner & Howarth, 1975), several authors have emphasized that monomethodological measurements do not necessarily derive any more valid representations of individuals' personalities and that convergence should be welcomed from other research disciplines (Kubinger, 2002).

As a consequence, the experimenters do not believe that the RPD task should replace those already developed. Although a number of previous tools assess risk propensity, probably none of them addresses all of what is relevant for risk taking. General personality traits, specific self-reports, and behavioral tests all sample important facets of the construct. However, to grasp it in all its richness and complexity, its definition should probably rest on the results from a battery of instruments based on different types of measures, procedures of assessment, contexts, and so on. The RPD contributes to this effort by including a variety of contexts, by employing a successive-steps approach to the estimation of the IV, by profiting from computerized test features (automatic selection of values for the proposals), and by including different amplitude intervals and decision frames. Furthermore, the RPD is easily adapted to different contexts of those employed for the dilemmas of the test used in the present research (Table 1).

Another benefit from the present line of research concerns the practical problem of response distortions in selection processes. Since Jackson and Messick (1967) concluded that response differences in MMPI might reflect little more than individual differences in the tendency to answer according to social desirability or acquiescence, many pieces of research have studied the role of voluntary distortion in self-reports, particularly in personnel selection contexts (Ones, Viswesvaran, & Reiss, 1996; Viswesvaran & Ones, 1999). This has become the most important criticism against personality questionnaires, particularly when they are used as the unique measurement of personality dimensions. An alternative to traditional questionnaires are the objective tests, in Cattell's sense (Cattell & Warburton, 1967). These task-based tests keep their aims masked and do not give any feedback of an individual's performance.

Voluntary biases are then prevented (see Santacreu, Rubio, & Hernández, 2006, for a detailed description of objective task-based tests).

Candidates in many selection processes are assessed as to their RP. However, many candidates think it is not beneficial for them to appear as too risky or too conservative. As a consequence, they previously over-train in the tasks more often employed to assess the trait in the selection processes. Having available a larger catalogue of interchangeable tasks not only allows for alternating tasks, but it also reduces the risk of over-training for social desirability and for lying or faking.

In short, the RPD task should be added to the catalogue of tasks sensitive to risk propensity, since it conveys several benefits. One benefit is that it helps to build an ideal battery that jointly maps as many facets as possible from any risk-propensity trait. Another benefit is that it can be alternated with other tasks to avoid over-training with specific measures.

References

- AREND, I. C., BOTELLA, J., CONTRERAS, M. J., HERNANDEZ, J. M., & SANTACREU, J. (2003). A betting dice test to study the interactive style of risk-taking behavior. *The Psychological Record*, *53*, 217-230.
- BECHARA, A. (2003). Risky business: emotion, decision-making and addiction. *Journal of Gambling Studies*, *19*, 23-51.
- BENJAMIN, A. M., & ROBBINS, S. J. (2007). The role of framing effects in performance on the balloon analogue risk task (BART). *Personality and Individual Differences*, *43*, 221-230.
- BOSTIC, R., HERRNSTEIN, R. J., & LUCE, R. D. (1990). The effect on the preference-reversal phenomenon of using choice indifferences. *Journal of Economic Behavior*, *13*, 193-212.
- CATTELL R.B., & WARBURTON, E.W. (1967). *Objective personality and motivation Tests*. Urbana: University of Illinois Press.
- EYSENCK, H. J., & EYSENCK, S. B. G. (1998). *Eysenck Personality Scales*. Kent: Hooper & Stoughton.
- GROL, R., WHITEFIELD, M., DE MAESENEER, J., & MOKKINK, H. (1990). Attitudes to risk taking in medical decision making among British, Dutch and Belgian general practitioners. *British Journal of General Practice*, *40*, 134-136.
- HARRISON, J. D., YOUNG, J. M., BUTOW, P., SALKELD, G., & SOLOMON, M. J. (2005). Is it worth the risk? A systematic review of instruments that measure risk propensity for use in the health setting. *Social Science & Medicine*, *60*(6), 1385-1396.
- HERNANDEZ, J. M. (2000). *La personalidad*. Madrid: Biblioteca Nueva.
- HUNDLEBY, J. D. (1973). The measurement of personality by objective tests. In P. KLINE (Ed.), *New approaches in psychological measurements*. New York: John Wiley & Sons.
- JACKSON, D. N., & MESSICK, S. (1967). Response styles and the assessment of psychopathology. In D.N. JACKSON & S. MESSICK (Eds.). *Problems in human assessment* (pp. 541-555). New York: McGraw-Hill.
- KAHNEMAN, D., & TVERSKY, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, *47*, 263-291.
- KOGAN, N., & WALLACH, M. A. (1964). *Risk-taking: A study in cognition and personality*. New York: Holt, Rinehart & Winston.

- KUBINGER, K. D. (2002). Psychology's challenge when personality questionnaires are applied for individual assessment. *Psychology Science*, *44*, 3-9.
- KUHBERGER, A. (1998). The influence of framing on risky decision: A meta-analysis. *Organizational Behavior and Human Decision Processes*, *75*, 23-55.
- LEIGH, B. C. (1999). Peril, chance, and adventure: Concepts of risk, alcohol use, and risky behavior in young adults. *Addiction*, *94*, 371-383.
- LEJUEZ, C. W., READ, J. P., KAHLER, C. W., RICHARDS, J. N., RAMSEY, S. E., STUART, G. L., STRONG, D. L., & BROWN, R. A. (2002). Evaluation of a behavioral measure of risk taking: The balloon analogue risk task (BART). *Journal of Experimental Psychology: Applied*, *8*, 75-84.
- LICHTENSTEIN, S. & SLOVIC, P. (1971). Reversals of preference between bids and choices in gambling decisions. *Journal of Experimental Psychology*, *89*, 46-55.
- NICHOLSON, N., SOANE, E., FENTON-O'CREEVY, M., & WILLIAM, P. (2004). Personality and domain-specific risk taking. *Journal of Risk Research*, *8*, 157-176.
- ONES, D. S., VISWESVARAN, C., & REISS, A.D. (1996). The role of social desirability in personality testing for personnel selection: The red herring. *Journal of Applied Psychology*, *81*, 660-679.
- PERVIN, L. A. (1996). *The Science of Personality*. New York: Wiley & Sons.
- RUBIO, V.J., HERNANDEZ, J.M., ZALDIVAR, F., MARQUEZ, M.O., & SANTACREU (submitted). Can we predict risk-taking behavior? Two behavioral tests for predicting guessing tendencies in a multiple choice test.
- RUBIO, V. J., SANTACREU, J., & HERNANDEZ, J. M. (2004). The objective assessment of personality: An alternative to self-report based assessment. *Análisis y Modificación de Conducta*, *30*, 827-840.
- RUBIO, V. J., SANTACREU, J., & HERNANDEZ, J. M. (2006). Die Erfassung der individuellen risikotendenz mit objektiver persönlichkeits-tests. In T.M. Ortner, R. Proyer & K.D. Kubinger (Eds.) *Objektiver Persönlichkeitstest*. Bern: Hans Huber.
- SANTACREU, J., RUBIO, V. J., & HERNANDEZ, J. M. (2006). The objective assessment of personality: Cattell's T-data revisited and more. *Psychology Science*, *48*, 53-68.
- SANTE, L., & SANTACREU, J. (2001). La eficacia (o la suerte) como moduladora del estilo interactivo tendencia al riesgo. *Acta Comportamental*, *9*, 163-188.
- SITKIN, S. B., & WEINGART, L. R. (1995). Determinants of risky decision-making behavior: A test of the mediating role of risk perceptions and propensity. *Academy Management Journal*, *38*, 1573-1592.
- SKEEL, R.L., NEUDECKER, J., PILARSKI, C., & PYTLAK, K. (2006). The utility of personality variables and behaviorally-based measures in the prediction of risk-taking behavior. *Personality and Individual Differences*, *43*, 203-214.
- SKINNER, N. S. F. & HOWARTH, E. (1975). Cross media independence of questionnaire and objective test personality factors. *Multivariate Behavioral Research*, *8*, 23-40
- SLOVIC, P. (1962). Convergent validation of risk taking measures. *Journal of Abnormal Social Psychology*, *65*, 68-71.
- TVERSKY, A., & KAHNEMAN, D. (1981). The framing of decisions and the psychology of choice. *Science*, *211*, 453-458.

- VIGIL-COLET, A. (2007). Impulsivity and decision making in the balloon analogue risk-taking task. *Personality and Individual Differences, 43*, 37-45.
- VISWESVARAN, C., & ONES, D.S (1999). Meta-analyses of fakability estimates: Implications for personality measurement. *Educational and Psychological Measurement, 59*, 197-210.
- WEBER, E. U., BLAIS, A., & BETZ, N. E. (2002). A domain-specific risk-attitude scale: measuring risk perception and risk behaviors. *Journal of Behavioral Decision Making, 15*, 263-290.
- ZUCKERMAN, M. (1994). *Behavioral expressions and biosocial bases of sensation seeking scales*. New York: Cambridge University Press.