Journal of Behavioral Decision Making, *J. Behav. Dec. Making* (2013)
Published online in Wiley Online Library (wileyonlinelibrary.com) **DOI:** 10.1002/bdm.1784

Individual Differences in Risky Decision Making: A Meta-analysis of Sensation Seeking and Impulsivity with the Balloon Analogue Risk Task

MARCO LAURIOLA[1]*, ANGELO PANNO[1], IRWIN P. LEVIN[2] and CARL W. LEJUEZ[3]
[1]*Department of Social and Developmental Psychology, University of Rome "Sapienza", Italy*
[2]*Department of Psychology, University of Iowa, USA*
[3]*Department of Psychology, University of Maryland, USA*

SUMMARY

To represent the state-of-the-art in an effort to understand the relation between personality and risk taking, we selected a popular decision task with characteristics that parallel risk taking in the real world and two personality traits commonly believed to influence risk taking. A meta-analysis is presented based on 22 studies of the Balloon Analogue Risk Task from which correlations with sensation seeking and impulsivity assessments could be obtained. Results calculated on a total of 2120 participants showed that effect size for the relation of sensation seeking with risk taking was in the small–moderate range ($\bar{r} = .14$), whereas the effect size for impulsivity was just around the small effect size threshold ($\bar{r} = .10$). Although we considered participants' demographics as moderators, we found only significantly larger effect sizes for the older adolescents and young adults compared with other ages. The findings of the present review supported the view that inconsistencies in personality–risk research were mostly due to random fluctuations of specific effect sizes, rather than to lack of theoretical ties or to measurement unreliability. It is also concluded that studies aimed at relating individual differences in personality to performance in experimental decision tasks need an appropriate sample size to achieve the power to produce significant results. Copyright © 2013 John Wiley & Sons, Ltd.

KEY WORDS risky decision making; individual differences; Balloon Analogue Risk Task; sensation seeking; impulsivity

Many personality-like concepts have been considered to account for individual differences in judgment and decision-making processes. However, recent excellent reviews concluded that empirical research on this topic yielded inconsistent and contradictory results, thus dismissing—or diminishing—the role of personality in decision making as well as increasing the gap between behavioral decision-making studies and studies of personality (Appelt, Milch, Handgraaf, & Weber, 2011; Mohammed & Schwall, 2009). Consequently, a fragmented and discouraging picture emerged, especially when attempting to relate individual differences in risk-related personality traits to risky choices in experimental tasks (e.g., Lauriola & Levin, 2001)

Possible reasons for inconsistent and contradictory findings may lie in theoretical and methodological differences between the personality literature and the decision-making literature. Studies of risky decision making were traditionally based on a variety of hypothetical gambles and economic games administered in laboratory settings to disclose systematic violations of the principles of rationality involved in making decisions (e.g., Weber & Johnson, 2008). By contrast, studies of personality were focused on self-report measures of risk-related traits, such as sensation seeking and impulsivity, to understand and explain individual differences in real-world risky behaviors (e.g., Boyer, 2006; De Wit, 2009; Verdejo-Garcia, Lawrence, & Clark, 2008; Zuckerman & Kuhlman, 2000).

Recently, Mohammed and Schwall (2009) encouraged decision researchers interested in individual differences to design experiments that minimize the power of the experimental situation, thus increasing their chance of detecting any smaller but consistent effect of individual difference factors. For example, manipulating the outcome valence according to some classic risky decision-making paradigms has a larger effect on risky choices (e.g., Cohen's $d = .57$; Kuhberger, 1998, p. 38) than individual difference variables that mask their effect on risk taking in experimental tasks.

In fact, individual differences interact with situational or task characteristics, thereby producing low or nonsignificant coefficients when risk-taking tendencies are assessed across different tasks or domains (e.g., Figner & Weber, 2011). Moreover, Appelt et al. (2011) also recommended a more systematic approach to the study of individual differences as well as a shift toward theoretically sound measures. In agreement with both these views, we maintain that not only do different studies of personality and decision making often use different measures of the same construct or "ad hoc" variations of personality scales whose reliability and validity are questionable but also that in most cases, decision-making researchers merely added to a variety of experimental tasks an unsystematic set of personality measures.

In the last few years, however, new experimental tasks became increasingly popular in decision making as well as in personality research. In one class of tasks, people are required to make repeated choices where risk levels escalate as a result of one's previous decisions (e.g., Lejuez et al., 2002; Pleskac, 2008; Figner, Mackinlay, Wilkening, & Weber, 2009; Gardner & Steinberg, 2005; Slovic, 1966). This feature not only is considered as one of the main characteristics of real-world risky behaviors (Goldberg & Fischhoff, 2000; Leigh, 1999; Moore & Gullone, 1996; Schonberg, Fox, & Poldrack, 2010; Weber & Johnson, 2009) but also provides a link to decision-making accounts of risky behaviors dealing with sequential outcomes

* Correspondence to: Marco Lauriola, Department of Social and Developmental Psychology, University of Rome, Via dei Marsi 78, 00185 Rome, Italy.
E-mail: marco.lauriola@uniroma1.it

(e.g., the "hot or cold hand" or the "gambler's fallacy" heuristics). Examples include investing in riskier stocks to recover from earlier investments or increasing the size of gambles after a string of prior losses (e.g., Johnson, Tellis, & Macinnis, 2005).

Consistent with all these premises, we devised the present review to examine whether and to what extent one can find consistent personality–risk relations if one reduces the variety of task conditions and systematically selects theoretically sound individual difference variables (Appelt et al., 2011; Mohammed & Schwall, 2009). Hence, we focused on a popular decision task with characteristics that parallel risk taking in the real world and two personality traits commonly believed to influence real-world risk taking. The choice of task and variables was also driven by the need for a relatively large number of investigations with common characteristics from which to glean reliable findings. With these criteria in mind, we chose the Balloon Analogue Risk Task (BART), which has a 10-year history of use, and the personality variables of sensation seeking and impulsivity, which are thought to be valid and strong predictors of risk-taking behavior in different real-world risky domains (e.g., Adams & Moore, 2007; Dunlop & Romer, 2010; Dahlen, Martin, Ragan, & Kuhlman, 2005; Hittner & Swickert, 2006; Johansson, Grant, Kim, Odlaug, & Gotestam, 2009; Nelson, Lust, Story, & Ehlinger, 2008; Frankenberger, 2004.

The Balloon Analogue Risk Task as a valid behavioral measure of risk taking

Expanding on the basic framework of the Devil Task (Slovic, 1966), Lejuez and colleagues developed the BART (Lejuez et al., 2002). BART is a computerized task that models real-world risk behavior through the conceptual frame of balancing the potential for reward and harm (Leigh, 1999; Lejuez et al., 2002). In the task, the participant is presented with a balloon and asked to pump it by clicking a button on the screen. With each click, the balloon inflates .3 cm, and money is added to the participant's temporary winnings; however, balloons also have explosion points, which can be varied both across and within studies. Before the balloon pops, the participant can press "Collect $$$," which saves his or her earnings to a permanent bank. If the balloon pops before the participant collects the money, all earnings for that balloon are lost, and the next balloon is presented. Thus, each pump confers not only greater risk but also greater potential reward.

The primary BART score is the average number of pumps on unexploded balloons, with higher scores indicative of greater risk-taking propensity. In the original version of the task, each pump was worth $.05, and there were 30 total balloons for each of three different balloon colors, with each color having a different probability of exploding on the first pump (1/8, 1/32, and 1/128, respectively). If the balloon did not explode after the first pump, the probability that the balloon would explode on the next pumps increased linearly until the last pump at which the probability of an explosion was 1.00.

Participants were given no information about the breakpoints and the absence of this information allowed for the examination of participants' initial responses to the task and to changes as they experienced the contingencies related to payout collections and balloon explosions. Results of the original study showed that the average number of pumps on unexploded balloons, also referred to as average adjusted pumps, was associated with some real-world risky behaviors occurring outside the laboratory (e.g., smoking and theft) as well as with self-report measures of personality traits, including impulsivity and sensation seeking. The original study also established some relationships between key outcome variables and BART scores that were most evident at the 1/128 explosion probability, which then became the level most commonly used in subsequent studies. Accordingly, the average number of pumps that would maximize earnings should be equal to 64 pumps, with lower and higher numbers describing risk-advantageous and risk-disadvantageous strategies, respectively (Lejuez et al., 2002).

Different types of studies supported the validity of BART. First, there are experimental studies whose goal was evaluating how one's performance on BART changed as a function of altering the psycho-physiological homeostasis of the body such as by sleep deprivation, medical therapies, or craving. These studies, which compared experimental groups versus control groups, provided evidence that the average number of pumps was affected by the experimental manipulations. Overall, this literature showed a connection between specific psycho-physiological processes and behavioral expression of risk taking, although enhanced or suppressed by different moderators in each study (Reynolds, Ortengren, Richards, & de Wit, 2006; Acheson, Richards, & de Wit, 2007; Acheson & de Wit, 2008; White, Lejuez, & de Wit, 2007; Killgore, 2007; Killgore et al., 2008; Lighthall, Mather, & Gorlick, 2009; Reed, Levin, & Evans, 2010).

The second type of study is based on quasi-experimental designs whose main goal was assessing how one's risky behavior in BART differed between groups composed of participants selected from a population whose risk of addiction was extremely high compared with healthy controls (e.g., crack or cocaine users, marijuana smokers, and alcohol or tobacco dependent people). Again, these studies supported the validity of BART by showing that different "at risk" groups often displayed greater average pumps than control groups (Bishara et al., 2009; Bornovalova, Daughters, Hernandez, Richards, & Lejuez, 2005; Coffey, Schumacher, Baschnagel, Hawk, & Holloman, 2011; Crowley, Raymond, Mikulich-Gilbertson, Thompson, & Lejuez, 2006; Hunt, Hopko, Bare, Lejuez, & Robinson, 2005; Ledgerwood, Alessi, Phoenix, & Petry, 2009). This effect was robust as shown by studies that controlled for demographics and other relevant individual difference factors, including impulsive (e.g., Lejuez, Aklin, Jones, et al., 2003a; Lejuez, Aklin, Bornovalova, & Moolchan, 2005; Aklin, Lejuez, Zvolensky, Kahler, & Gwadz, 2005).

Finally, like classic personality studies of risk taking, most publications also reported significant correlations between the average adjusted pump and real-world risky behaviors, such as drug and alcohol use, smoking, gambling,

DOI: 10.1002/bdm

aggression, psychopathic tendencies, and unprotected sex (Aklin et al., 2005; Hunt et al., 2005; Lejuez et al., 2002; Lejuez, Aklin, Jones, et al., 2003a; Lejuez et al., 2007; Skeel, Pilarski, Pytlak, & Neudecker, 2008; Bornovalova et al., 2009; Mishra, Lalumière, & Williams, 2010; Swogger, Walsh, Lejuez, & Kosson, 2010; MacPherson, Magidson, Reynolds, Kahler, & Lejuez, 2010).

Taken together, this literature provides strong support for the validity of BART as a behavioral measure of risk taking. Hence, utilizing the BART based on the strengths noted, the goal of the current study was to examine the nature of its relationship with key personality constructs. Consistent with the guidelines for the study of individual differences in decision making (Appelt et al., 2011; Mohammed & Schwall, 2009) as well as with the goals of the present review, we narrowed our interest to sensation seeking and impulsivity as these individual difference variables have the greatest potential to account for a consistent and non-negligible amount of risk-taking variance both outside and inside the laboratory setting.

Sensation seeking, impulsivity, real-world risk taking, and risky decision making

The sensation seeking trait is defined by individual differences "in the seeking of varied, novel, complex, and intense sensations and experiences, and the willingness to take physical, social, legal, and financial risks for the sake of such experience" (Zuckerman, 1994, p. 27). Accordingly, the theoretical underpinning for its association with risk taking is based on the presumed excitement and arousal that can be provided to sensation seekers by specific stimulating experiences, most of which necessarily involve a high element of risk (Zuckerman, 2007).

Although different instruments have been developed for assessing this trait (e.g., Arnett, 1994; Hoyle, Stephenson, Palmgreen, Pugzles Lorch, & Donohew, 2002; Zuckerman, Kuhlman, Joireman, Teta, & Kraft, 1993), the Sensation Seeking Scales (Zuckerman, Eysenck, & Eysenck, 1978) are the most extensively used in personality–risk research. In general, the sensation seeking trait correlated so reliably with one's involvement in real-world risky activities, including thrill-seeking sports, unsafe sex, recreational smoking and drinking, drug use, and so forth, that it is acknowledged among personality psychologists as a measure of risk taking itself (e.g., Adams & Moore, 2007; Dunlop & Romer, 2010; Dahlen et al., 2005; Hittner & Swickert, 2006; Johansson et al., 2009; Nelson et al., 2008; Frankenberger, 2004).

However, it is worth noting that most of this literature equated risk taking with recreational risky activities of high stimulating value for adolescents, whereas less is known about the association of sensation seeking with other types of risks, including one's performance in escalating risk tasks, like BART. For instance, it has been documented that extreme risk takers in a recreational domain (i.e., bungee jumpers) only take a moderate amount of risk when making financial decisions (Hanoch, Johnson, & Wilke, 2006). Thus, risk-taking behavior not only is domain specific, but there might also be specific personality–risk relations within each domain, as documented by a recent study of personality traits and likelihood of engaging in risky activities (Weller & Tikir, 2011).

Expanding on Zuckerman (1994), Glicksohn and Abulafia (1998) pointed out that the sensation seeking trait should be reconsidered in terms of two narrow aspects, which may account for specific personality–risk relations. On the one hand, the so-called "nonimpulsive, socialized mode of sensation seeking," measured by the Thrill and Adventure Seeking subcomponent, is the personality characteristic most likely involved in taking types of risk motivated by the need for stimulation (e.g., bungee jumping). On the other hand, the "impulsive, unsocialized mode of sensation seeking," resulting from a combination of Disinhibition, Boredom Susceptibility, Experience Seeking with facets of Impulsivity, and Psychoticism, can be involved in other types of risks, such as those resulting from ignoring stop signals in dangerous reward-seeking behaviors (e.g., gambling).

Like financial investments, hypothetical gambles and economic games administered in laboratory settings often lack any element of arousal, and this might account for inconsistent results when relating personality to such risky decisions. By contrast, modern escalating risk tasks seem to provide some elements of arousal, which may disclose more systematic relations of sensation seeking with risk taking in experimental tasks (e.g., de Haan et al., 2011; Penolazzi, Gremigni, & Russo, 2012).

However, as regards the relation of sensation seeking with BART, some studies have found the expected positive relation (Lejuez et al., 2005; Lejuez et al., 2002; Lejuez et al., 2007; MacPherson et al., 2010), whereas others failed to find it (Aklin et al., 2005; Benjamin & Robbins, 2007; Lejuez et al., 2003a; Lejuez, Aklin, Zvolensky, & Pedulla, 2003b; Killgore, 2007). Thus, although sensation seeking is considered as one of the major personality determinants of real-world recreational risky behaviors during adolescence, whether and to what extent this trait has a consistent main effect on risk taking in controlled experimental risk tasks is still an open question. We strongly believe that the reason for inconsistent results is the random fluctuation of effect sizes around the "true" sensation seeking average effect size, and the present review confirms that.

Other personality–risk research has indicated that impulsivity rather than sensation seeking is a major personality characteristic involved in the occurrence of real-world risky behaviors. Unlike sensation seeking, it is hard to find a common theoretical definition of impulsivity, and this label "is applied somewhat indiscriminately to individuals, behavior, and cognitive processes" (Enticott & Ogloff, 2006, p. 4). Consequently, the impulsivity trait has been operationally defined in many ways, ranging from multidimensional personality scales (e.g., Dickman, 1990; Eysenck, Pearson, Easting, & Allsopp, 1985; Patton, Stanford, & Barratt, 1995; Tellegen, 1982; Reynolds et al., 2006) to behavioral assessments, such as measuring one's ability to inhibit responses in stop signal tasks (e.g., Logan, Schachar, & Tannock, 1997) or inferring one's individual discount rates from observed choices in hypothetical delay discount tasks

DOI: 10.1002/bdm

(e.g., Baker, Johnson, & Bickel, 2003; Madden, Petry, Badger, & Bickel, 1997; Bickel, Odum, & Madden 1999). Furthermore, some authors have gone so far as to consider BART itself as a behavioral impulsivity measure rather than a decision-making task of escalating risk (e.g., Reynolds et al., 2006; Acheson et al., 2007; Acheson & de Wit, 2008; Cross, Copping, & Campbell, 2011), whereas others provided evidence that BART, impulsivity, and sensation seeking scores loaded on separate factors (Meda et al., 2009).

As a result of this fragmentation, the theoretical underpinning for the impulsivity–risk relation is more complex and still debated, but regardless of the conceptual framework, real-world risk taking is believed to be the behavioral expression of impulsivity (Enticott & Ogloff, 2006). Accordingly, the impulsivity trait has been a reliable predictor of behaviors, such us drug abuse, risky driving, unprotected sex, and problem gambling (e.g., Chambers & Potenza, 2003; Dahlen et al., 2005; De Wit, 2009; Hoyle, Fejfar, & Miller, 2000). However, if one gets more into this literature, one can see that the association of impulsivity with risk taking is often inferred from studies of severe health, financial, or social risks, which more often occurred among special populations, such as psychiatric disordered or developmentally disabled individuals (e.g., Holmes et al., 2009; Hunt et al., 2005; Moeller, Barratt, Dougherty, Schmitz, & Swann, 2001; Tull et al., 2009) as well as types of socially deviant individuals (e.g., Hopko et al., 2006).

Most studies of Impulsivity and BART also failed to reveal the expected relation, although there was a tendency for a positive correlation (Hunt et al., 2005; Reynolds et al., 2006; Lejuez et al., 2007; Bornovalova, Gwadz, Kahler, Aklin, & Lejuez, 2008; Pleskac et al., 2008; Skeel et al., 2008; Romer et al., 2009; Cyders et al., 2010; Marini & Stickle, 2010). Also in this case, we believe that sampling variability has contributed to random fluctuations of effect sizes.

Although sensation seeking and impulsivity may each affect risk taking, they also proved to be conceptually and empirically distinct constructs (Kirby & Finch, 2010; Steinberg, 2007, 2008; Steinberg et al., 2008; Whiteside & Lynam, 2001). Hence, another open question is whether Sensation Seeking and Impulsivity may have a joint effect as personal determinants of risk taking. As Steinberg et al. (2008) pointed out, "Not all impulsivity leads to stimulating or even rewarding experiences (e.g., impulsively deciding to end a friendship), and not all sensation seeking is done impulsively (e.g., purchasing advance tickets to ride a roller coaster or sky dive)" (p. 1765). Likewise, not all real-world risky behaviors are motivated by the need for arousal (e.g., financial investments or ethical risks), although there might be situations that may turn out to be risky just because of inhibitory failure, delay aversion, increased autonomic arousal, or deficient forethought (e.g., forgetting precautions for preventing accidents or crossing the road impulsively). As to this point, it is worth noting that Zuckerman and Kuhlman (2000) expanded on the original notion of Sensation Seeking and Impulsivity by positing the existence of a synergic relation of specific facets of the two traits, which may also account for their "marriage" at a higher level of analysis, as it is evident from intercorrelations and factor analyses (Joireman & Kuhlman, 2004).

As a consequence, people who are high on both traits are supposed to take more risk than people who are high on only one trait, because they are attracted by rewards (including arousal), while at the same time they may neglect signs of potential punishment or losses, as predicted by impulsivity theory (Roberti, 2004). It is worth noting that some studies of the BART included in this review were in keeping with these theoretical developments and predicted risk taking with positive results by using combined scores of impulsivity and sensation seeking items (i.e., ImpSS) instead of separated measures of the two traits (e.g., Bornovalova et al., 2009; Lejuez et al., 2005). Interestingly, one of these studies also varied the level of reward associated with each pump. As the reward increased, there was a drop in risk taking that was particularly evident for people low in ImpSS, whereas—as expected—people high in ImpSS were largely insensitive to variation in reward/loss magnitude (Bornovalova et al., 2009).

Our goal then is to complete a comprehensive meta-analysis on a relatively large set of studies of BART from which a clear sense of the relationship between the task and key personality constructs of sensation seeking and impulsivity could be obtained. In keeping with our general hypothesis, we maintain that inconsistent personality–risk relations in the literature can be due at least in part to the unsystematic use of different personality assessments (Appelt et al., 2011) in combination with different experimental designs, which can induce large situational effects overwhelming smaller but consistent effects of individual differences (Mohammed & Schwall, 2009). In addition, sampling variability may account for some deviation of each specific study's effect size from the average population effect size, especially for studies carried out on relatively small samples. However, with the literature that we have reviewed so far, we expect different effect size for sensation seeking and impulsivity with risk taking on BART. In fact, a larger consensus exists in personality research on the relation of sensation seeking scales with common recreational risks during adolescence and young adulthood. By contrast, there is less agreement on which type of risks is associated with impulsivity (recreational vs. severe health, social and financial risks) as well as whether other extreme person characteristics, such as one's level of psychopathology, social deviance, or developmental disability, might be involved as factors that may favor the behavioral expression of risk taking.

METHOD

Sample of studies

The initial search for journal articles was conducted with the database PsycINFO, which has a broad coverage of both psychology and social science journals. Search terms included the keywords "BART" or "Balloon Analogue Risk Task" and "sensation seeking" or "impuls*". The following search limits were imposed: (i) human populations only, (ii) English language only, (iii) peer-reviewed journal, and (iv) articles published between 2002 and 2011. In addition, other relevant articles were included in the working list if

DOI: 10.1002/bdm

they cited the original article by Lejuez et al. (2002). This search yielded 67 abstracts.

Abstracts were then screened, and any articles failing to meet the following criteria were removed: (i) the study was empirical, (ii) self-reported psychometric and/or behavioral measures of sensation seeking and impulsivity were used, (iii) impulsivity was measured as an independent construct (for instance, some common attention-deficit/hyperactivity disorder checklists amalgamate hyperactivity and impulsivity into a single dimension and report a single combined measure; such scales were excluded), (iv) data were presented or potentially available from which an effect between sensation seeking or impulsivity and BART could be calculated, and (v) the effects of impulsivity or sensation seeking on BART were not affected by experimental manipulations (e.g., sleep deprivation and drugs or placebos) or by clinical conditions (alcoholism, psychoactive drug use, psychiatric diagnosis, etc.). In the latter cases, the effect size was calculated only for research participants in control groups or conditions, if any.

Cases in which abstracts did not provide sufficient information to establish whether the articles met with the inclusion criteria were included in a next stage of selection based on the reading of the full manuscript. A total of 65 articles were downloaded or requested through interlibrary loan. Among them there were 30, 18, and 7 articles dealing with impulsivity and sensation seeking, impulsivity only, and sensation seeking only, respectively. If an article met the inclusion criteria but lacked sufficient data for an effect size to be computed, authors were contacted by e-mail. Ultimately, 22 articles were included in the meta-analysis, including 18 and 19 studies for sensation seeking and impulsivity, respectively (Tables 1 and 2).

Coding the studies

To meet the main goals of this meta-analysis, two datasets were completed. One included studies with sensation seeking measures; the other included studies with impulsivity measures. For each study, the following information was coded: (i) all statistics relevant to the assessment of the effect size (correlation, *N*, *p*-values, reliability), (ii) the sample mean age, (iii) sample gender composition (over 55% female participants, over 55% male participants, or balanced gender), (iv) sample ethnic composition (over 50% Caucasian vs. over 50% Ethnic Minorities), (v) the type of population studied (school or college students vs. community samples), (vi) the type of assessment instrument employed in the study (e.g., Eysenck Impulsivity Scale and Barrat Impulsivity Scale), and (vii) the publication year of the study. Finally, (i) the sample mean and SD for the average number of pumps on unexploded balloons were coded to distinguish different levels of BART performance. Only studies in which the average number of pumps was higher than 64 (for 1/128 starting probability of explosion) may disclose a relation between personality traits and a risk-disadvantageous strategy. When the average number of pumps was below this point, which was the case for all studies included in this review, the relation of personality with risk taking is in terms of risk-advantageous choices. The coding of categorical variables was undertaken by two coders. Cohen's kappa was calculated as a measure of inter-rater agreement and ranged from .83 to 1.00. Discrepancies were checked and resolved by agreement between the two coders.

Statistical analyses

Effect size and moderator analyses were carried out according to the methods described by Lipsey and Wilson (2001). However, all publication bias analyses were in keeping with Field and Gillett (2010).

Aggregated effect size

Because our meta-analysis was based on correlational evidence, it was logical to adopt the Pearson's correlation coefficient, *r*, as the common metric for comparing studies and for computing aggregate effect sizes for sensation seeking and impulsivity with risk taking. According to the conventional standards (Cohen, 1988), *r*-based effect size was appraised as follows: .10 is "small," .25 is "medium," and .40 or greater is "large."

The requirement of independence of observations means that the same sample cannot count multiple times when computing an aggregate effect size. However, some of the studies used multiple measures of impulsivity and/or sensation seeking as well as different combinations of behavioral and/or self-report scales. Multiple correlation coefficients from a single sample were indeed combined within each study according to Hunter and Schmidt (2004), and the resulting composite correlations for Sensation Seeking and Impulsivity were analyzed in two separate datasets (Tables 1 and 2).

The aggregate effect size for each dataset was computed according to the standard Inverse Variance weighting method, hereafter referred to as IV, so that the relative contribution of each study was proportionate to the statistical precision of its effect size (Lipsey & Wilson, 2001, p. 64). Heterogeneity among studies in each dataset was tested on the basis of the Q statistic, which has a chi-square distribution with $K-1$ degrees of freedom, with K indicating the total number of studies in each dataset (Hedges & Olkin, 1985). The null hypothesis assumes that the variance between studies is due to sampling variability. Consequently, if the Q statistic turns out to be statistically significant, one can hypothesize other sources of between-study variance, which might be either due to systematic differences between studies or to random differences.

Heterogeneity, if any, can also be incorporated into effect size estimates by switching from a fixed effect model to a random effect one. As the conceptual background of this review suggested that there might be relatively high heterogeneity within the various personality domains and measures, a random effect model was first implemented. However, because the heterogeneity turned out to be very limited and not statistically significant, the fixed effect approach was followed.

Hunter and Schmidt (2004) suggested a different approach to meta-analysis of correlations aimed at controlling for methodological artifacts, such as measurement unreliability, which is likely to affect study conclusions based on psychometric scales. In keeping with Lipsey and Wilson (2001, pp 109–110), we implemented the Hunter–Schmidt approach, hereafter

DOI: 10.1002/bdm

Table 1. Studies included in the meta-analysis of the relationship between sensation seeking and BART. Effect sizes, weights, continuous and categorical moderators are also reported.

	Study	ES *r*	IVW	SS rtt	HS ES *r*	HS IVW	Mean age	Performance level	Time lag (years)	Scale	Gender	Ethnicity	Sample
1	Aklin et al. (2005)	0.27	48	0.71	0.34	29.99	14.80	26.40	3	Z[a]	Bal	Min	HS
2	Benjamin and Robbins (2007)	0.24	33	0.78	0.29	22.65	20.20	40.67	5	Z[a]	F	Cau	Und
3	Bornovalova et al. (2009)	−0.04	37	0.78	−0.05	25.40	19.80	35.22	7	Z[a]	Bal	Cau	Und
4	Cyders et al. (2010)	0.08	101	0.80	0.10	71.10	19.22	41.89	8	M[b]	Bal	Cau	Und
5	Jones and Lejuez (2005)	0.02	57	0.82	0.02	41.13	19.06	30.28	3	Z[a]	F	Cau	Und
6	Killgore (2007)	0.26	51	0.76	0.32	34.11	23.50	35.60	5	H[c]	Bal	Cau	Com
7	Lejuez et al. (2002)	0.31	83	0.87	0.35	63.54	20.90	29.40	0	MM[d]	Bal	Cau	Com
8	Lejuez, Aklin, Jones, et al. (2003a)	0.11	57	0.80	0.13	40.13	20.10	39.80	1	Z[a]	Bal	Cau	Und
9	Lejuez, Aklin, Zvolensky, et al. (2003b)	0.33	23	0.80	0.39	16.19	15.23	33.00	1	Z[a]	F	Min	HS
10	Lejuez et al. (2005)	0.18	122	0.75	0.22	80.52	15.10	32.48	3	Z[a]	F	Min	HS
11	Lejuez et al. (2007)	0.20	95	0.78	0.24	65.21	14.80	35.10	5	Z[a]	Bal	Min	HS
12	MacPherson et al. (2010)	0.14	254	0.72	0.18	160.93	11.00	35.43	8	H[c]	Bal	Min	HS
13	Marini and Stickle (2010)	0.17	147	0.79	0.20	102.19	15.10	38.64	8	Z[a]	M	Cau	Inm
14	Mishra et al. (2010)	0.16	112	0.84	0.19	82.79	20.30	33.00	8	Z[a]	Bal	Cau	Und
15	Reynolds et al. (2006)	0.16	67	0.82	0.19	48.35	22.92	35.60	4	MM[e]	Bal	Cau	Und
16	Romer et al. (2009)	0.03	384	0.74	0.04	250.06	11.40	24.60	7	H[c]	Bal	Cau	HS
17	Skeel et al. (2008)	0.07	97	0.71	0.09	60.61	19.22	36.20	6	M[f]	F	Cau	Und
18	Swogger et al. (2010)	0.22	116	0.82	0.26	83.71	27.92	31.52	8	M[g]	M	Min	Inm

ES *r* = Effect Size *r*; IVW = Inverse Variance Weight; SS rtt = Sensation Seeking Reliability used for correction; HS ES *r* = Hunter–Schmidt Effect Size *r*; HS IVW = Hunter–Schmidt Inverse Variance Weight. Z[a] = SSS Form V, total (Zuckerman et al., 1978); M[b] = UPPS, Sensation Seeking subscale (Whiteside & Lynam, 2001); H[c] = BSSS Brief Sensation Seeking Scale, total (Hoyle et al., 2002); MM[d] = SSS Form V, total (Zuckerman et al., 1978), I7, Venturesomeness subscale (Eysenck et al., 1985), and MPQ, Behavioral Constraint subscale (Tellegen, 1982); MM[e] = I7, Venturesomeness subscale (Eysenck et al., 1985), and MPQ, Behavioral Constraint subscale, (Tellegen, 1982); M[f] = TPQ, Novelty Seeking subscale (Cloninger, 1986); M[g] = SSRT: Sensation Seeking Real-world Risk Taking, total (Swogger et al., 2010). M = over 55% male sample; F = over 55% female sample; Bal = gender balanced sample. Cau = over 50% Caucasian sample; Min = over 50% minority sample. HS = high-school sample; Und = undergraduate sample; Com = community sample; Inm = inmate sample.

DOI: 10.1002/bdm

Table 2. Studies included in the meta-analysis of the relationship between impulsivity and BART. Effect sizes, weights, continuous and categorical moderators are also reported.

	Study	ES *r*	IVW	Imp rtt	HS ES *r*	HS IVW	Age	Performance Level	Time Lag	Scale	Gender	Etnicity	Sample
1	Aklin et al. (2005)	0.23	48	0.72	0.29	30.41	14.80	26.40	3	E^a	Bal	Min	HS
2	Bornovalova et al. (2009)	0.05	37	0.84	0.06	27.35	19.80	35.22	7	E^a	Bal	Cau	Und
3	Cyders et al. (2010)	0.12	101	0.70	0.15	62.22	19.22	41.89	8	M^b	Bal	Cau	Und
4	Hunt et al. (2005)	0.08	77	0.86	0.09	58.27	18.90	35.00	3	M^c	F	Cau	Und
5	Jones and Lejuez (2005)	0.16	57	0.84	0.19	42.13	19.06	30.28	3	E^a	F	Cau	Und
6	Killgore (2007)	0.32	51	0.78	0.39	35.01	23.50	35.60	5	M^d	Bal	Cau	Com
7	Lejuez et al. (2002)	0.27	83	0.90	0.30	65.74	20.90	29.40	0	MM^e	Bal	Cau	Com
8	Lejuez, Aklin, Jones, et al. (2003a)	0.01	57	0.84	0.01	42.13	20.10	39.80	1	E^a	Bal	Cau	Und
9	Lejuez, Aklin, Zvolensky, et al. (2003b)	0.21	23	0.84	0.24	17.00	15.23	33.00	1	E^a	F	Min	HS
10	Lejuez et al. (2005)	0.13	122	0.75	0.16	80.52	15.10	32.48	3	E^a	F	Min	HS
11	Lejuez et al. (2007)	0.07	95	0.72	0.09	60.19	14.80	35.10	5	E^a	Bal	Min	HS
12	Lorian and Grisham (2010)	0.33	52	0.81	0.39	37.07	20.24	37.23	8	M^f	F	Min	Und
13	Marini and Stickle (2010)	0.04	147	0.58	0.06	75.03	15.10	38.64	8	M^g	M	Cau	Inm
14	Mishra et al. (2010)	0.10	112	0.78	0.12	76.88	20.30	33.00	8	E^a	Bal	Cau	Und
15	Reynolds et al. (2006)	0.07	67	0.90	0.08	53.06	22.92	35.60	4	MM^h	Bal	Cau	Und
16	Romer et al. (2009)	0.02	384	0.62	0.03	209.51	11.40	24.60	7	E^a	Bal	Cau	HS
17	Skeel et al. (2008)	0.01	97	0.87	0.01	74.26	19.22	36.20	6	M^i	F	Cau	Und
18	Upton et al. (2011)	0.11	95	0.78	0.12	78.71	20.20	42.50	9	E^a	Bal	Cau	Und
19	Weafer et al. (2011)	0.10	18	0.78	0.11	14.91	22.30	44.10	9	M^j	M	Cau	Com

ES *r* = Effect Size *r*; IVW = Inverse Variance Weight; IMP rtt = Impulsivity Reliability used for correction; HS ES *r* = Hunter–Schmidt Effect Size *r*; HS IVW = Hunter–Schmidt Inverse Variance Weight. E^a = Eysenck Impulsiveness Scales, Impulsivity subscale (Eysenck et al., 1985); M^b = UPPS Impulsive Behavior Scale (Whiteside & Lynam, 2001); M^c = BIS-10 Barratt Impulsiveness Scale, total score (Barratt, 1985); M^d = EVAR Evaluation of Risks Scale, Impulsivity subscale (Killgore et al., 2006); MM^e = I7 Impulsivity subscale score (Eysenck et al., 1985) and BIS-11 Barratt Impulsiveness Scale, total score (Patton et al., 1995); M^f = BIS/BAS Behavioral Inhibition System subscale (Carver & White, 1994): Mg = APSD Antisocial Process Screening Device, Impulsivity factor score (Frick & Hare, 2001); MM^h = BIS-11 Motor Impulsiveness subscale (Patton et al., 1995) and I7 Impulsivity subscale (Eysenck et al., 1985); M^i = TPQ Harm Avoidance subscale (Cloninger, 1986); M^j = Delay Ocular Return Task (Godijn & Theeuwes, 2003). M = over 55% male sample; F = over 55% female sample; Bal = gender balanced sample. Cau = over 50% Caucasian sample; Min = over 50% minority sample. HS = high-school sample; Und = undergraduate sample; Com = community sample; Inm = inmate sample.

DOI: 10.1002/bdm

referred to as HS, by correcting each study's effect size and its weight for the unreliability of both variables. Where a study used a single personality measure, reliability estimates declared by the authors for that study were used. Where a study used multiple measures, a composite reliability was estimated according to Hunter and Schmidt (2004). Where information about reliability was not available, the average reliability resulting from the whole dataset was used as our best guess. Likewise, for BART, whose reliability has not been directly assessed in most studies, we considered the average test–retest reliability ($r_{tt} = .82$) reported by Swogger et al. (2010, p. 444).

The standard IV method and the HS method are both appropriate for meta-analyses of correlations. However, they have different goals. Whereas the IV method estimates the pooled effect across studies "as it was measured," the HS method tries to estimate the strength of the underlying relationship under ideal conditions, such as for instance perfect reliability (Wilson, 2012). As the focus of the present review is investigating personality–risk relation at the construct level as well as at the level of its empirical indicators, we reported effect sizes obtained by both methods.

Effect size difference

Because we estimated two separate effect sizes in this review, a logical question to be answered is whether the sensation seeking and the impulsivity relations with risk taking were statistically different. In keeping with Cohen (1988, pp 109–143), one can compute a statistic to assess the magnitude of the difference between two effect sizes based on the difference between two *z*-transformed correlation coefficients (i.e., the Cohen's *q*). However, this formula only provides an effect size estimate itself, which can be only appraised as small, medium, or large, whereas it is not possible to assess its statistical significance. In addition, it is worth noting that the Cohen's *q* does apply only to the comparison of independent effect sizes, such as those based on different levels of a moderator variable in a single dataset. By contrast, in the present review, we analyzed two separate datasets, which share some degree of dependency as there were *K* common studies of correlated personality traits. Hence, using the Cohen's *q* formula as well as comparing the confidence intervals of each effect size estimates could be misleading.

Meng, Rosenthal, and Rubin (1992) proposed a method for comparing correlated correlation coefficients, such as the case for the aggregated effect sizes resulting from the sensation seeking and impulsivity dataset. To briefly summarize, we adjusted the difference between two *z*-transformed correlations by taking into account the average intercorrelation of sensation seeking and impulsivity and sample size. In so doing, we first selected common studies and tested whether this selection status moderated the effect size estimates. As this test was negative, we estimated the correlation between impulsivity and sensation seeking as the weighted average correlation in the selected studies. Finally, we adjusted the estimated effect size difference and formally tested its statistical significance according to Meng et al. (1992, Equations 1 and 4).

Heterogeneity and moderator analysis

Significant heterogeneity is not a prerequisite for conducting a moderator analysis (e.g., Rosenthal & Di Matteo, 2000). Thus, we have carried out a set of analyses to investigate whether some of the study characteristics (also coded in Table 1) might have affected the relation of personality traits with risk taking in some systematic way. As there were continuous moderators, such as participants' age, BART performance level, and publication year, as well as categorical moderators, such as gender, ethnicity, type of population, and type of personality scale, a twofold approach was used. We carried out separate meta-regressions with effect size as the dependent variable for continuous predictors, whereas we used dummy variables to test possible contrasts for different levels of categorical moderators. In both cases, the statistical significance of each moderator variables was tested on the basis of the *Q statistic*. Details of computational procedures are provided by Lipsey and Wilson (2001, pp 133–142).

Publication bias

The possibility that the conclusions of this review might have been affected by publication bias was a main concern of this study. First of all, the classical Rosenthal's (1979) fail-safe *N* was calculated. This index estimates how many unpublished studies with a null effect size would be necessary to turn a significant population effect size estimate into a nonsignificant one based on the *Stouffer Z-test*. However, as the fail-safe *N* is an absolute measure of publication bias, its size depends on how many *K* studies are included in the meta-analysis. Rosenthal (1979) recommended the fail-safe *N* to be smaller than a $5K + 10$ benchmark, whereas more recently, Mullen, Muellerleile and Bryant (2001, p 1454) recommended the fail-safe ratio $N/(5K + 10)$ to be greater than 1 to rule out a publication bias. In our review, the critical values $5K + 10$ were 100 and 105 for sensation seeking and impulsivity, respectively.

Criticism of the fail-safe *N* led meta-analysts to develop a somewhat different approach to examine publication bias. The so-called "funnel plot" is a graphical technique in which the standard error of each study's effect size is plotted against the standardized effect size itself. Lack of publication bias is demonstrated by a symmetrical cloud of studies centered around the population effect size, with increasing variability at increasing levels of standard error. This is because there should be about as many studies providing no significant results as those providing significant ones at each specific level of standard error, whereas studies with smaller standard errors should also be closer to the population effect size. However, because the "funnel plot" interpretation might be difficult, Field and Gillett (2010) suggested that the rank-order correlation between the effect size and its associated standard error be calculated. In fact, as far as there is symmetry in the plot, the resulting coefficient is rather small and not statistically significant.

Field and Gillett (2010) also provided R-code implementing a sensitivity analysis according to Vevea and Woods (2005). This method tested to what extent different types of publication bias might have affected the aggregate effect size (i.e., one-tailed or two-tailed bias and severe or moderate bias). In the

DOI: 10.1002/bdm

present review, we considered only the one-tailed types of biases (both severe and moderate) because we aimed at modeling either a severe or moderate bias against the publication of results disconfirming our initial hypotheses.

RESULTS

As a preliminary analysis, we assessed how the average adjusted pump varied between the selected studies. As stated in the introduction, a performance level higher or lower than the maximizing earnings point (i.e., 64 pumps) can separate studies in which the majority of respondents made predominantly risk-disadvantageous or risk-advantageous choices. This notion fits well with the idea that different motivational processes differentiated sensation seeking risk taking from impulsive risk taking. The former type of risk is supposed to be more associated with arousal resulting from repeatedly taking risk-advantageous choices; the latter type is expected to be more associated with neglecting loss of reward signals when repeatedly making risk-disadvantageous choices. Unfortunately, the BART performance level in the 22 studies included in this review was largely below the maximizing earnings point, as it varied from 24.60 to 44.10 (with a weighted SD = 5.93). The aggregated effect sizes for performance level were 35.60 (SE = 0.28) and 33.06 (SE = 1.34) for random effect and fixed effect estimates, respectively. This finding showed that there were no studies in which a majority of participants made consistent risk-disadvantageous choices from a normative point of view. Differently, one can think of 35.60 or of 33.06 as the average performance level for research participants for which the effects of sensation seeking or impulsivity on BART were not affected by experimental manipulations or clinical conditions. Because there was some degree of heterogeneity for performance level in the reviewed studies ($Q = 445.36$; df = 21; $p < .001$), we also examined whether other moderators coded for this review might have accounted for between-study variability. Only the sample's mean age affected the average performance level, with studies composed of older adolescents or young adults resulting in a relatively higher but still suboptimal performance than studies of preadolescents ($Q = 111.36$; df = 1; $p < .001$).[1]

Next, we examined the aggregated effect size for the relation of sensation seeking and impulsivity with BART performance level (Tables 1 and 2) to test the study's main hypothesis. Recall that we expected personality traits considered as determinants of risk taking in real-world situations to have a consistent main effect on risk taking in modern escalating risk tasks. The answer was overall positive for both sensation seeking and impulsivity, as we found two aggregated effect sizes significantly different from zero. However, as we appraise effect sizes according to the well-established Cohen's criteria, the relation of sensation seeking with risk taking was in the small–moderate range ($\bar{r} = .14$; $CI_{95\%} = .09–.18$; $Z = 5.88$; $p < .001$), whereas the relation of impulsivity with risk taking was just around the "small" effect size threshold ($\bar{r} = .10$; $CI_{95\%} = .05–.15$; $Z = 3.85$; $p < .001$). In both cases, the Q values were not statistically significant ($Qs = 14.11$ and 13.53 for sensation seeking and impulsivity, respectively), thereby showing that there was a good fit between the fixed effect model and the data as well as that the inclusion criteria applied to the studies yielded a homogeneous set of personality studies.[2]

The results provided so far were based on the pooled effect size "as it was measured." The HS method estimates the strength of the underlying relations under ideal measurement conditions, that is what the aggregated effect size would have been if all studies had been free of methodological artifacts, such as unreliability of psychometric scales. Although there was a general improvement in both effect sizes, the results obtained by the HS method mirrored those reported earlier. Again, the aggregated effect sizes for the relation of sensation seeking and impulsivity with risk taking were in both cases in the small–moderate range, with the sensation seeking effect size being relatively larger than the impulsivity effect size ($\bar{r} = .16$; $CI_{95\%} = .11–.22$; $Z = 5.92$; $p < .001$ and $\bar{r} = .12$; $CI_{95\%} = .06–.18$; $Z = 3.96$; $p <. 001$). Also in this analysis, there was no significant heterogeneity of effect sizes. Hence, the unreliability of psychometric methods was not a likely account for inconsistent results of BART and personality traits.

We followed up with our investigation of study characteristics that might have potentially affected the relation of risk taking with personality traits. For the sensation seeking dataset (Table 1), there was a significant effect for type of population studied ($Q = 3.95$; df = 1; $p < .05$), with studies carried out on community samples showing a larger effect size than studies carried out on student samples (both high school and undergraduates). In addition, participant's age and ethnicity approached the conventional levels of statistical significance ($Qs = 2.47$ and 2.60; dfs = 1; $ps = .08$ and .07), showing that the effect size trend tended to increase with participant's age and with a greater representation of ethnic minorities in the study.

For the impulsivity dataset (Table 2), the meta-regression equation for participant's age as moderator was statistically significant ($Q = 3.48$; df = 1; $p < .05$), showing that the effect size trend for impulsivity increased with participant's age as it did for sensation seeking. Likewise, there was a significant moderating effect for type of population studied ($Q = 4.55$; df = 1; $p < .05$), again with an increasing effect size resulting from studies of community samples relative to student samples. Despite recent literature indicating that men are more

[1]Although there was a tendency for studies based on predominantly male gender samples ($K = 3$) to result in a higher performance level than studies based on predominantly female gender samples ($K = 7$), this gender gap was not statistically significant. Likewise, neither ethnicity composition nor the type of population accounted for a significant portion of performance level heterogeneity. As we moved forward to investigating continuous moderators, we found a significant effect of the sample's mean age on performance level, with older ages also taking more risk on BART.

[2]As a result of the low effect size heterogeneity, the random effect variance component in the random effect model provided negative estimates, and it was set to zero. Consequently, the random and the fixed effect model produced the same effect size estimates, and the fixed effect model was chosen because of its parsimony (Lipsey & Wilson, 2001, p 120).

DOI: 10.1002/bdm

apt to take risks and to self-report higher sensation seeking and impulsivity than women (e.g., Cross et al., 2011; de Haan et al., 2011), participant's gender did not moderate the relation of personality with risk taking on BART. Likewise, no significant moderating effect of ethnicity was detected.

One important moderator to be considered in this review is the publication year of each study, as its statistical significance might be helpful in detecting a historically decreasing trend of effect size. It is well documented that studies providing significant findings get published in an average shorter time than studies providing null or not significant results, and this "time lag" may severely bias a systematic review like ours (Hopewell, Clarke, & Mallett, 2005). The meta-regression equation for publication year was not statistically significant for both sensation seeking ($Q = 2.30$; df = 1; $p = .10$) and impulsivity ($Q = 1.58$; df = 1; $p = .15$), thus showing that the effect size trend tended to decrease over time only to a very limited extent.

Finally, we concluded our review by assessing whether and to what extent a publication bias possibly affected the results reported so far. In fact, there might still be unpublished studies reporting findings contrary to our hypotheses that may threaten the validity of our conclusions. The fail-safe *N*s were 228 and 118 for the sensation seeking and the impulsivity effect sizes, respectively. In both cases, these values were above Rosenthal's critical numbers of 100 and 105, respectively. More importantly, the fail-safe ratio was greater than 1.00 in both cases (i.e., 2.28 and 1.12), thereby showing that the "weight of evidence"—reported in this review—"does appear sufficiently tolerant for future results" (Mullen et al., 2001 p.1454).

The funnel plots showed an approximately symmetrical cloud of studies centered around the population effect size. The funnel shaped cloud was, however, more evident for the sensation seeking effect size (Figure 1(a)), whereas it was slightly skewed for the impulsivity effect size (Figure 1 (b)). In addition, the rank-order correlation between the effect size and its associated standard error was not statistically significant (rs = .14 and .27, ps = .82 and .10, for sensation seeking and impulsively, respectively). As a final test of publication bias, we carried out a sensitivity analysis hypothesizing both moderate and severe one-tailed selection of studies (Vevea & Woods, 2005). The effect size for sensation seeking was empirically robust as it never fell below the "small" effect size threshold even if a severe publication bias was present (Raw $\bar{r} = .14$; Moderate bias $\bar{r} = .12$; Severe bias $\bar{r} = .10$). For impulsivity, the effect sizes tended to fall below the small threshold even for moderate one-tailed selection bias (Raw $\bar{r} = .10$; Moderate bias $\bar{r} = .08$; Severe bias $\bar{r} = .06$). With these analyses, we concluded that a publication bias seems a very unlikely threat to the effect size estimates for sensation seeking and risk taking, whereas some concern might arise for impulsivity, although it is worth remembering that different tests of publication bias were in most cases negative.

All analyses conducted so far showed that the sensation seeking effect size was larger than the impulsivity effect size. Thus, a logical question was whether these effect sizes were also statistically different. Although the Cohen's q was small and the confidence intervals were in most cases overlapping (both for the IV and the HS analyses), we could not rule out the possibility that sensation seeking and impulsivity had a significantly different effect size as both ways to look at significant differences are suitable for comparing statistically independent effect sizes, only. It is worth noting that we have built two separate datasets to control for statistical dependency, which is due to a common number of research participants ($N = 1526$) as well as to the average positive intercorrelation between sensation seeking and impulsivity measures ($r = .36$ across 15 studies). A more conservative test for the effect size difference was indeed based on Meng et al. (1992), who published equations for comparing correlated correlation coefficients. Also, in this case, we could not reject the null hypothesis ($Z = 1.45$). Therefore, although different in their range of interpretability, the two effect sizes were not statistically different when we formally tested this hypothesis.

These latter results seem to support the hypothesis that the effect size for sensation seeking cannot be quantitatively distinguished from the effect size for impulsivity, although some qualitative differences emerged. Recall that theoretical developments in the personality literature suggested that there is a particular type of "impulsive, unsocialized sensation

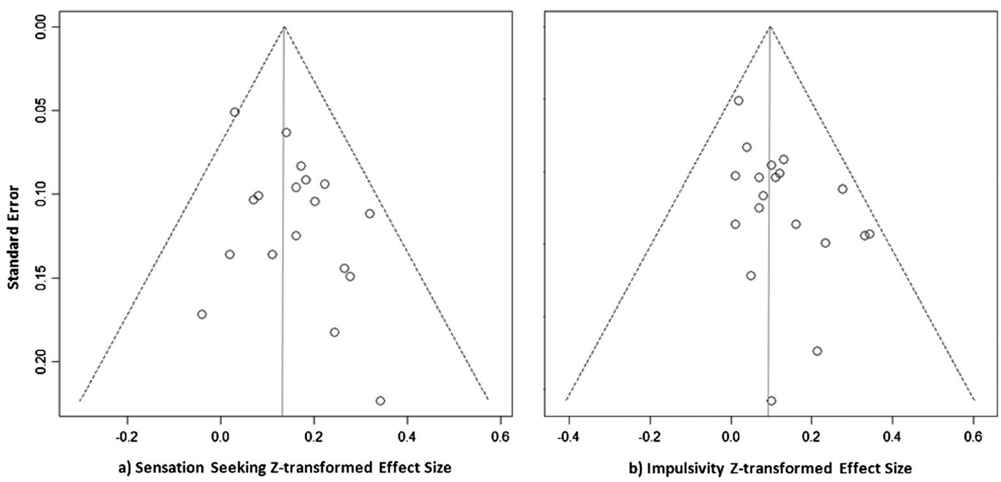

Figure 1. Funnel plots for (a) sensation seeking and (b) impulsivity studies. Lack of publication bias is demonstrated by a symmetrical cloud of studies centered around the population effect size, with increasing variability at increasing levels of standard error

DOI: 10.1002/bdm

seeking" (e.g., Joireman & Kuhlman, 2004), which might account for different types of risk besides one's involvement in recreational risky activities for the mere sake of arousal. Unfortunately, only two studies of BART actually used a composite ImpSS score to predict risk taking (e.g., Bornovalova et al., 2009; Lejuez et al., 2005).

To guess how large an ImpSS effect size might be, we have calculated a composite correlation of sensation seeking and impulsivity according to Hunter and Schmidt equations (2004) removing—where it was possible—the Thrill and Adventure Seeking score of the Zuckerman Sensation Seeking Scale, which is presumed to be a measure of nonimpulsive socialized sensation seeking (e.g., Glicksohn & Abulafia, 1998). The resulting estimates were analyzed, and the results were very close to those obtained for the sensation seeking dataset, but with a lesser number of studies ($K = 15$) ($\bar{r} = .13$; $CI_{95\%} = .08–.18$; $Z = 4.91$; $p < .001$ and $\bar{r} = .16$; $CI_{95\%} = .09–.22$; $Z = 4.75$; $p < .001$, for IV and HS, respectively).

DISCUSSION

Some excellent reviews recently concluded that emphasis on the role played by decision-maker stable characteristics is at best "oversized" considering the little consensus on the significance and the interpretation of the results of existing empirical studies (Appelt et al., 2011). We indeed devised the present review to show that contradictory evidence in the literature might be due at least in part to the unsystematic association of different experimental tasks and conditions with different personality constructs, which made it difficult to retrieve in the literature a coherent set of studies upon which to generate reliable conclusions. In addition, we hypothesized that a careful selection of meaningful personality variables might have increased the chance for detecting significant relations of personality with risk taking in risky decision making.

To test our hypotheses, we systematically reviewed a selected sample of studies, including only those meeting with the following directives. First, we held constant the variety of experimental tasks and conditions by narrowing our interest to empirical studies of risk taking based on a modern, popular, and specific task of escalating risks, which has a 10-year history of use. This strategy helped reduce the effect on risk taking attributable to the variety of situational or experimental factors manipulated in decision-making research, while still counting a relatively large number of studies, thereby providing a greater chance of detecting smaller but consistent personality effects on risk taking in experimental tasks (Mohammed & Schwall, 2009). Second, we reduced the variety of personality constructs potentially associated with risk taking by including in this review only studies of sensation seeking and impulsivity, which are commonly believed to influence real-world risk taking (e.g., Steinberg et al., 2008).

Borrowing from Appelt et al. (2011), we strongly believe that only measures with a theoretical tie with risky decision making are likely to result in consistent findings both inside and outside the laboratory setting. Thus, for instance, it has been documented that modern sequential risk tasks, such as BART, are able to trigger arousal (e.g., Figner et al., 2009; Schonberg et al., 2010), which is a specific task characteristic motivating risk-taking behavior in sensation seekers. In addition, sequential risk tasks are also able to model another typical characteristic of real world behaviors, in which risk taking is rewarded up to a point beyond which taking risk is likely to result in diminishing returns and increasing potential losses. This feature is closely tied to impulsivity theory (Enticott & Ogloff, 2006; Roberti, 2004), which posits that impulsive risk taking is likely due to failure to inhibiting dangerous reward-seeking behaviors, such as making a sequence of risk-disadvantageous choices.

With a total of 2120 participants taking BART with different personality measures and distributed across 22 different studies in the two personality domains, we concluded that the sensation seeking trait was associated with risk taking in BART with an effect size in the small–moderate range, whereas the effect size assessed for impulsivity was just around the small effect size threshold. It should be kept in mind, however, that predicting behavior from personality ratings is highly problematic, and it is beyond the aims of this paper to interpret our findings in terms of forecasting a specific behavior, such as risk taking on a single balloon task. Therefore, whatever the effect size of personality with BART, there might be other domains of risk in which this prediction is not warranted.

Nevertheless, the findings of this study have some implications for both the decision-making and the personality literature. First, they help reconcile inconsistencies in the reviewed studies by showing that sampling variability and low statistical power—rather than lack of theoretical ties or measurement unreliability—were the most likely causes of contradictory findings. For instance, if one compares the sensation seeking effect size with BART, with the effect size typically resulting from a reliable risky choice framing manipulation, such as the Asian Disease problem (e.g., Kuhberger, 1998, p. 38), one can see that the latter effect size is about twice that of the former one (Cohen's $d_{\text{sensation seeking}} = 0.28$ vs. Cohen's $d_{\text{asian disease}} = 0.57$). Thus, the question is not whether individual differences matter as personal determinants of risk taking in experimental tasks, but rather to what extent they are able to predict risky choices when compared with manipulation of experimental factors (see also Mohammed & Schwall, 2009).

The second implication of the present positive findings is the demonstration that if one capitalizes on a consistent body of literature and if one selects personality variables on a sound theoretical base, then a consistent main effect of personality on risky decision making can be found in experimental tasks as well as in real-world risky behaviors (Appelt et al., 2011; Mohammed & Schwall, 2009). Unlike hypothetical gambles, modern risk tasks, such as BART, involve elements of arousal (e.g., Schonberg et al., 2010), which is also considered one of the motivating variables for sensation seekers taking recreational risks (e.g., de Haan et al., 2011). Hence, our cumulative analysis of BART studies provided positive evidence for the hypothesis that when personality psychologists and behavioral decision scientists are able to properly select tasks and variables, consistent relations can be found in both areas.

DOI: 10.1002/bdm

Third, the effect size estimates for sensation seeking and impulsivity, although not statistically different, were, however, qualitatively different. Not only was there a consistent tendency for sensation seeking to systematically result in slightly higher estimates than impulsivity, but also the former effect size was more empirically robust when publication bias was taken into account. We interpreted these findings in the light of our analysis of BART performance level. As performance levels above the maximizing earning point indicated risk-disadvantageous choices, such studies might have uncovered a more tight association with impulsivity in that they reflected a neglect of loss of reward signals or just needless pursuit of gains. As in all of the reviewed studies, the average behavior was largely below the normatively expected earning point (i.e., 35.60 ± 5.93 vs. 64); this might have limited the association of high trait impulsivity with risk-disadvantageous choices.

Another possible interpretation for qualitatively different results can be that the association of impulsivity with suboptimal risk taking might be due to its relatively large correlation with sensation seeking ($r = .36$ across 15 studies). Thus, although we could not detect an association of impulsivity with disadvantageous risk taking, there still might be in the selected studies a "trace" of an impulsivity effect, mediated by its association with sensation seeking. In fact, as we combined impulsivity and sensation seeking in a single dataset; the resulting effect size was as large as those assessed for sensation seeking alone. We should, however, acknowledge that to study the effect of "normal" personality on risky decision making, we excluded studies in which BART performance was affected by experimental manipulations (e.g., sleep deprivation and consumption taking medication) or by clinical conditions (e.g., alcoholism, psychoactive drug use and psychiatric diagnosis). As the association of impulsivity with real-world risky behaviors was mostly supported by studies of special populations (e.g., psychiatric patients or socially deviant individuals) taking severe health, social, and financial risks (e.g., Adams & Moore, 2007; Chambers & Potenza, 2003; Dahlen et al., 2005; De Wit, 2009; Dunlop & Romer, 2010; Frankenberger, 2004; Hittner & Swickert, 2006; Hoyle et al., 2000; Johansson et al., 2009; Nelson et al., 2008), this might explain why the effect size for impulsivity with risk taking was qualitatively different from the sensation seeking one.

Beyond effect size, the analysis of data also revealed that the selected sample of studies was highly homogeneous in terms of personality–risk relations. Although this result provides a post hoc answer to the criticism of combining different kind of studies in meta-analysis (Lipsey & Wilson, 2001; Field & Gillett, 2010), it limited the likelihood of disclosing systematic effects of moderator variables when formally testing effect size differences. We should, however, acknowledge that—although studies based on the same sample of participants were counted only once—the reviewed studies were not completely independent, as about half were produced by a single research team. This byproduct of the current state of the literature might have indeed deflated the dataset heterogeneity. In spite of this, we detected some effects when contrasting specific levels of moderator variables. In particular, the reviewed studies, which covered an age range from 11 to 23 years old, yielded a relatively larger effect size for the older ages. In other words, sensation seeking and impulsivity tendencies were both more likely to influence the behavioral expression of risk taking in escalating risk tasks during middle–late adolescence and young adulthood than during early adolescence or preadolescence. As Steinberg et al. (2008) pointed out, personality scales tapping into sensation seeking are of a limited utility for assessing this trait during preadolescence, thereby suggesting a methodological account for the moderating effect of age. However, we also interpreted this finding in the light of recent developmental findings. In fact, age differences in sensation seeking tendencies are linked to pubertal maturation and develop according to a curvilinear pattern, with a rapid increase between 10 and 15 years and a slower decline thereafter. By contrast, impulsive tendencies tend to decline linearly from childhood to the adult age (Steinberg et al., 2008). Thus, whereas during early adolescence, there is relatively lower sensation seeking and relatively higher impulsivity, during middle adolescence, both impulsivity and sensation seeking are relatively higher compared with later ages, thereby increasing one's vulnerability to risk. Likewise, Weller, Levin and Denburg (2011) documented a similar decline of risk preferences to achieve a gain in an experimental decision task designed to assess risk attitude from the preschool years to adult age. In addition, as recently demonstrated by Gardner and Steinberg (2005), peer influences can more easily elicit risk seeking choices for adolescents (13–16 years old) than for youths (18–22 years old) or adults (over 24 years old).

Taken together, one can conclude from these studies that the moderating effect of age on personality–risk relations reflected the common decline of sensation seeking, impulsive, and risk-taking tendencies during late adolescence and adulthood, compared with childhood and preadolescence. That people develop more mature or stable risk (or arousal) preferences over time is widely documented by personality studies of recreational risks (e.g., Adams & Moore, 2007; Dunlop & Romer, 2010; Dahlen et al., 2005; Hittner & Swickert, 2006; Johansson et al., 2009; Nelson et al., 2008; Frankenberger, 2004).

A somewhat unexpected finding resulting from our review is the absence of any moderating effect due to decision-maker gender, whereas the issue of gender differences in sensation seeking, impulsivity, and risk taking is a topic of long-standing interest. For instance, Cross et al. (2011) concluded that men obtain significantly higher scores on sensation seeking scales as well as on behavioral risk tasks. Likewise, consistent gender differences emerged in large population studies using surveys and personality scales (de Haan et al., 2011; Dohmen et al., 2011). With the studies included in the present review, we concluded that the correlation of personality with risk taking on BART was the same for both men and women. Thus, for instance, although men are acknowledged as being greater risk takers and sensation seekers than women, the correlation between sensation seeking and risk taking was about the same within each gender, that is, men or women higher on sensation seeking were

DOI: 10.1002/bdm

more apt to take risks than their same gender counterparts lower on this trait. As a limitation to the present findings, we should however acknowledge that gender was balanced in the majority of the reviewed studies and this experimental control might have undermined gender moderation.

The findings of the present review also have important implications for personality psychology as they suggest that risk taking at least on this type of measure may be associated with both impulsivity and sensation seeking. In keeping with Zuckerman and Kuhlman (2000), we not only replicated a statistical relation of specific facets of the two traits, as it was evident from intercorrelations and factor analyses (Joireman & Kuhlman, 2004) but also reinforced the view that aspects of the so-called impulsive unsocialized sensation seeking had the same positive effect when a behavioral risk task is used as criterion for validity generalization.

In terms of the practical implications of the present review, we may index the following considerations. On the one hand, the moderator analyses suggest that its utility may be stronger for older adolescents and adults than for younger adolescents, suggesting also that risk taking may be influenced more by contextual factors than personality among the youngest group. Given that our analyses suggest a relatively independent link of sensation seeking, impulsivity, and BART scores on real-world risky behaviors, it will be very important to understand both the aspects of personality and risk-taking propensity on BART that are overlapping and unique when considering the complex relationship of these variables and their relationship to real-world risk taking. On the other hand, the results of our review seem to exclude the possibility that BART might be efficiently used as a proxy of an impulsive personality, as it is believed by some researchers (see Cross et al., 2011, for a review of types of impulsivity measures). Rather, it looks like BART, as other modern escalating risk tasks with a shorter history of use (e.g., Pleskac, 2008; Figner et al., 2009; Gardner & Steinberg, 2005), may add to self-reported and behavioral impulsivity to reconcile findings from behavioral decision-making and personality studies (Meda et al., 2009).

Before discussing the major implications for research on individual differences in risky decision making, it is worth noting that the effect sizes assessed in the present study were robust to time-lag bias as well as to publication bias, which is a documented tendency for studies reporting nonsignificant findings to be published later in peer-reviewed journals and/or just not published at all (Hopewell et al., 2005). The decision to include in this review only studies assessing sensation seeking and impulsivity with an incremental risk task with a 10-year history of use allowed us to assess and eventually rule out the possibility that studies with significant personality–risk relations were overrepresented in this review. This does not mean that the BART is considered in this review as the "Gold Standard" for modern behavioral research on antecedents and consequences of risk taking; rather, we exploited the relative popularity of this task to gather a relatively large sample of studies. In fact, the reliable association of personality traits with risk taking is in our review limited to risk taking in tasks that can offer high emotional arousal to research participants and (perhaps) may offer in principle a chance for disadvantageous risk-taking choices. Less is known about other potentially reliable associations of other personality traits with different tasks, such as the Columbia Card Task (Figner et al., 2009), which can dissociate "hot" and "cold" risky decisions. For instance, a recent study from our laboratory (Panno, Lauriola, & Figner, 2013) demonstrated that stable individual differences in emotion regulation strategies were able to predict risk taking in deliberative decision making, but the relatively recent publication of the CCT and the novelty of the emotion regulation paradigm in risky decision making limited the possibility of generalization of this latter conclusion through a meta-analytic inquiry.

Along with the excellent guidelines suggested in earlier literature reviews (Appelt et al., 2011; Mohammed & Schwall, 2009), as a final remark, we strongly recommend that decision researchers study the effects of individual differences on samples providing the appropriate statistical power to detect small but consistent personality effects on risk taking as well as on other decision-making phenomena. For instance, according to Cohen (1988), to detect a small–moderate correlation (e.g., $r = 14$) as with the sensation seeking–risk relation, a sample of about 270 cases will provide 80% power to discover that the correlation is statistically significant at the 0.05 level. Hence, studies of individual difference and decision making should adhere to the excellent existing guidelines (Appelt et al., 2011; Mohammed & Schwall, 2009) and also should seriously take into account the sample size and power issues. If all these issues are addressed, we should see more compelling evidence of the role of systematic individual differences in risky decision making as well as in personality and developmental psychology.

ACKNOWLEDGMENTS

The authors discussed the contents of this article together. Marco Lauriola and Irwin Levin elaborated on the theoretical framework and the research hypotheses. Angelo Panno and Marco Lauriola devised the methodological content and analyzed the data. Carl Lejuez provided a significant contribution to the interpretation of the results of the reported study. The final version of the manuscript was written by M. Lauriola, A. Panno, I. P. Levin, and C. W. Lejuez. The authors would like to thank the following colleagues: Ashely Acheson, Brady A. Reynolds, Daniel Upton, Donald A. Hantula, Glahn C. David, Harriet de Wit, Heather A. Jones, Jessica Weafer, Julie C. Stout, Lara Ray, Laura Betancourt, Laura MacPherson, Marc T. Swogger, Mark T. Fillmore, Martin A., Melissa A. Cyders, Reid L. Skeel, Sandeep Mishra, Schepis S. Ty, Scott Coffey, Steven J. Robbins, Timothy R. Stickle, Thomas Hills, Tim J. Pleskac, Will M. Aklin, Wlliam D. S. Killgore, and Zacny James for providing information or data; Maja Osolnik for helping the implementation of data analyses by SPSS programming; and Frank L. Schmidt and David B. Wilson for helpful advice concerning their meta-analysis methods.

DOI: 10.1002/bdm

REFERENCES

Acheson, A., Richards, J. B., & de Wit, H. (2007). Effects of sleep deprivation on impulsive behaviors in men and women. *Physiology & Behavior*, *91*, 579–87.

Acheson, A., & de Wit, H. (2008). Bupropion improves attention but does not affect impulsive behavior in healthy young adults. *Experimental and Clinical Psychopharmacology*, *16*, 113–23.

Adams, T., & Moore, M. (2007). High-risk health and credit behavior among 18- to 25-year-old college students. *Journal of American College Health*, *56*, 101–108.

*Aklin, W. M., Lejuez, C. W., Zvolensky, M. J., Kahler, C. W., & Gwadz, M. (2005). Evaluation of behavioral measures of risk-taking propensity with inner city adolescents. *Behavioral Research Therapy*, *4*, 215–228.

Appelt, K. C., Milch, K. F., Handgraaf, M. J. J., & Weber, E. U. (2011). The decision making individual differences inventory and guidelines for the study of individual differences in judgment and decision-making research. *Judgment and Decision Making*, *6*, 252–262.

Arnett, J. (1994). Sensation seeking: A new conceptualization and a new scale. *Personality and Individual Differences*, *16*, 289–296.

Baker, F., Johnson, M. W., & Bickel, W. K. (2003). Delay discounting in current and never-before cigarette smokers: Similarities and differences across commodity, sign, and magnitude. *Journal of Abnormal Psychology*, *112*, 382–392.

Barratt, E. S. (1985). Impulsiveness subtraits: Arousal and information processing. In J. T. Spence & C. E. Itard (Eds.), *Motivation, emotion, and personality* (pp. 137–146). Amsterdam: Elsevier.

*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.

Bickel, W. K., Odum, A. L., & Madden, G. J. (1999). Impulsivity and cigarette smoking: delay discounting in current, never, and ex-smokers. *Psychopharmacology*, *146*, 447–454.

Bishara, A. J., Pleskac, T. J., Fridberg, D. J., Yechiam, E., Lucas, J., Busemeyer, J. R., et al. (2009). Similar processes despite divergent behavior in two commonly used measures of risky decision making. *Journal of Behavioral Decision Making*, *22*, 435–454.

Bornovalova, M. A., Daughters, S. B., Hernandez, G. D., Richards, J. B., & Lejuez, C. W. (2005). Differences in impulsivity and risk-taking propensity between primary users of crack cocaine and primary users of heroin in a residential substance-use program. *Experimental and Clinical Psychopharmacology*, *13*, 311–318.

*Bornovalova, M., Cashman-Rolls, A., O'Donnell, J., Ettinger, K., Richards, J., deWit, H., et al. (2009). Risk-taking differences on a behavioral task as a function of potential reward/loss magnitude and individual differences in impulsivity and sensation seeking. *Pharmacology Biochemistry and Behavior*, *93*, 258–262.

*Bornovalova, M. A., Gwadz, M., Kahler, C. W., Aklin, W. M., & Lejuez, C. W. (2008). Sensation seeking and risk-taking propensity as mediators in the relationship between childhood abuse and HIV-related risk behavior. *Child Abuse & Neglect*, *32*, 99–109.

Boyer, T. W. (2006). The development of risk-taking: A multi-perspective review. *Developmental Review*, *26*, 291–345.

Carver, C. S, & White, T. L. (1994). Behavioral inhibition, behavioral activation, and affective responses to impending reward and punishment: The BIS/BAS scales. *Journal of Personality and Social Psychology*, *67*, 319–333.

Chambers, R. A., & Potenza, M. N. (2003). Neurodevelopment, impulsivity, and adolescent gambling. *Journal of Gambling Studies*, *19*, 53–84.

Cloninger, C. R. (1986). A unified biosocial theory of personality and its role in the development of anxiety states. *Psychiatric Developments*, *3*, 167–226.

Coffey, S. F., Schumacher, J. A., Baschnagel, J. S., Hawk, L. W., & Holloman, G. (2011). Impulsivity and risk-taking in borderline personality disorder with and without substance use disorders. *Personality Disorders: Theory, Research, and Treatment*, *2*, 128–141.

Cohen, J. (1988). Statistical power analysis for the behavioral sciences (2nd edn). Hillsdale, NJ: Erlbaum.

Cross, C. P., Copping, L. T., & Campbell, A. (2011). Sex differences in impulsivity: A meta-analysis. *Psychological Bulletin*, *137*, 97–130.

Crowley, T. J., Raymond, K. M., Mikulich-Gilbertson, S. K., Thompson, L. T., & Lejuez, C. W. (2006). A risk-taking "set" in a novel task among adolescents with serious conduct and substance problems. *Journal of the American Academy of Child and Adolescent Psychiatry*, *45*, 175–183.

*Cyders, M. A., Zapolski, T. C., Combs, J. L., Settles, R. F., Fillmore, M. T., & Smith, G. T. (2010). Experimental effect of positive urgency on negative outcomes from risk-taking and on increased alcohol consumption. *Psychology of Addictive Behaviors*, *24*, 367–375.

Dahlen, E. R., Martin, R. C., Ragan, K., & Kuhlman, M. M. (2005). Driving anger, sensation seeking, impulsiveness, and boredom proneness in the prediction of unsafe driving. *Accident Analysis and Prevention*, *37*, 341–348.

De Wit, H. (2009). Impulsivity as a determinant and consequence of drug use: A review of underlying processes. *Addiction Biology*, *14*, 22–31.

Dickman, S. J. (1990). Functional and dysfunctional impulsivity: Personality and cognitive correlates. *Journal of Personality and Social Psychology*, *58*, 95–102.

Dohmen, T., Falk, A., Huffman, D., Sunde, U., Schupp, J., & Wagner, G. G. (2011). Individual risk attitudes: Measurement, determinants, and behavioral consequences. *Journal of the European Economic Association*, *9*, 522–550.

Dunlop, S., & Romer, D. (2010). Adolescent and young adult crash risk: Sensation seeking, substance use propensity and substance use behaviors. *Journal of Adolescent Health*, *46*, 90–92.

Enticott, P. G., & Ogloff, J. R. P. (2006). Elucidation of impulsivity. *Australian Psychologist*, *41*, 3–14.

Eysenck, S. B. G., Pearson, P. R., Easting, G., & Allsopp, J. F. (1985). Age norms for impulsiveness, venturesomeness and empathy in adults. *Personality and Individual Differences*, *6*, 613–619.

Field, A. P., & Gillett, R. (2010). How to do a meta-analysis. *British Journal of Mathematical and Statistical Psychology*, *63*, 665–694.

Figner, B., Mackinlay, R. J., Wilkening, F., & Weber, E. U. (2009). Affective and deliberative processes in risky choice: Age differences in risk-taking in the Columbia Card Task. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*, 709–730.

Figner, B., Weber, E. U. (2011). Who takes risk when and why? Determinants of risk-taking. *Current Directions in Psychological Science*, *20*, 211–216.

Frankenberger, K. D. (2004). Adolescent egocentrism, risk perceptions, and sensation seeking among smoking and nonsmoking youth. *Journal of Adolescent Research*, *19*, 576–590.

Frick, P. J., & Hare, R. D. (2001). *Antisocial Process Screening Device*. Toronto: Multi-Health Systems.

Gardner, M., & Steinberg, L. (2005). Peer influence on risk-taking, risk preference, and risky decision making in adolescence and adulthood: An experimental study. *Developmental Psychology*, *41*, 625–635.

Glicksohn, J., & Abulafia, J. (1998). Embedding sensation seeking within the big three. *Personality and Individual Differences*, *25*, 1085–1099.

Godijn, R., & Theeuwes, J. (2003) The relationship between exogenous and endogenous saccades and attention. In: J. Hyona, R. Radach, & H. Deubel (eds), *The mind's eye: cognitive and applied aspects of eye movement research*. (pp 3–26). New York, NY: Elsevier.

Goldberg, J., & Fischhoff, B. (2000). The long-term risks in the short-term benefits: Perceptions of potentially addictive activities. *Health Psychology*, *19*, 299–303.

DOI: 10.1002/bdm

de Haan, L., Kuipers, E., Kuerten, Y., van Laar, M., Olivier, B., & Verster, J. C. (2011). The RT-18: A new screening tool to assess young adult risk-taking behavior. *International Journal of General Medicine*, *4*, 575–584.

Hanoch, Y., Johnson, J. G., & Wilke, A. (2006). Domain-specificity in experimental measures and participant recruitment: An application to risk-taking behavior. *Psychological Science*, *17*, 300–304.

Hedges, L. V., & Olkin, I. (1985). Statistical methods for meta-analysis. Orlando: Academic Press.

Hittner, J. B., & Swickert, R. (2006). Sensation seeking and alcohol use: A meta-analytic review. *Addictive Behaviors*, *31*, 1383–1401.

Holmes, K. M., Bearden, C. E., Barguil, M., Fonseca, M., Serap Monkul, E., Nery, F. G., et al. (2009). Conceptualizing impulsivity and risk-taking in bipolar disorder: importance of history of alcohol abuse. *Bipolar Disorders*, *11*, 33–40.

Hopewell, S., Clarke, M., & Mallett, S. (2005). Grey literature and systematic reviews. In H. R. Rothstein, A. J. Sutton, & M. Borenstein (Eds.), Publication bias in meta analysis: Prevention, assessment, and adjustments (pp. 48–72). West Sussex: Wiley.

Hopko, D., Lejuez, C., Daughters, S., Aklin, W., Osborne, A., Simmons, B., et al. (2006). Construct validity of the Balloon Analogue Risk Task (BART): Relationship with 59 MDMA use by inner-city drug users in residential treatment. *Journal of Psychopathology and Behavioral Assessment*, *28*, 95–101.

Hoyle, R. H., Fejfar, M. C., & Miller, J. D. (2000), Personality and sexual risk-taking: A quantitative review. *Journal of Personality*, *68*, 1203–1231.

Hoyle, R. H., Stephenson, M. T., Palmgreen, P., Pugzles Lorch, E., & Donohew, R. L. (2002) Reliability and validity of a brief measure of sensation seeking. *Personality and Individual Differences*, *32*, 401–414.

*Hunt, M., Hopko, D., Bare, R., Lejuez, C. W., & Robinson, E. (2005). Construct validity of the Balloon Analog Risk Task (BART): Associations with psychopathy and impulsivity. *Assessment*, *12*, 416–428.

Hunter, J. E., & Schmidt, F. L. (2004). Methods of meta-analysis: Correcting error and bias in research findings (2nd edn). Thousand Oaks, CA: Sage.

Johansson, A., Grant, J. E., Kim, S. W., Odlaug, B. L., & Gotestam, K. G. (2009). Risk factors for problematic gambling: a critical literature review. *Journal of Gambling Studies*, *25*, 67–92.

Johnson, J., Tellis, G. J., & Macinnis, D. J. (2005). Losers, winners, and biased trades. *Journal of Consumer Research*, *32*, 324–329.

Joireman, J., & Kuhlman, D. M. (2004). The Zuckerman–Kuhlman personality questionnaire: origin, development, and validity of a measure to assess an alternative five-factor model of personality. In R. M. Stelmack (Ed.), On the psychobiology of personality: Essays in honor of Marvin Zuckerman (pp. 49–64). New York: Elsevier.

*Jones, H. A., & Lejuez, C. W. (2005). Personality correlates of caffeine dependence: The role of sensation seeking, impulsivity, and risk-taking. *Experimental and Clinical Psychopharmacology*, *13*, 259–266.

Killgore, W. D. S., Vo, A. H., Castro, C. A., & Hoge, C. W. (2006). Assessing risk propensity in American soldiers: preliminary reliability and validity of the Evaluation of Risks (EVAR) scale–English version. *Military Medicine*, *171*, 233–9.

*Killgore, W. D. S. (2007). Effects of sleep deprivation and morningness-eveningness traits on risk-taking. *Psychological Reports*, *100*, 613–626.

Killgore, W. D., Grugle, N. L., Killgore, D. B., Leavitt, B. P., Watlington, G. I., McNair, S., et al. (2008). Restoration of risk-propensity during sleep deprivation: caffeine, dextroamphetamine, and modafinil. *Aviation, Space, and Environmental Medicine*, *79*, 867–874.

Kirby, K. N., & Finch, J. C. (2010). The hierarchical structure of self-reported impulsivity. *Personality and Individual Differences*, *48*, 704–713.

Kuhberger, A. (1998). The Influence of framing on risky decisions: A meta-analysis. *Organizational Behavior and Human Decision Processes*, *75*, 23–55.

Lauriola, M., & Levin, I. P. (2001). Personality traits and risky decision-making in a controlled experimental task: An exploratory study. *Personality and Individual Differences*, *31*, 215–226.

Ledgerwood, D. M., Alessi, S. M., Phoenix, N., & Petry, N. M. (2009). Behavioral assessment of impulsivity in pathological gamblers with and without substance use disorder histories versus healthy controls. *Drug and Alcohol Dependence*, *105*, 89–96.

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., Aklin, W. M., Bornovalova, M. A., & Moolchan, E. T. (2005). Differences in risk-taking propensity across inner-city adolescent ever- and never-smokers. *Nicotine & Tobacco Research*, *7*, 71–79.

*Lejuez, C. W., Aklin, W. M., Daughters, S. B., Zvolensky, M. J., Kahler, C. W., & Gwadz, M. (2007). Reliability and validity of the youth version of the Balloon Analogue Risk Task (BART-Y) in the assessment of risk-taking behavior among inner-city adolescents. *Journal of Clinical Child and Adolescent Psychology*, *36*, 106–111.

*Lejuez, C. W., Aklin, W. M., Jones, H. A., Richards, J. B., Strong, D. R., Kahler, C. W., et al. (2003a). The Balloon Analogue Risk Task (BART) differentiates smokers and nonsmokers. *Experimental and Clinical Psychopharmacology*, *11*, 26 –33.

*Lejuez, C. W., Aklin, W. M., Zvolensky, M. J., & Pedulla, C. M. (2003b). Evaluation of the Balloon Analogue Risk Task (BART) as a predictor of adolescent real-world risk-taking behaviours. *Journal of Adolescence*, *26*, 475–479.

*Lejuez, C. W., Read, J. P., Kahler, C. W., Richards, J. B., Ramsey, S. E., Stuart, G. L., et al. (2002). Evaluation of a behavioral measure of risk-taking: The Balloon Analogue Risk Task (BART). *Journal of Experimental Psychology. Applied*, *8*, 75–84.

Lighthall, N. R., Mather, M., & Gorlick, M. A.(2009). Acute stress increases sex differences in risk seeking in the Balloon Analogue Risk Task. *PLoS One*, *4*, e6002. doi: 10.1371/journal.pone.0006002

Lipsey, M. W., & Wilson, D. B. (2001). Practical meta-analysis. Thousand Oaks, California: Sage Publications.

Logan, G. D., Schachar, R. J., & Tannock, R. (1997). Impulsivity and inhibitory control. *Psychological Science*, *8*, 60–64.

*Lorian, C. N., & Grisham, J. R. (2010). The safety bias: Risk-avoidance and social anxiety pathology. *Behaviour Change*, *27*, 29–41.

*MacPherson, L., Magidson, J. F., Reynolds, E. K., Kahler, C. W., & Lejuez, C. W. (2010). Changes in sensation seeking and risk-taking propensity predict increases in alcohol use among early adolescents. *Alcoholism, Clinical and Experimental Research*, *34*, 1400–08.

Madden, G. J., Petry, N. M., Badger, G. J., & Bickel, W. K. (1997). Impulsive and self-control choices in opioid-dependent patients and non-drug-using control participants: Drug and monetary rewards. *Experimental and Clinical Psychopharmacology*, *5*, 256–262.

*Marini, V. A., & Stickle, T. R. (2010). Evidence for deficits in reward responsivity in antisocial youth with callous-unemotional traits. *Personality Disorders: Theory, Research, and Treatment*, *1*, 218–29.

Meda, S. A., Stevens, M. C., Potenza, M. N., Pittman, B., Gueorguieva, R., Andrews, M. M., et al. (2009). Investigating the behavioral and self-report constructs of impulsivity domains using principal component analysis. *Behavioural Pharmacology*, *20*, 390–399.

DOI: 10.1002/bdm

Meng, X. L., Rosenthal, R., & Rubin, D. B. (1992). Comparing correlated correlation coefficients. *Psychological Bulletin*, *111*, 172–175.

*Mishra, S., Lalumière, M. L., & Williams, R. J. (2010). Gambling as a form of risk-taking: Individual differences in personality, risk-accepting attitudes, and behavioral preferences for risk. *Personality and Individual Differences*, *49*, 616–21.

Moeller, F. G., Barratt, E. S., Dougherty, D. M., Schmitz, J. M., Swann, A. C.. (2001). Psychiatric aspects of impulsivity. *The American Journal of Psychiatry*, *158*, 1783–1793.

Mohammed, S., & Schwall, A. (2009). *Individual differences and decision making: What we know and where we go from here*. In G. P. Hodgkinson, & J. K. Ford (Eds.), *International Review of Industrial and Organizational Psychology*, *24* (pp. 249–312). John Wiley & Sons.

Moore, S., & Gullone, E. (1996). Predicting adolescent risk behavior using a personalized cost benefit analysis. *Journal of Youth and Adolescence*, *25*, 343–359.

Mullen, B., Muellerleile, P., & Bryant, B. (2001). Cumulative meta-analysis: A consideration of indicators of sufficiency and stability. *Personality and Social Psychology Bulletin*, *27*, 1450–1462.

Nelson, M. C., Lust, K., Story, M., & Ehlinger, E. (2008). Credit card debt, stress and key health risk behaviors among college students. *American Journal of Health Promotion*, *22*, 400–412.

Panno, A., Lauriola, M., & Figner, B. (2013). Emotion regulation and risk taking: Predicting risky choice in deliberative decision making. *Cognition & Emotion*, *27*, 326–334.

Patton, J. H., Stanford, M. S., & Barratt, E. S. (1995). Factor structure of the Barratt Impulsiveness Scale. *Journal of Clinical Psychology*, *51*, 768–774.

Penolazzi, B., Gremigni, P., & Russo, P. M. (2012). Impulsivity and reward sensitivity differentially influence affective and deliberative risky decision making. *Personality and Individual Differences*, *53*(5), 655–659.

Pleskac, T. J. (2008). Decision making and learning while taking sequential risks. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*, 167–185.

*Pleskac, T. J., Wallsten, T. S., Wang, P., & Lejuez, C. W. (2008). Development of an automatic response mode to improve the clinical utility of sequential risk-taking tasks. *Experimental and Clinical Psychopharmacology*, *16*, 555–564

Reed, S. C., Levin, F. R., & Evans, S. M. (2010). The effects of progesterone pretreatment on the response to oral d-amphetamine in women. *Hormones and Behavior*, *58*, 533–543.

*Reynolds, B., Ortengren, A., Richards, J. B., & de Wit, H. (2006). Dimensions of impulsive behavior: Personality and behavioral measures. *Personality and Individual Differences*, *40*, 305–15.

Roberti, J. W. (2004). A review of behavioral and biological correlates of sensation seeking. *Journal of Research in Personality*, *38*, 256–279.

*Romer, D., Betancourt, L., Giannetta, J. M., Brodsky, N. L., Farrah, M., & Hurt, H. (2009). Executive cognitive functions and impulsivity as correlates of risk-taking and problem behavior in preadolescent. *Neuropsychologia*, *47*, 2916–2926.

Rosenthal, R. (1979). The "file drawer problem" and tolerance for null results. *Psychological Bulletin*, *86*, 638–641.

Rosenthal, R., & Di Matteo, M. R. (2000). Meta analysis: Recent developments in quantitative methods for literature reviews. *Annual Review of Psychology*, *52*, 59–82.

Schonberg, T., Fox, C. R., & Poldrack, R. A. (2010). Mind the gap: Bridging economic and naturalistic risk-taking with cognitive neuroscience. *Trends in Cognitive Sciences*, *15*, 11–19.

*Skeel, R. L., Pilarski, C., Pytlak, K., & Neudecker, J. (2008). Personality and performance-based measures in the prediction of alcohol use. *Psychology of Addictive Behaviors*, *22*, 402–09.

Slovic, P. (1966). Risk-taking in children: Age and sex differences. *Child Development*, *37*, 169–176.

Steinberg, L. (2007). Risk-taking in adolescence: New perspectives from brain and behavioral science. *Current Directions in Psychological Science*, *16*, 55–59.

Steinberg, L. (2008). A social neuroscience perspective on adolescent risk-taking. *Developmental Review*, *28*, 78–106.

Steinberg, L., Albert, D., Cauffman, E., Banich, M., Graham, S., & Woolard, J. (2008). Age differences in sensation seeking and impulsivity as indexed by behavior and selfreport: Evidence for a dual systems model. *Developmental Psychology*, *44*, 1764–1778.

*Swogger, M. T., Walsh, Z., Lejuez, C. W., & Kosson, D. S. (2010). Psychopaty and risk-taking among jailed inmates. *Criminal Justice and Behavior*, *37*, 439–52.

Tellegen, A. (1982). Brief manual for the Multidimensional Personality Questionnaire. Minneapolis, MN: University of Minnesota.

Tull, M. T., Trotman, A., Duplinsky, M. S., Reynolds, E. K., Daughters, S. B., Potenza, M. N., et al. (2009). The effect of posttraumatic stress disorder on risk-taking propensity among crack/cocaine users in residential substance abuse treatment. *Depression and Anxiety*, *26*, 1158–1164.

Upton, D. J., Bishara, A. J., Ahn, W.-Y., & Stout, J. C. (2011). Propensity for risk taking and trait impulsivity in the Iowa Gambling Task. *Personality and Individual Differences*, *50*, 492–495.

Verdejo-Garcia, A., Lawrence, A. J., Clark, L. (2008). Impulsivity as a vulnerability marker for substance-use disorders: Review of findings from high-risk research, problem gamblers and genetic association studies. *Neuroscience and Biobehavioral Reviews*, *32*, 777–810.

Vevea, J. L., & Woods, C. M. (2005). Publication bias in research synthesis: Sensitivity analysis using a priori weight functions. *Psychological Methods*, *10*, 428–443.

Weber, E. U., & Johnson, E. J. (2008). Decisions under uncertainty: Psychological, economic and neuroeconomic explanations of risk preference. In P. Glimcher, C. Camerer, E. Fehr, & R. Poldrack (Eds.), Neuroeconomics: Decision making and the brain (127–44). New York: Elsevier.

Weber, E. U., & Johnson, E. J. (2009). Mindful judgment and decision making. *Annual Review of Psychology*, *60*, 53–86.

Weafer, J., Milich, M., & Fillmore, M. T. (2011). Behavioral components of impulsivity predict alcohol consumption in adults with ADHD and healthy controls. *Drug and Alcohol Dependence*, *113*, 139–146.

Weller, J. A., Levin, I. P., & Denburg, N. L. (2011), Trajectory of risky decision making for potential gains and losses from ages 5 to 85. *Journal of Behavioral Decision Making*, *24*, 331–344.

Weller, J. A., & Tikir, A. (2011). Predicting domain-specific risk taking with the HEXACO personality structure. *Journal of Behavioral Decision Making*, *201*, 180–201.

White, T. L., Lejuez, C. W., & de Wit, H. (2007). Personality and gender differences in effects of d-amphetamine on risk-taking. *Experimental and Clinical Psychopharmacology*, *15*, 599–609.

Whiteside, S. P., & Lynam, D. R. (2001). The Five Factor Model and impulsivity: Using a structural model of personality to understand impulsivity. *Personality and Individual Differences*, *30*, 669–689.

Wilson, D. B. (2012). Personal communication.

Zuckerman, M. (1994). Behavioral expressions and biosocial bases of sensation seeking. New York: Cambridge University Press.

Zuckerman, M. (2007). Sensation seeking and risky behavior. Washington, DC: American Psychological Association.

Zuckerman, M., Eysenck, S. B. J., & Eysenck, H. J. (1978). Sensation seeking in England and America: Cross-cultural, age, and sex comparisons. *Journal of Consulting and Clinical Psychology*, *46*, 139–149.

DOI: 10.1002/bdm

Zuckerman, M., & Kuhlman, D. M. (2000). Personality and risk-taking: Common biosocial factors. *Journal of Personality*, *68*, 999–1029.

Zuckerman, M., Kuhlman, D., Joireman, J., Teta, P., Kraft, M. (1993). A comparison of the three structural models for the personality: The big three, the big five and the alternative five. *Journal of Personality and Social Psychology*, *65*, 757–768.

Authors' biographies:

Marco Lauriola, Ph.D. is Associate Professor of Psychology at University of Rome, "Sapienza". His research is focused on individual differences in personality and decision-making. In this line, he published a number of original contributions relating cognitive styles and personality traits to judgment processes and choice behaviors under uncertainty, ambiguity and risk, including health risks and safety hazards.

Angelo Panno, Ph.D., graduated in Social Psychology at University of Rome, "Sapienza", Italy. He defended a dissertation titled "Self-regulation and emotions: predicting risky choice". His research is focused on strategies people use to regulate emotions during the decision process. He is also interested in studying the effects of anticipated regret, both as predictor and as mediator of risk taking in behavioral tasks.

Irwin P. Levin is Professor Emeritus in the Department of Psychology and the Department of Marketing at the University of Iowa. His interests are in individual differences in decision making, particularly risky decision making. His recent work includes age-related differences and neuropsychological correlates of behavioral differences.

Carl W. Lejuez, Ph.D. is Professor of Psychology at University of Maryland. His research is translational in nature, applying laboratory methods to understand real world clinical problems and then applying this knowledge to develop novel assessment and treatment strategies. His research also spans the clinical domains of addictions, personality pathology, and mood disorders, and he is most interested in the common processes across these conditions. Dr. Lejuez is Founder and Director of the Center for Addictions, Personality, and Emotion Research (CAPER).

Authors' addresses:

Marco Lauriola and Angelo Panno, Department of Social and Developmental Psychology, University of Rome "Sapienza", Italy

Irwin P. Levin, Department of Psychology, University of Iowa, USA

Carl W. Lejuez, Department of Psychology, University of Maryland, USA

DOI: 10.1002/bdm