Developmental Cognitive Neuroscience 25 (2017) 221–234

Contents lists available at ScienceDirect

Developmental Cognitive Neuroscience

journal homepage: http://www.elsevier.com/locate/dcn

Beyond dual systems: A genetically-informed, latent factor model of behavioral and self-report measures related to adolescent risk-taking

K. Paige Harden[a,b,*], Natalie Kretsch[a], Frank D. Mann[a], Kathrin Herzhoff[c], Jennifer L. Tackett[c], Laurence Steinberg[d], Elliot M. Tucker-Drob[a,b]

[a] *Department of Psychology, University of Texas at Austin, Austin, TX, United States*
[b] *Population Research Center, University of Texas at Austin, Austin, TX, United States*
[c] *Department of Psychology, Northwestern University, Evanston, IL, United States*
[d] *Department of Psychology, Temple University, Philadelphia, PA, United States*

ARTICLE INFO

Article history:
Received 8 March 2016
Received in revised form
21 November 2016
Accepted 20 December 2016
Available online 26 December 2016

Keywords:
Dual systems
Cognitive control
Reward seeking
Impulsivity
Self-control
Intelligence
Delay discounting
Risk-taking
Adolescence

ABSTRACT

The *dual systems* model posits that adolescent risk-taking results from an imbalance between a cognitive control system and an incentive processing system. Researchers interested in understanding the development of adolescent risk-taking use a diverse array of behavioral and self-report measures to index cognitive control and incentive processing. It is currently unclear whether different measures commonly interpreted as indicators of the same psychological construct do, in fact, tap the same underlying dimension of individual differences. In a diverse sample of 810 adolescent twins and triplets (*M* age = 15.9 years, *SD* = 1.4 years) from the Texas Twin Project, we investigated the factor structure of fifteen self-report and task-based measures relevant to adolescent risk-taking. These measures can be organized into four factors, which we labeled *premeditation*, *fearlessness*, *cognitive dyscontrol*, and *reward seeking*. Most behavioral measures contained large amounts of task-specific variance; however, most genetic variance in each measure was shared with other measures of the corresponding factor. Behavior genetic analyses further indicated that genetic influences on cognitive dyscontrol overlapped nearly perfectly with genetic influences on IQ ($r_A = -0.91$). These findings underscore the limitations of using single laboratory tasks in isolation, and indicate that the study of adolescent risk taking will benefit from applying multimethod approaches.

"Why do people not think, when they are grown up, as I do when I am young?"

"Oh dear," said Merlin, "You are making me feel confused. Suppose you wait till you are grown up and know the reason?"

"I don't think that is an answer at all," replied the Wart, justly.

–T.H. White, *The Once and Future King*

1. Introduction

1.1. The dual systems model of adolescent risk-taking

Adolescence is a health paradox. Although teenagers are stronger, fitter, more fertile, and more resistant to disease than children or adults, they are also uniquely vulnerable to being hurt or killed because of their risky behavior. The top three leading causes of death for adolescents and young adults ages 15–24 are due to behaviors – unintentional injury (such as falls or car accidents), suicide, and homicide (Centers for Disease Control, 2015). The long-standing question, then, is *why* teenagers are prone to risk-taking behaviors.

Central to any explanation of the development of adolescent risk-taking is the construct of *self-control*. Self-control can be defined as behavioral tendencies that produce "actions aligned with valued, longer-term goals" (Duckworth and Steinberg, 2015) or as "the ability to suppress inappropriate emotions, desires, and actions in favor of alternative, appropriate ones" (Casey, 2015). Given that adolescent risk-taking generally is not considered valued or appropriate by adult society, it is practically a tautology to say that adolescents are prone to risk-taking because they lack self-control. In the past decade, however, research has begun to push past this tautology and offered new insights into (1) what are the component psychological processes that contribute to

* Corresponding author at: 108 E. Dean Keeton Stop #A8000, Austin, TX 78712, United States.
E-mail address: harden@utexas.edu (K.P. Harden).

https://doi.org/10.1016/j.dcn.2016.12.007
1878-9293/

self-controlled versus risk-taking behavior, (2) how do these component processes change, on average, during adolescence, and (3) what neural systems underlie these developmental changes?

A critical distinction in this work has been between processes that enable self-controlled behavior versus those processes that disrupt it (Duckworth and Steinberg, 2015). And, this distinction between "volitional" versus "impulsigenic" processes has also been central to neurodevelopmental theories of adolescent risk-taking. According to the *dual systems model* (Steinberg, 2010; Shulman et al., 2016) and the related *maturational imbalance* model (Casey et al., 2011; Casey, 2015), adolescent risk-taking can be understood in terms of two separable dimensions – incentive processing and cognitive control – that mature according to separate timescales and that are undergirded by distinct neurobiological circuits. First, an *incentive processing* system, involving sub-cortical brain regions such as the striatum, is hypothesized to undergird impulsigenic processes, such as reward sensitivity, novelty seeking, and sensation seeking. The incentive processing system is hypothesized to develop rapidly from early to middle adolescence. Multiple studies have found, for instance, that adolescents show stronger reactivity in the ventral striatum in response to rewards than do children or adults, and that the extent of ventral striatum reactivity correlates with real-world risk-taking (Galvan, 2010; Shulman et al., 2016).

In contrast, a *cognitive control* system, involving the prefrontal cortex, is hypothesized to undergird volitional psychological processes that facilitate self-controlled behavior, such as planning, abstract reasoning, and impulse control. Unlike the incentive processing system, the cognitive control system appears to mature more slowly from adolescence through emerging adulthood. For example, white matter volume increases until the mid-twenties, particularly in areas of the brain involved in high-level cognitive control (Geier, 2013). This neural maturity gap, therefore, is thought to result in adolescents experiencing heightened sensitivity to rewards and emotions without a concomitant increase in capacity for controlling their behavior.

1.2. Measuring dimensions of self-control: the problem of construct validity

The dual systems model has become a popular and generative theoretical framework, and there has been a corresponding explosion in the number of empirical papers that test or challenge its predictions (see, for example, Casey et al., 2016; Pfeifer and Allen, 2016; Shulman et al., 2016; van den Bos and Eppinger, 2016 for reviews and commentary). One critical challenge for this rapidly expanding literature is the problem of *construct validity* (Cronbach and Meehl, 1955): What constructs account for performance on commonly used laboratory tasks and survey measures? The field currently suffers from three core limitations to understanding construct validity. First, investigators commonly attempt to measure key constructs using a single behavioral task, which confounds idiosyncratic task-specific variation with systematic, construct-relevant variation that is of true interest. Second, the tasks used to measure the same ostensible underlying construct differ from study to study and from lab to lab, which precludes testing how different measures of the same putative construct agree, if at all. Third, performance on the same task is sometimes described and interpreted in terms of different underlying constructs. For example, the same task (delay discounting) has been used to measure both reward sensitivity (Weigard et al., 2014) and cognitive control (van den Bos et al. (2015), but see also commentary by Steinberg and Chein (2015)). Previous reviewers have noted that these limitations severely limit research progress. In the absence of specificity in how constructs are defined and measured, a model can allow "for the generation of a virtually unlimited number of testable hypotheses, which may be inconsistent with each other but that are all consistent with this general claim" (van den Bos and Eppinger, 2016).

In their classic paper, Cronbach and Meehl (1955) proposed that the construct validity of a task or set of tasks could be investigated by testing a *nomological network*, the "interlocking system of laws which constitute a theory." The nomological network includes laws regarding how "observable properties or quantities" (i.e., observed measures) are related to each other, how theoretical constructs are related to measures, and how different theoretical constructs relate to each other. As Pfeifer and Allen (2016) described, "the absence of such an [nomological network] analysis can be a critical barrier to progress if studies that are putatively investigating a particular construct are not in fact doing so" (p. 134). Putting this into modern statistical terms, some predictions of a theorized nomological network can be tested as a latent factor model, in which the pattern of covariation among multiple measures is accounted for by a smaller number of latent factors that represent theoretical constructs, which may, in turn, be correlated with one another. Consider, for example, self-reports of sensation seeking and preference for immediate rewards on a delay discounting task. Both measures have been used by investigators interested in the rapid development of an incentive processing system, and thus might be presumed to tap the same underlying construct. In the context of a latent factor model, this presumption can be tested by examining whether these measures can be used as indicators of the same latent factor.

1.3. Dimensions of self-reported impulsivity

The multivariate pattern of associations among self-report and behavioral measures commonly used in the study of adolescent risk-taking has not yet been widely studied using a latent variable approach. However, within the personality literature, there have been extensive factor analytic studies of self-report measures that tap tendencies toward self-controlled versus disinhibited behavior. This line of research has identified four dimensions of disinhibited personality traits: (1) *urgency*, the "tendency to commit rash or regrettable actions as a result of intense negative affect"; (2) *premeditation*, the "tendency to delay action in favor of careful thinking and planning"; (3) *perseverance*, the tendency to "remain with a task until completion and avoid boredom", and (4) *sensation seeking*, "the tendency to seek excitement and adventure" (UPPS; Whiteside and Lynam, 2001, p. 677).

Some studies testing the predictions of the dual systems model have used self-report personality measures that clearly map onto the UPPS model. In a series of studies using longitudinal data from the National Longitudinal Survey of Youth (Harden and Tucker-Drob, 2011; Harden et al., 2012a,b; Quinn and Harden, 2013; Shulman et al., 2015a,b; Shulman et al., 2016), we investigated age-related changes in sensation seeking (measured using items such as "life with no danger would be too dull for me") and in what we called "impulsivity" or "impulse control." Items used to measure impulsivity in these studies included "I often get in a jam because I do things without thinking" – content that matches the UPPS dimension of premeditation. Consistent with the predictions of the dual-systems model, impulse control gradually improved over the course of adolescence, whereas sensation seeking increased from early to mid-adolescence, peaking around age 16, and then slowly declined through young adulthood (Harden and Tucker-Drob, 2011). Furthermore, individual differences in changes in sensation seeking and impulse control were only weakly associated with one another, evidence for their discriminant validity (Campbell and Fiske, 1959).

Other investigators have used different self-report measures as indices of the sensitivity of the incentive processing system, such as the Fun-Seeking Scale of the BIS/BAS scales (Carver and

White, 1994, used in van Duijvenvoorde et al., 2014) or the Present-Hedonism scale of Zimbardo's Time Perspective Inventory (Zimbardo and Boyd, 1999, used by van den Bos et al., 2014). Upon closer inspection, however, the content of these scales substantially overlaps with the UPPS dimension of sensation seeking. For example, BIS/BAS Fun-Seeking includes the item, "I crave excitement and new sensations," and the item with the highest factor loading on the Present-Hedonism scale is "I take risks to put excitement in my life." Crucially, making the inference that van den Bos et al. (2015), van Duijvenvoorde et al. (2014), and Harden and Tucker-Drob (2011) were all measuring the same personality construct (or, at least, similar constructs) would allow us to the bridge the developmental and neuroscience literatures: Average levels of sensation seeking increase during adolescence, individual differences in sensation seeking are heritable, and individual differences in sensation seeking are correlated with greater connectivity between the ventral striatum and other subcortical areas (hippocampus, amygdala) *and* with greater rewards-related activation in the ventral striatum and medial PFC (Harden and Mann, 2015).

1.4. Convergence between self-report and behavioral measures

In addition to self-report measures, studies of adolescent risk-taking also commonly use gambling tasks such as the Balloon Analogue Risk Task (BART; Lejuez et al., 2007), the Iowa Gambling Task (IGT; Bechara, 2007), and the Stoplight task (Steinberg et al., 2008). All of these tasks require participants to make decisions that involve some (unknown) possibility of reward and some (unknown) possibility of loss. Yet another commonly used behavioral measure of impulsivity is the delay discounting task (Green and Myerson, 2004), which asks participants to choose between a smaller reward received sooner or a larger reward received later. Performance on all of these tasks is associated with "real world" risk-taking behavior, particularly substance use (Bechara et al., 2001; Bickel and Marsch, 2001; Lejuez et al., 2002, 2003; Lejuez et al., 2003; Kim-Spoon et al., 2016; MacKillop et al., 2011).

Despite the commonsense intuition that these behavioral tasks are measuring similar aspects of impulsivity as self-report measures, previous research has found that the two methods show minimal agreement. For example, in a small sample of university students ($n = 70$), performance on neither the BART nor the IGT showed any correlation with any self-report measures tapping premeditation (BIS-11, I-7 Impulsivity, MPQ Constraint) or sensation seeking (I-7 Venturesomeness; Reynolds et al., 2006). On the basis of these results, the researchers concluded that "self-report and behavioral tasks probably measure different constructs, and... even among the behavioral measures, different tasks measure different, perhaps unrelated, components of impulsive behavior" (Reynolds et al., 2006, p. 305–306). Similarly, Cyders and Coskunpinar (2011) meta-analyzed correlations from 27 research studies reporting at least one association between self-report and behavioral measure of impulsivity. They concluded that, "practically, the relationship is small" (p. 965). Most recently, MacKillop et al. (2016) conducted one of the few latent factor analyses of self-report and behavioral measures of impulsivity in a well-powered sample of young adults ($n = 1252$). Consistent with previous research focusing on pairwise correlations, they concluded that a factor representing "impulsive choice," indicated by choices on various delay discounting tasks, was only minimally related to a factor representing self-reported impulsive personality traits on the UPPS and Barratt Impulsiveness Scale ($r = 0.10$).

Additionally, behavioral measures may correspond poorly – or not at all – with each other. In undergraduates, the association between BART and delay discounting is negligible ($r = -0.049$, Reynolds et al., 2006). Two studies (of high school students and university students) found that performance on the BART is unrelated to performance on the IGT (Aklin et al., 2005; Buelow and Blaine, 2015). A similar study did find a modest correlation between the two ($r = 0.25$, Skeel et al., 2007); however, all of these studies had small samples ($n \leq 70$) and were thus both underpowered to detect small effects and imprecise with respect to the confidence intervals surrounding their effect size estimates. Moreover, unlike work in the domain of self-report personality scales, correspondence among behavioral tasks has largely been investigated by simply examining pairwise correlations. No study has yet investigated the relations among survey and behavioral measures of impulsive traits relevant to risk-taking using a latent variable approach in a well-powered sample of adolescents.

1.5. The utility of behavioral genetic data

In addition to examining the observed *phenotypic* correlations among different measures, a nomological network can be further interrogated by using a behavioral genetic design, which uses different types of biological relatives (e.g., identical versus fraternal twins) to estimate how variation in observed test performance is associated with genetic differences between people. A behavioral genetic analysis can inform three issues. First, and most simply, a twin design can indicate whether a test or measure is sensitive to genetic differences. Previous behavioral genetic research has confirmed that genes are a major contributor to individual differences in risk-taking and externalizing behavior in adolescence (Krueger et al., 2002). However, the specific genetic variants that underlie this omnibus genetic effect, as well as the paths that mediate genetic effects on complex behavior, are largely unknown (Manolio et al., 2009). The rapidly expanding field of imaging genetics seeks to inform these questions by building models of gene → brain → behavior relationships (e.g., Ge et al., 2013; Medland et al., 2014). Measures that are not sensitive to genetic differences between people are expected to be of limited utility for imaging genetics studies, and neglect a major source of individual differences in risk-taking.

Second, in addition to estimating the heritability of a single measure, behavior genetic designs can test the extent to which the *association* between measures is mediated by genetic influences that are shared between them. Measures may be only modestly associated at the phenotypic level, yet may tap strongly overlapping sets of genetic predispositions. Alternatively, observed performance on two measures may correspond rather strongly, but this correspondence could be primarily attributable to environmental influences. The strength of the genetic correlation between measures has become increasingly relevant in the era of genome-wide association studies (GWAS) that have begun to identify specific genetic variants for many commonly studied traits (e.g., Okbay et al., 2016a,b). GWASs require enormously large samples, and many measures commonly used in developmental cognitive neuroscience are too lengthy or difficult to be feasibly administered to samples of $n > 100{,}000$ people. However, if there is a substantial genetic correlation between measures, then a polygenic score calculated from the results of a GWAS of measure #1 (such as a self-report personality scale) may be useful for predicting performance on measure #2 (such as an in-lab behavioral task), or even for predicting neural activity during that task. In this way, researchers would be able to probe the specific biological mechanisms of gene-behavior realtionships in much less extravagantly-sized samples.

Third, although twin studies are most commonly thought of as tools for understanding genetic influence, they are also informative regarding how environmental influences are stratified in a population. Specifically, twin designs decompose environmentally-influenced variation in a phenotype into two components. The first component, termed the non-shared environment, quantifies the extent to which even identical twins differ from one another, and reflects both the effect of idiosyncratic, twin-specific environmental influences, as well as non-systematic error variance. The second component, termed the shared environment, reflects the extent to which twins raised in the same home are similar to one another (above and beyond any effects of genetic resemblance). In addition to the home rearing environment, possible sources of shared environmental influence include schools, neighborhoods, and shared peer groups. The extent of shared environmental variance in measures of self-control is a particularly interesting question. Previous behavioral genetic research on personality in adolescence and young adulthood has found minimal evidence for shared environmental influence for most traits (Briley and Tucker-Drob, 2014). At the same time, educators and policymakers increasingly identify self-control, as well as other "non-cognitive" skills, as appropriate metrics for evaluating school quality (Heckman and Rubinstein, 2001; Tough, 2012). Behavioral genetic analyses can indicate whether naturally occurring variation in family-wide sociocontextual influences accounts for variation in measures of impulsivity and planning, as might be expected given the policy emphasis on school differences (c.f. Tucker-Drob et al., 2016).

1.6. Goals of the current paper

In this paper, we present results from a series of latent factor models that examine the pattern of associations among self-report and behavioral measures in a multivariate assessment battery. Drawing on previous research on adolescent risk-taking, this battery includes self-reports of sensation seeking, impulse control, planning for the future, and the perceived negative and positive consequences of risky behaviors; gambling tasks that require participants to make decisions that involve the possibility of rewards and losses; cognitive tests that require participants to plan and reason abstractly; and a delay discounting task. This battery, therefore, broadly samples measures commonly used in "dual systems" research on adolescent risk-taking, and was administered to a population-based, ethnically and socioeconomically diverse sample of adolescent twins. By using a twin sample, we are further able to examine the extent of genetic and shared environmental influence on each measure and on their associations.

2. Method

2.1. Participants

Data for the current study were drawn from an on-going in-laboratory study of high-school age twins that forms one of the core components of the Texas Twin Project (Harden et al., 2013). Participants were identified and recruited from public school rosters in the Austin, TX and Houston, TX metropolitan areas. The analytic sample consisted of 810 adolescents nested within 398 families (385 twin pairs, 12 sets of triplets, and 1 set of quadruplets). Each set of triplets contributes 3 pairwise combinations, and the quadruplets set contributes 6 pairwise combinations, resulting in 427 pairs: 153 monozygotic (MZ) pairs (84 female; 69 male) and 284 DZ pairs (59 female, 83 male, and 132 opposite-sex). Zygosity was classified based on ratings of physical similarity (Heath et al., 2003); see the Supplement for more information. Participants ranged in age from 13.6 to 20.1 years (M = 15.9 years, SD = 1.4). Most participants (89%) were between the ages of 14 and 18. Fifty-seven (57%) percent of twins reported that they were non-Hispanic White, 18% reported that they were Hispanic/Latino, 15% reported that they were Black/African American, 5% reported that they were Asian/Asian American, and the remaining 5% reported that they were another race/ethnicity.

2.2. Measures

We administered three self-report questionnaires and five behavioral measures, as well as an IQ test and a self-report measure of pubertal development, which are summarized in Table 1. The battery of measures was designed to capture a broad array of beliefs, predispositions, motivations, and abilities relevant to risk taking, including risk perception, reward sensitivity, planning, and impulse control.

All measures were administered in-laboratory. Surveys were administered in-laboratory using REDCap (Research Electronic Data Capture), a secure, web-based application designed to support data capture for research studies (Harris et al., 2009). The IGT and Delay Discounting tasks were programmed and administered using E-Prime version 2.0. A computerized version of Tower of London was obtained from Sanzen Neuropsychological Assessment (www.catstests.com). The Balloon Analogue Risk Task-Youth version and the Stoplight task are distributed by the task authors as free-standing executable programs.

Table 2 describes summary statistics (Ms and SDs) for the dependent variables (DVs). Additional information on the scoring of the IGT and the Delay Discounting task is provided in the Supplement. Average time to first move on the Tower of London was log-transformed so that the distribution was more normal. Standardized DVs were used for all subsequent analyses. See the Supplement for density plots of the distributions of all DVs in males and females, and for a table of the zero-order correlations among all DVs. As expected, girls reported more advanced pubertal development than boys (females: M = 3.38, SD = 0.81; males: M = 2.91, SD = 0.71), and pubertal development was correlated with age for boys (r = 0.41) and girls (r = 0.42).

2.3. Analyses

Two sets of analyses were conducted. First, we fit a series of Exploratory Structural Equation Models (ESEM; Asparouhov and Muthén, 2009). ESEM integrates aspects of traditional exploratory factor analysis (EFA) within a confirmatory factor analysis/structural equation modeling (CFA/SEM) framework (Marsh et al., 2014), and allows one to estimate a factor loading matrix while simultaneously estimating the association between the latent factors (free of measurement error) and covariates.

Second, informed by the ESEM results, we fit behavior genetic models that decomposed variation unique to each measure and variation common across measures into additive genetic influences (A), environmental influences shared by twins raised in the same home (C), and non-shared environmental influences unique to each twin (E). For measured variables, but not latent factors, the E component also includes measurement error. E also includes any postzygotic differences between MZ twins in genetic sequence (e.g., mutations), although this is expected to be trivial in proportion to the total amount of genetic sequence variability within humans. For a complete introduction to the logic and parameterization of twin models, including a discussion of all relevant assumptions, please see Neale and Maes (2004) or Plomin et al. (2013).

All analyses were conducted in M*plus* version 7.1 (Muthén and Muthén, 1998–2015). Model fit was primarily evaluated using the root mean square error of approximation (RMSEA; Steiger, 1990), with values < 0.05 indicating good model fit. For phenotypic analyses that treated each individual as a case, standard errors and model fit statistics were corrected for nesting within families using the TYPE = COMPLEX option in M*plus* (McNeish and Stapleton, 2016). The behavior genetic models were parameterized as multiple-group, multilevel models, using TYPE = TWOLEVEL in M*plus*, where the groups were defined by zygosity (MZ vs. DZ) and the levels defined by clustering within vs. between twin families. Multilevel

Table 1
Summary of Measurement Battery.

Measure	Source	Paradigm	Dependent Variable(s)
Self-Report Questionnaires			
UPPS Impulsivity Scale[a]	Whiteside and Lynam (2001)	45-item self-report survey tapping 4 dimensions of impulsivity: *urgency* (e.g., "Sometimes when I feel bad, I can't seem to stop what I am doing, even though it is making me feel worse"), *premeditation* (e.g., "I like to stop and think things over before I do them"), *perseverance* (e.g., "I finish what I start"), and *sensation seeking* (e.g., "I generally seek new and exciting experiences and sensations").	Urgency subscale score Premeditation subscale score Sensation Seeking subscale score
Future Orientation scale	Steinberg et al. (2009)	15-item self-report survey that measures 3 dimensions of future orientation: *planning ahead* (e.g., "I like to plan things out one step at a time"), *time perspective* (e.g., "I often think about what my life will be like 10 years from now"), and *anticipation of future consequences* (e.g., "I usually think about the consequences before I do something.̈)	Planning subscale score Time Perspective subscale score Anticipation of Future Consequences subscale score
Risk Perception scale	Benthin et al. (1993)	28-item self-report survey that asks participants to imagine 7 risky activities (e.g., "Having sex without a condom", "Trying a new drug") and answer 4 questions for each. *Perceived harms* score based on 3 items per activity regarding being frightened, being personally at risk, and expecting harmful effects to be serious. *Perceived benefits* vs. *harms* scores based on 1 item per activity on benefits or pleasures being greater than the risks.	Perceived Harms subscale score Perceived Benefits vs. Harms subscale score
Pubertal Development scale	Petersen et al. (1988)	5 sex-specific items rating growth in height, growth of body hair, and skin changes (males and females), growth of facial hair and deepening of voice (males), and growth of breasts and menstruation (females). The menstruation item was coded to be consistent with the 4-point scale of the other items (1 = *No*, 4 = *Yes*).	PDS scale score
Behavioral Tasks			
Iowa Gambling Task	Cauffman et al. (2010) and Bechara et al. (1994)	Individuals are given the opportunity to "play" or "pass" from 4 decks of cards. Two decks are advantageous ("good"), in that repeated play will ultimately result in winning money, whereas two decks are disadvantageous ("bad"), in that repeated play will result in losing money.	Play on good decks Play on bad decks
Delay Discounting Task	Steinberg et al. (2009)	Participants choose between smaller, immediate rewards and a larger, delayed reward. All rewards are hypothetical. The value of the delayed reward is held constant at $1000, but the length of the delay varies across six blocks (1 week, 1 month, 6 months, 1 year, 5 years, 15 years). On the first trial of each block, participants are presented with the choice between a delayed reward of $1000 or an immediate reward of either $200, $500, or $800 (randomly determined). If the participant selects the delayed reward, the immediate reward option on the next trial is intermediate between the previous value and $1000 (i.e., increased on the next trial); in contrast, if the participant selects the immediate reward, then the next trial offers an immediate reward that is intermediate between the previous value and $0. This is repeated until participants' responses converge on an *indifference point* – the value of an immediate reward that is equivalent to $1000 received after some delay. For example, a person might value $400 received today equivalently to $1000 received after 6 months. The indifference points for the six time intervals are averaged.	Average indifference point
Balloon Analogue Risk Task – Youth version	Lejuez et al. (2007)	Individuals decide how much air to "pump" into a balloon on the computer screen. For each successful pump of air, more points are accrued; however, at some point, the addition of more air causes the balloon to burst, leading the participant to lose all points accrued during that trial.	Average number of pumps on trials during which balloon did not explode
Stoplight	Steinberg et al. (2008)	Individuals "drive" a car to a destination under time pressure. Along the way are a series of crossroads, and at each one the person decides whether to run a yellow light, which turns red after a variable amount of time, or to stop and wait for the light to turn red and then green. Time is saved if the person successfully runs the yellow light, whereas time is lost when the light turns red and the person crashes into another car at the intersection (more time than if the person had stopped and waited for the light to turn green).	Percent of crossroads at which person failed to stop
Tower of London	Sanzen Neuropsychological Tests (2003–2015); Shallice (1982)	Individuals view a series of balls on a peg and the balls that must be moved to a pre-specified goal configuration on another peg. Participants are instructed to replicate the goal configuration using the smallest number of moves. The task measures how well an individual can organize sequential behavior to reach a goal and ability to inhibit acting before a plan is fully formed.	Total excess number of moves made (relative to perfect solution) Average time to first move (log-transformed)
Wechsler Abbreviated Scale of Intelligence	Psychological Corporation (1999)	Test of general cognitive ability that estimates full scale IQ using four subtests: Vocabulary, Similarities, Matrix Reasoning, and Block Design.	Full-scale IQ (FSIQ)

[a] Preliminary analyses indicated that one UPPS dimension, lack of perseverance, had no significant relationship with any behavioral measure (all correlations less than 0.10), and all factor analyses including lack of perseverance had serious convergence problems. As failure to complete tasks (i.e., impulsive quitting) also does not have any straightforward theoretical relationship with the dual systems model, all subsequent analyses focused on the other three subscales of the UPPS (*sensation seeking*, *lack of premeditation*, and *urgency*).

Table 2
Summary Statistics and Twin Correlations for Self-Control Measures.

	Summary Statistics				*Twin Pair Intraclass Correlations*	
Dependent Variable(s)	*M*	*SD*	*Min*	*Max*	*MZ*	*DZ*
Self-Report[a]						
UPPS: Urgency	2.15	0.58	1.00	3.91	0.38	0.18
UPPS: Premeditation	2.96	0.48	1.36	4.00	0.42	0.16
UPPS: Sensation Seeking	2.85	0.58	1.17	4.00	0.51	0.17
Future Orientation: Planning	2.91	0.62	1.00	4.00	0.24	0.12
Future Orientation: Time Perspective	22	0.59	1.00	4.00	0.21	0.08
Future Orientation: Future Consequences	3.07	0.58	1.00	4.00	0.34	0.16
Risk Perception: Benefits vs. Harms	0.49	0.51	0.00	3.00	0.34	0.14
Risk Perception: Harms	3.39	0.43	1.00	4.00	0.43	0.30
Behavioral Measures						
IGT: Good decks[b]	0.78	0.13	0.13	1.00	0.32	0.23
IGT: Bad decks[b]	0.66	0.13	0.13	1.00	0.34	0.17
BART: Avg. adjusted pumps	29.98	12.24	1.00	84.60	0.28	0.16
Stoplight: proportion intersections	0.44	0.19	0.00	1.00	0.29	0.13
Delay Discounting: Avg. indiff. point[c]	493.04	213.75	1.00	999.00	0.46	0.29
TOL: Excess moves	2.61	2.28	0.00	12.00	0.30	0.23
TOL: Avg. time to first move (s) [d]	1.59	0.26	0.97	2.91	0.31	0.14
Full-scale IQ (FSIQ)[e]	102.21	13.09	61.00	146.00	0.76	0.48

Note: MZ = monozygotic, DZ = dizygotic.
[a] All survey items were rated on a 1–4 scale.
[b] Statistics reported for total proportion of plays from good and bad decks across all blocks. All subsequent analyses used standardized random intercepts estimated from mixed effects models centered on the final block. See Supplement for more information.
[c] See Supplement for more information on scoring of the Delay Discounting task.
[d] Average time to first move in seconds was log-transformed for subsequent analyses.
[e] Individuals with FSIQ scores <70 ($n = 6$, <1% of the sample) were excluded from subsequent analyses.

models require that each cluster (family) is assigned to one and only one grouping variable. One set of triplets had pairs of multiple zygosity types (1 set of identical twins and their fraternal twin); this family was dropped from analyses, resulting in $n = 807$ participants for behavior genetic analyses.

8. Results

8.1. ESEM

8.1.1. Testing dimensionality

The initial ESEMs estimated five models (with one-, two-, three-, four-, and five-factors) and estimated the loadings of all DVs on the factors using geomin rotation. These ESEMs are equivalent to traditional EFA models with one- to five-factors, respectively. One-, two-, and three-factor ESEMs all had poor fit to the data (RMSEAs > 0.05, CFI < 0.95), whereas a four-factor model had acceptable fit (RMSEA = 0.042, CFI = 0.96), and fit the data significantly better than a three-factor model (Satorra-Bentler scaled χ^2 difference test = 8.64, $p = 0.003$). In contrast, a five-factor model did not fit significantly better than a four-factor solution (Satorra-Bentler scaled χ^2 difference test = 3.59, $p = 0.05$), and no DV had a significant loading on the fifth factor. Therefore, a four-factor solution was selected as the best and most parsimonious representation of the data.

8.1.2. Results from 4-factor model: factor loading pattern

Next, we estimated a four-factor ESEM that included age, pubertal development, sex, and FSIQ as covariates of the factors. Parameter estimates from this model are summarized in Table 3. In describing these results, we focus on all loadings that were significantly different than zero at $p < 0.05$ or had absolute values ≥ 0.15.

Indicators for the first factor, which we label *Premeditation*, were largely composed of self-report questionnaires: the Premeditation scale of the UPPS and the three scales of the Future Orientation questionnaire had the highest loadings on this factor. UPPS urgency and sensation seeking also had small negative loadings. Adolescents scoring highly on the Premeditation factor describe themselves as careful and conscientious; they report that they plan and consider potential future outcomes. Although the largest loadings were for self-report measures, average time to first move on the Tower of London also had a small positive loading on premeditation, which is consistent with previous research (Steinberg et al., 2008).

The strongest indicators for the second factor, which we label *Fearlessness*, were the subscales of the Benthin Risk Perception scale. Adolescents scoring highly on this factor report being less frightened by the potential negative consequences of risk-taking, they find these potential harms to be less serious and less likely to happen to them personally, and they consider the potential benefits of risk-taking to outweigh any potential harms. In addition, UPPS Urgency and Sensation Seeking had smaller positive loadings on this factor: fearless adolescents report that they like dangerous or thrilling activities and do things they later regret when they are upset. Somewhat surprisingly, average time to first move on Tower of London also had a small positive loading on fearlessness.

The strongest indicators for the third factor, which we label *Reward-Seeking*, were plays on the good decks of the IGT, pumps on the BART, and running through intersections on the Stoplight task. In addition, there were smaller positive loadings from plays on bad decks on the IGT, delay discounting, and self-reported sensation seeking.

Indicators for the fourth factor, which we label *Cognitive Dyscontrol*, also included both self-report and behavioral measures: a greater number of excess moves (i.e., errors) and a shorter average time to first move on the Tower of London, greater play from bad decks on the Iowa Gambling Task, a lower indifference point on the delay discounting task (i.e., greater discounting of delayed rewards), and less time perspective on the future orientation scale. Adolescents scoring highly on this factor, therefore, tended to make bad decisions on a variety of in-lab tasks; they make errors and fail to learn from them. They also describe themselves as being present oriented with little regard for the future.

Table 3
Parameter Estimates from Four-Factor ESEM with Covariates.

	Standardized Factor Loadings											
	Premeditation			Fearlessness			Reward seeking			Cognitive dyscontrol		
Variables	*Loading*	*SE*	*p-value*	*Loading*	*SE*	*p-value*	*Loading*	*SE*	*p-value*	*Loading*	*SE*	*p-value*
UPPS: Urgency	**−0.21**	**0.05**	**<0.001**	**0.20**	**0.06**	**<0.001**	−0.09	0.09	0.33	0.10	0.08	0.19
UPPS: Premeditation	**0.80**	**0.03**	**<0.001**	0.00	0.03	0.91	−0.03	0.07	0.70	0.01	0.04	0.85
UPPS: Sensation Seeking	**−0.22**	**0.05**	**<0.001**	**0.21**	**0.05**	**<0.001**	**0.16**	**0.09**	**0.08**	−0.08	0.07	0.25
Future Orientation: Planning	**0.74**	**0.04**	**<0.001**	0.01	0.03	0.64	0.00	0.07	0.99	0.05	0.08	0.56
Future Orientation: Time Perspective	**0.47**	**0.05**	**<0.001**	−0.01	0.05	0.82	0.03	0.09	0.73	**−0.19**	**0.09**	**0.04**
Future Orientation: Future Consequences	**0.79**	**0.03**	**<0.001**	−0.05	0.03	0.17	0.01	0.05	0.83	0.00	0.04	0.91
Risk Perception: Perceived Harms	0.04	0.03	0.25	**−0.83**	**0.05**	**<0.001**	−0.04	0.06	0.42	0.07	0.02	0.01
Risk Perception: Perceived Benefits vs. Harms	0.02	0.02	0.31	**0.79**	**0.05**	**<0.001**	−0.03	0.04	0.48	0.12	0.06	0.04
IGT: Good decks	0.05	0.04	0.17	0.03	0.04	0.48	**0.56**	**0.14**	**<0.001**	−0.04	0.08	0.65
IGT: Bad decks	0.01	0.03	0.66	0.01	0.04	0.72	**0.22**	**0.09**	**0.02**	**0.42**	**0.07**	**<0.001**
Delay Discounting	0.05	0.05	0.30	−0.10	0.05	0.07	**0.15**	**0.09**	**0.10**	**−0.25**	**0.09**	**0.01**
BART: Avg. adjusted pumps	−0.10	0.06	0.09	−0.03	0.04	0.40	**0.47**	**0.09**	**0.00**	0.07	0.07	0.28
Stoplight: intersections	−0.02	0.06	0.80	0.06	0.06	0.28	**0.32**	**0.15**	**0.04**	0.09	0.11	0.41
TOL: Excess moves	0.03	0.04	0.52	−0.01	0.04	0.79	0.04	0.05	0.40	**0.42**	**0.06**	**<0.001**
TOL: Avg. time to first move	0.13	0.06	0.02	0.13	0.06	0.03	−0.07	0.11	0.56	**−0.20**	**0.10**	**0.06**

	Standardized Regressions on Covariates											
	Premeditation			Fearlessness			Reward seeking			Cognitive dyscontrol		
Covariates	*β*	*SE*	*p-value*	*β*	*SE*	*p-value*	*β*	*SE*	*p-value*	*β*	*SE*	*p-value*
Age (years)	**0.11**	**0.03**	**<0.001**	**0.17**	**0.03**	**<0.001**	0.05	0.05	0.31	**−0.18**	**0.07**	**0.01**
Male Sex (vs. Female)	**−0.28**	**0.10**	**<0.001**	**0.44**	**0.10**	**<0.001**	**0.50**	**0.21**	**0.02**	−0.35	0.25	0.16
Pubertal Development	0.01	0.05	0.88	0.01	0.05	0.88	**0.19**	**0.08**	**0.01**	**−0.20**	**0.08**	**0.01**
FSIQ	0.04	0.05	0.42	0.04	0.05	0.42	**0.27**	**0.08**	**<0.001**	**−0.59**	**0.06**	**<0.001**

	Residual Factor Correlations			
	Premeditation	Fearlessness	Reward seeking	Cognitive dyscontrol
PREMEDITATION	1.00			
FEARLESSNESS	**−0.47**	1.00		
REWARD SEEKING	−0.09	0.12	1.00	
COGNITIVE DYSCONTROL	−0.08	0.09	0.06	1.00

Note: Regression coefficients for FSIQ and pubertal development standardized with respect to both predictor and outcome. Effects of age and sex standardized with respect only to outcome. Standard errors in parentheses. Factor loadings with an absolute value greater than 0.15 are in bold type and were retained in subsequent behavioral genetic modeling (see Fig. 1). Regression coefficients and factor correlations that were significantly different than zero at $p < 0.05$ are in bold type.

8.1.3. *Results from 4-factor model: covariates*

The pattern of sex differences in the four factors was as expected, with females showing greater premeditation, lower fearlessness, and lower reward seeking than males. There was, however, no significant sex difference for cognitive dyscontrol. The four factors showed different age- and puberty-related mean trends. Importantly, because age and puberty were included as simultaneous predictors, effects of age and puberty are unique of one another. Compared to younger adolescents, older adolescents showed higher levels of premeditation and lower levels of cognitive dyscontrol. On the other hand, older adolescents also reported more fearlessness than younger adolescents. Finally, consistent with the dual systems model, which posits that the maturation of the incentive processing system is more closely tied to pubertal development than chronological age, reward seeking was not uniquely associated with age, but was positively associated with self-reported pubertal development incremental to age. Unexpectedly, cognitive dyscontrol also showed a correlation with pubertal development, with more developed adolescents showing better control. Finally, IQ was associated only with reward seeking and cognitive dyscontrol, with the latter association being especially strong.

Controlling for covariates, the residual correlations between the factors were mostly minimal, except for that between premeditation and fearlessness ($r = -0.47$). As these factors were largely determined by self-report surveys, the correlation might reflect shared method variance. However, it is notable that reward-seeking and cognitive dyscontrol (i.e., the factors on which behavioral measures had the largest loadings) were negligibly correlated with another, as this suggests that the ESEM results were not simply driven by method variance (self-report vs. behavioral).

8.2. *Behavior genetic analyses*

Table 2 summarizes the twin pair intraclass correlations for each measure, separately for MZ and DZ twin pairs. The twin correlations for FSIQ (0.76 in MZs, 0.48 in DZs) are highly consistent with previous behavioral genetic research on intelligence in twins (for meta-analyses see Tucker-Drob and Briley, 2014; Tucker-Drob et al., 2013). In all cases, the MZ twin correlation exceeded the DZ same-sex correlation, indicating genetic influence. For 9 out of 15 measures of self-control, the DZ twin correlation was less than half the MZ correlation, indicating negligible shared environmental influence.

The measurement portion of the behavioral genetic model was specified as illustrated in Fig. 1. For model parsimony, factor loadings with an absolute value less than 0.15 in the ESEM (Table 3) were fixed to zero, with one exception: Average time to first move on the Tower of London had two significant loadings (on premeditation and fearlessness) that were less than 0.15, whereas its loading on cognitive dyscontrol exceeded 0.15 but was not significant. In the behavior genetic model, therefore, average time to first move was allowed to load on all three of these factors.[1]

[1] We tested an alternative model that used stricter criteria for including a measures as an indicator of a latent factor (loadings $\geq |.20|$ at $p < 0.05$). However, this

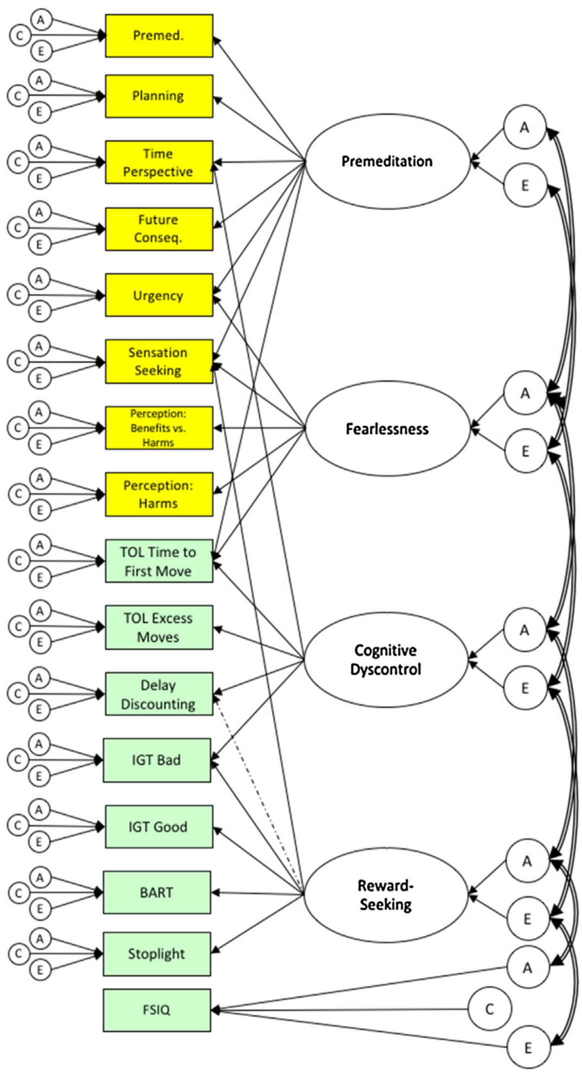

Fig. 1. Behavior Genetic Model.
Note: Double-lined arrows indicate that all correlations between *A* factors and all correlations between *E* factors were estimated. Self-report measures in yellow; behavioral measures in green. *A* = additive genetic, *C* = shared environmental, *E* = non-shared environmental. Age and sex were controlled by regressing these variables from the factors. All factor loadings were estimated in the final model to be significantly different from zero at $p < 0.05$, except for the loading of delay discounting on reward seeking (shown with a dashed line).

Factor variances and DV residual variances (i.e., test-specific variance not shared with the factor) were decomposed into genetic, shared environmental, and non-shared environmental components. This model also decomposed the correlations among the factors and the correlations between the factors and FSIQ into genetic, shared environmental, and non-shared environmental components. Behavior genetic models controlled for age and sex by partialling these covariates from the factors.

Preliminary models that estimated the intraclass correlations for the four latent factors in MZs versus DZs indicated that shared environmental influences on the factors were minimal (MZ correlations were no less than twice the DZ correlations); shared environmental variances in the factors, therefore, were not estimated in order to facilitate model convergence. Additionally, if in initial models a unique biometric variance component was estimated to be negative and not significantly different from zero, this component was fixed to zero for the final model estimation.

Results from the behavioral genetic analysis are summarized in Table 4. The model fit the data well (RMSEA = 0.041). Overall, factor-level variances were more heritable (ranging from 31% for premeditation to 66% for reward-seeking) than the task-specific variances (ranging from 0% to 38% of the total variance in the DV, median = 11%). Most of the task-specific variance was non-shared environmental in origin (differing even within identical twins), a quantity that includes measurement error. The relationships between FSIQ and the factors were primarily attributable to shared genetic influences. That is, the genetically-influenced variation in FSIQ was positively correlated with genetically-influenced variation in fear/avoidance, cognitive dyscontrol, and reward-seeking, whereas the environmentally-influenced variances in these phenotypes were largely unrelated to each other. This was particularly striking for FSIQ and cognitive dyscontrol; the genetic correlation between these constructs exceeded −0.9.

9. Discussion

This study investigated the factor structure and genetics of a multivariate battery of self-report and behavioral measures relevant to risk-taking in a sample of adolescent twins. We examined (a) the extent to which tasks that are commonly interpreted in terms of the same underlying construct can, in fact, be considered indicators of the same latent factor and (b) the extent to which phenotypically correlated measures tap the same underlying genetic signal. Overall, we found that the structure of the battery was quite complex. Four dimensions were required to account for the pattern of covariation among the measures, and several measures had cross-loadings on multiple dimensions. Most behavioral measures contained large amounts of unique (task-specific) variance; however, most *genetic* variance in each measure was shared with other measures, with factor heritabilities ranging from 0.31 to 0.66.

9.1. Epistemological issues in latent factor models

Before we discuss the specific results, it is important to contextualize the findings with a number of general caveats about interpreting latent factor models. First, although a four-factor solution was the best and most parsimonious representation of the data, the number of factors extracted depends on which (and how many) DVs are entered into the model. Our measurement battery sampled broadly from commonly used behavioral tasks and self-report surveys, but the sensitivity of these results to the inclusion of alternative measures is an empirical question to be interrogated in future research. The behavioral tasks, in particular, are flexible measures that can be scored in a variety of ways, and the specific DV used for each varies across studies and labs. For example, performance on the BART can be measured using the average adjusted number of pumps (as done in the original study by Lejuez et al. (2002), and in the current paper), number of explosions (e.g., Braams et al., 2015), total money earned (e.g., Peper et al., 2013), or, less commonly, in terms of parameters estimated from a cognitive decision model (e.g., Ashenhurst et al., 2014; Wallsten et al., 2005; van Ravenzwaaij et al., 2011).

Flexibility in scoring and data analysis complicates issues surrounding scientific validity and reproducibility. The evidence supporting either the interchangeability (or the differential validity) of alternative measures of task performance is often quite thin. In the absence of a strong consensus among researchers regarding optimal scoring, the availability of alternative scoring procedures increases "researcher degrees of freedom" and the likelihood of

alternative model fit significantly worse than the model shown in Fig. 1 (Satorra-Bentler χ^2 difference test = 6.85, $p = 0.009$). This suggests that even small factor loadings were necessary to account for the observed covariances.

Table 4
Parameter Estimates from Behavioral Genetic Models.

	Variances								
	Additive Genetic (A)			Shared Environmental (C)			Non-Shared Environmental (E)		
Measure	Variance	*SE*	*p*-value	Variance	*SE*	*p*-value	Variance	*SE*	*p*-value
Factor									
Premeditation	**0.313**	**0.089**	**<0.001**				**0.687**	**0.089**	**<0.001**
Fearlessness	**0.355**	**0.103**	**0.001**				**0.645**	**0.103**	**<0.001**
Cognitive Dyscontrol	**0.626**	**0.230**	**0.006**				0.374	0.230	0.103
Reward-Seeking	**0.664**	**0.206**	**0.001**				0.336	0.206	0.104
FSIQ	**0.719**	**0.127**	**<0.001**	0.040	0.113	0.726	**0.241**	**0.036**	**<0.001**
Test-Specific									
UPPS: Urgency	0.282	0.170	0.098	0.028	0.137	0.838	**0.541**	**0.062**	**<0.001**
UPPS: Premeditation	**0.113**	**0.039**	**0.004**				**0.278**	**0.041**	**<0.001**
UPPS: Sensation Seeking	**0.379**	**0.066**	**<0.001**				**0.442**	**0.060**	**<0.001**
Future Orientation: Planning	0.045	0.045	0.322				**0.417**	**0.053**	**<0.001**
Future Orientation: Time Perspective				0.034	0.042	0.421	**0.712**	**0.055**	**<0.001**
Future Orientation: Future Consequences	0.021	0.036	0.555				**0.332**	**0.050**	**<0.001**
Risk Perception: Perceived Benefits vs. Risks	0.030	0.048	0.524				**0.445**	**0.106**	**<0.001**
Risk Perception: Perceived Harms				**0.097**	**0.042**	**0.021**	0.109	0.058	0.061
TOL: Excess moves	0.101	0.202	0.618	0.060	0.152	0.691	**0.633**	**0.085**	**<0.001**
TOL: Avg. time to first move	**0.233**	**0.072**	**0.001**				**0.689**	**0.091**	**<0.001**
Delay Discounting	0.206	0.194	0.290	0.110	0.142	0.438	**0.555**	**0.074**	**<0.001**
IGT: Bad decks	0.154	0.086	0.073				**0.676**	**0.083**	**<0.001**
IGT: Good decks	0.103	0.214	0.632	0.053	0.157	0.736	**0.618**	**0.093**	**<0.001**
BART: Avg. adjusted pumps	0.151	0.078	0.054				**0.683**	**0.096**	**<0.001**
Stoplight: intersections	0.153	0.079	0.053				**0.775**	**0.098**	**<0.001**

Factor Correlations [b]	Premeditation	Fearlessness	Cognitive Dyscontrol	Reward-Seeking	FSIQ
Premeditation	1	0.352 (0.119)	−0.251 (0.201)	−0.214 (0.225)	0.149 (0.088)
Fearlessness	0.746 (0.221)	1	−0.056 (0.173)	−0.151 (0.200)	−0.047 (0.082)
Cognitive Dyscontrol	−0.031 (0.221)	−0.415 (0.238)	1	0.671 (0.468)	0.118 (0.160)
Reward-Seeking	−0.502 (0.232)	−0.132 (0.233)	−0.319 (0.315)	1	0.059 (0.204)
FSIQ	0.102 (0.111)	0.302 (0.139)	−0.909 (0.208)	0.352 (0.119)	1

[a]Test-specific variances represent proportions of the total (standardized) variance in the DV.
Bold values are significant at $p < 0.05$.
[b] Genetic correlations are below the diagonal; non-shared environmental correlations are above the diagonal. *SE*s are reported in parentheses.

obtaining false positive results (Neuroskeptic, 2016; Simmons et al., 2011). In the current study, we made an effort to select our scoring procedures *ex ante*. If we have used this task in our previously published research, then we scored the current data to be consistent with our papers. Otherwise, we selected a scoring method that we judged to be the most common method in the literature.[2] Our results, therefore, are intended to illuminate the relations among measures as they are frequently used by researchers in this area (including by our own labs), but these results do not necessarily apply to alternative scorings of the behavioral tasks. Future research using these flexible measures may benefit from pre-registering DV selection and planned analyses.

Second, the labeling of latent factors is subjective, and these labels can and should be updated and refined with future research (Cronbach and Meehl, 1955). For instance, we labeled the factor representing covariance among the IGT, BART, Stoplight, and self-reported sensation seeking as “reward seeking”. This label was selected, because (a) behaviors on the in-lab tasks were motivated by the possibility of a reward, and (b) this label is consistent with how other investigators commonly interpret performance on these tests. Yet these behavioral tasks also involve some learning component, as individuals learn the reinforcement structure of the task. If future research finds that reward sensitivity tasks that do *not* involve learning also hang together with IGT, BART, and Stoplight tasks as indicators of the same factor, this would increase confidence in the “reward seeking” label, versus alternative conceptualizations of what these measures have in common.

Third, it is important to remember that latent factors (and correlations among latent factors) can emerge via multiple different mechanistic processes, and delineating these mechanisms cannot be accomplished with only cross-sectional, correlational data. To use an example from clinical psychology, people have debated whether comorbidity among psychiatric disorders, which can be represented statistically as a latent factor, represents the causal effect of a single underlying process on multiple disorders or emerges from a network of bidirectional relationships among specific symptoms (Cramer et al., 2010, see also commentary by Molenaar, 2010). In the same way, the current study documents how widely-used measures relevant to understanding risk-taking cluster together in adolescents, but how these associations emerge with development remains to be understood.

Overall, the current study is informative regarding how a number of commonly used measures converge and diverge, and should be viewed as a springboard for further refinement. In particular, how the constructs studied here are related to risk-taking outside the lab is a key step for establishing external validity, while longitudinal data will be necessary to trace how the covariances observed here emerge with development.

9.2. Cognitive (Dys)control and the ability-personality distinction

Ability and personality are the two major foci of individual differences research. The assessment of ability aims to characterize what a person *can* do, i.e., what is her maximum level of performance? In contrast, the assessment of personality (Cronbach, 1960) focuses on what a person *typically* does, i.e., what is her usual behavior? Within developmental cognitive neuroscience, conceptions of “cognitive control” often blend aspects of both personality and abil-

[2] See the Supplement for a more extended discussion of the choice of metric for the delay discounting task (average indifference point vs. discount rate).

ity. For example, Luna et al. (2015) describe cognitive control as "an *ability* to flexibly, voluntarily, and adaptively coordinate behavior in the service of internal goals..." (p. 152, emphasis added). They conceptualize the "core components" of cognitive control as inhibitory control, performance monitoring, and working memory – all abilities that are investigated in the behavioral literature as components of executive function (e.g., Engelhardt et al., 2015). In contrast, Shulman et al. (2015a,b) defined cognitive control more broadly as "undergird[ing] self-regulatory behavior" and measured it using self-report personality items that asked how people usually behave (e.g., "I often get into a jam because I do things without thinking.")

Certainly, both ability and personality are maturing in adolescence. However, our results indicate that cognitive dyscontrol (i.e., lacking the ability to learn from errors and avoid mistakes) is a separable dimension from premeditation (i.e., the motivated tendency to plan for the future and consider potential consequences). The phenotypic and genetic correlations between these dimensions were negligible. Furthermore, cognitive dyscontrol was strongly related to lower IQ, whereas the relationship between IQ and premeditation was modest. Future theoretical and empirical work should avoid conflating these separable dimensions of maximal ability and typical behavior when examining adolescent improvements in "cognitive control" and their relation to risk-taking behavior.

The phenotypic correlation between the cognitive dyscontrol factor and IQ masked an even higher genetic correlation, which was nearly perfect ($r = -0.9$) in this sample. This result has several intriguing implications. The genetic architecture of IQ is the subject of intense study, and specific genetic correlates of IQ are beginning to be uncovered (Davies et al., 2015; Ibrahim-Verbaas et al., 2016; Okbay et al., 2016a,b; Rietveld et al., 2013; Rietveld et al., 2014). Currently, one can predict childhood reading skills, adolescent educational achievement, adult financial success, and longevity from an individual's DNA using polygenic scores derived from GWAS (Belsky et al., 2016; Marioni et al., 2016; Selzam et al., 2016). The high genetic correlation between IQ and the cognitive dyscontrol factor suggests that these polygenic scores may also predict aspects of impulse control and impulse control disorders. On the flip side, the practice of "controlling for" IQ, by partialling out the variance in behavioral tasks that is shared with IQ, may be removing a substantial portion of the genetically-influenced variation in cognitive (dys)control, much of the variation that is shared with other measures of the same putative construct, and much of the variation that undergirds its relations with key outcomes such as risk-taking.

Interestingly, IQ was also *positively* associated with the factor that we labeled "reward seeking," although the magnitude of this correlation was substantially less than what was observed for the cognitive dyscontrol factor. The positive association between reward seeking and IQ observed here is consistent with previous studies of adolescents and adults, which have also found positive correlations between sensation seeking and IQ (Carrol and Zuckerman, 1977; Cohen et al., 1983; Kish and Busse, 1968; Russo et al., 1993), and might reflect (at least) two different processes. First, as mentioned previously, over the course of the IGT and BART, participants learn about the likelihood that a given behavior will be rewarded versus punished. As intelligence can be broadly conceptualized as the capacity to learn from experience (Neisser et al., 1996), it is not surprising that more intelligent adolescents learned how to earn a greater reward from the task. Indeed, the behaviors that we have labeled as "reward seeking" *were* generally rewarded.[3]

Second, higher reward seeking and/or sensation seeking may positively contribute to the development of intelligence, because youth with high levels of these traits are expected to seek out and create more varied environments for themselves (Raine et al., 2002). In children, ratings of "stimulation seeking" at age 3 (defined as exploring the physical environment, degree of verbalization and friendliness with an unfamiliar research assistant, and degree of active social play with other children) longitudinally predicted IQ at age 11 (Raine et al., 2002). Of course, in keeping with the thesis of this paper, the relationship between "stimulation seeking" in childhood and sensation seeking or reward seeking in adolescence needs to be clarified.

9.3. Reflexive versus reflective inhibition

Both cognitive dyscontrol and premeditation were further separable from yet another dimension, which we termed fearlessness, and which was primarily defined by considering the potential negative consequences of various risky activities to be "dangerous" and "scary." Low levels of fearlessness are expected to be inhibitory with regard to risk-taking behavior; adolescents who report being scared of negative consequences are less likely to engage in risky activities. Yet fearful behavior should not be conflated with self-controlled behavior, as there is a key difference between reflectively considering and avoiding negative future consequences versus reflexively shrinking from aversive stimuli. This distinction is consistent with Carver and colleagues' "two-mode model of self-regulation" (Carver et al., 2008, 2009), which distinguishes between "deliberative effortful control" and "reflexive punishment sensitivity." Both dimensions are hypothesized to be countervailing forces to "reflexive reward sensitivity;" however, for the fearful, "punishment sensitive" adolescent, deliberate control is necessary to overcome impulsive *inaction*. The separation between the dimensions of reward seeking and fear(lessness) in our results is further consistent with Ernst's (2014) triadic model of adolescent neurodevelopment, which, in addition to incentive processing and cognitive control, also highlights the development of an avoidance system, centered around the amygdala, hippocampus, and insula.

Notably, the age trends in fearlessness and cognitive dyscontrol were both positive: Older adolescents had better cognitive control, but also believed that the potential negative consequences of risk behaviors were less scary and dangerous. The positive age trend for fearlessness might be due, at least in part, to older adolescents having more experience with the behaviors in question. Models of extinction imply that exposure to a behavior or experience decreases fear (including fearful cognitions) about that behavior (Monfils et al., 2009). Additionally, opposing trends suggest that the "brakes" on adolescent risk-taking behavior might switch with development from more reflexive avoidance to more reflective control.

9.4. Sensation seeking: at the intersection of reward seeking and fearlessness

Self-reported sensation seeking had loadings on both the reward seeking factor and the fearlessness factor. This pattern of cross-loadings is intuitively sensible; finding dangerous activities enjoyable requires both not being overly frightened of the activ-

[3] The relation between risk-taking and reward on the BART is an inverse-U, and reward is maximized on the BART when pumps equal 64. The mean (average adjusted) pumps in our sample was around 30, and only 5 participants showed risk-taking on the BART in the range (pumps >64) in which increased risk-taking is actually disadvantageous. The correlation between pumps and money earned on the task was 0.90 ($p < 0.001$). Previous studies have also found that very few participants take disadvantageous risks on the BART; a meta-analysis of 22 studies found that the average adjusted pumps ranged from 24.6 to 44.1 across studies; the fixed effect meta-analytic estimate was 33.1.

ity and some motivation to approach rewards. Furthermore, this result is consistent with previous studies of sensation seeking and fearlessness, which have found that high sensation seekers report lower levels of trait anxiety and report feeling less fear (Franken et al., 1992), and show attenuated skin conductance and startle responses to aversive stimuli (Lissek et al., 2005).

Unlike the BART, IGT, and Stoplight games, the residual variance in sensation seeking (i.e., variance not shared with the factor) was also substantially heritable. The finding of genetic variance that is unique to self-reported sensation seeking is particularly important, given previous reports finding very strong genetic correlations between sensation seeking and delinquent behavior in adolescence (Harden et al., 2012a; Harden et al., 2012b; Mann et al., 2016). Self-reported sensation seeking appears to tap a unique genetic signal that is not shared with any other self-report or behavioral measure. For researchers interested in the neural underpinnings of sensation seeking, there is a need for behavioral tasks that can be adapted to the scanner and that adequately capture genetically-influenced individual differences in sensation seeking. One promising approach comes from Norbury (2015), who adapted a behavioral paradigm used in animal research, and presented participants with a series of trials in which they could win points or receive a "stimulating but not painful" electric shock. Across two studies, the value participants placed on receiving this electric stimulation was correlated with UPPS sensation seeking at $r \sim 0.3$, suggesting that this task may outperform many existing gambling paradigms in its convergence with self-report measures.

Interestingly, the reward-seeking factor was associated with pubertal development rather than with age. This finding is consistent with the predictions of the dual systems model (Icenogle et al., in press; Smith et al., 2013) and with neuroimaging studies that have found associations between puberty-related gonadal hormones and neural responses to reward (Braams et al., 2015; Op de Macks et al., 2011; Peper et al., 2013). Finally, the residual, task-specific variance in each reward-seeking behavioral task was quite substantial, and this task-specific variance was nearly entirely non-shared environmental in origin, a quantity that includes measurement error. Given that any behavioral task in isolation is a fairly poor indicator of the underlying construct it purports to measure, the common practice of relying on a single task to operationalize the construct of reward-seeking is clearly not supported.

9.5. *Individual-specific vs. family-wide environmental influences on self-control*

In behavioral genetic analyses, shared environmental influences on each of the four factors, as well as on IQ and on most individual measures, were minimal. This result indicates that, *within the range of environments sampled in this study,* endogenous variation in family-level environments – a category that includes the common home rearing environment, as well as neighborhood and school contexts shared by twins raised together – does not have an appreciable influence on factors presumed to influence risk-taking. The finding of minimal shared environmental influence could potentially be concerning for education policymakers, who have proposed using measures of self-regulation or "grit" as indicators of school quality. As a columnist for *Education Week* noted in response to the newest version of the federal Elementary and Secondary Education Act, "if states select [school quality] indicators that can't be accurately measured or influenced by the schools... the indicator requirement could lead to unintended consequences" (Blad, 2016). On the other hand, these results are informative only regarding the effects of *differences* in family-wide environments that are *naturally occurring within this sample.* The effects of environmental interventions or changes that are novel or outside the range of what is already being experienced, or the effects of environments (such as being enrolled in a public school in Texas) that are universal among the adolescents in this sample cannot be assessed in the current study.

Additionally, these estimates of zero shared environmental influence depend on the validity of the assumptions of the twins-reared-together design (see Neale and Maes, 2004), including (1) the equal environments assumption that identical twins are not treated more similarly, with regard to environments that affect the phenotype, simply because they are known to be identical, and (2) there is no assortative mating on the phenotype of interest in the parental generation. These assumptions are credible, because (1) previous studies have found that misidentified twins (i.e., twins who are genetically identical believe themselves to be fraternal) are as similar as accurately classified identical twins (Conley et al., 2013), and (2) non-zero assortative mating would lead to an over-estimate of c^2. Finally, analysis of data from twins-reared-together cannot simultaneously estimate both dominant genetic influences (*D*, which inflate the similarity of identical twins relative to all other pairs of biological relatives) and shared environmental influence. Previous research has found evidence for *D* variance in personality (Keller et al., 2005), and the MZ twin correlation exceeded twice the DZ correlations for many of the self-control phenotypes measured in this study, suggesting the need for future behavioral genetic research on self-control to incorporate data from more extended pedigrees (e.g., non-twin siblings, parents).

10. Conclusions

The dual systems model has been an enormously influential theoretical account of why adolescence is generally a time of heightened risk-taking. As research in this area has matured, commentators have called for greater specificity in dual systems model predictions, in an effort to increase falsifiability (Pfeifer and Allen, 2016) and to avoid overly vague heuristics (Casey et al., 2016). Our research used a psychometric, genetically-informed approach to analyze behavioral and self-report data relevant to risk-taking in adolescence. Our results extend and refine dual systems model research in four ways. First, the personality components of "cognitive control" are separable from ability components, despite being often conflated in descriptions of how adolescents mature. Second, it is problematic to operationalize a key theoretical construct using a single behavioral task, as most of the variance in any one task is non-systematic, shared neither across tasks nor across family members, and ambiguous in its level of specificity. Third, genetically-influenced variation in self-reported sensation seeking, which increases in adolescence and is one of the best predictors of involvement in socially problematic forms of risk-taking, is not adequately captured by commonly used gambling tasks. Fourth, some aspects of "cognitive control," including delay discounting, choosing bad decks on the IGT, self-reported urgency, and performance on the Tower of London, are strongly associated with genetically-influenced variation in IQ. As many labs have collected similar data as are described here, we look forward to learning whether the nomological network described in this paper replicates across groups.

Conflict of interest

None.

Acknowledgements

K.P.H. and E.M.T.D. are Faculty Research Associates of the Population Research Center at the University of Texas at Austin, which is supported by a grant, 5-R24-HD042849, from the Eunice Kennedy

Shriver National Institute of Child Health and Human Development. This research was supported by grants R21-AA023322 and R21-AA020588from the National Institute on Alcohol Abuse and Alcoholism, and by the Russell Sage Foundation Visiting Scholar program.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.dcn.2016.12.007.

References

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. Behav. Res. Ther. 43 (2), 215–228.

Ashenhurst, J.R., Bujarski, S., Jentsch, J.D., Ray, L.A., 2014. Modeling behavioral reactivity to losses and rewards on the Balloon Analogue Risk Task (BART): moderation by alcohol problem severity. Exp. Clin. Psychopharmacol. 22 (4), 298–306.

Asparouhov, T., Muthén, B., 2009. Exploratory structural equation modeling. Struct. Eq. Model. A Multidiscip. J. 16 (3), 397–438.

Bechara, A., Damasio, A.R., Damasio, H., Anderson, S.W., 1994. Insensitivity to future consequences following damage to human prefrontal cortex. Cognition 50 (1–3), 7–15, http://dx.doi.org/10.1016/0010-0277(94)90018-3.

Bechara, A., 2007. Iowa Gambling Task (IGT) Professional Manual. Lutz, FL: Psychological Assessment Resources.

Bechara, A., Dolan, S., Denburg, N., Hindes, A., Anderson, S.W., Nathan, PE, 2001. Decision-making deficits, linked to a dysfunctional ventromedial prefrontal cortex, revealed in alcohol and stimulant abusers. Neuropsychologia 39 (4), 376–389.

Belsky, D.W., Moffitt, T.E., Corcoran, D.L., Domingue, B., Harrington, H., Hogan, S., et al., 2016. The genetics of success how single-nucleotide polymorphisms associated with educational attainment relate to life-course development. Psychol. Sci. 27, 957–972.

Benthin, A., Slovic, P., Severson, H., 1993. A Psychometric study of adolescent risk perception. J. Adolesc. 16 (2), 153–168, http://dx.doi.org/10.1006/jado.1993.1014.

Bickel, W.K., Marsch, L.A., 2001. Toward a behavioral economic understanding of drug dependence: delay discounting processes. Addiction 96 (1), 73–86, http://dx.doi.org/10.1046/j.1360-0443.2001.961736.x.

Blad, E. (January 6, 2016). ESSA law broadens definition of student success. Education Week. (Accessed 02 August 2016) at http://www.edweek.org/ew/articles/2016/01/06/essa-law-broadens-definition-of-school-success.html.

Braams, B.R., van Duijvenvoorde, A.C.K., Peper, J.S., Crone, E.A., 2015. Longitudinal changes in adolescent risk-taking: a comprehensive study of neural responses to rewards, pubertal development, and risk-taking behavior. J. Neurosci. 35 (18), 7226–7238, http://dx.doi.org/10.1523/JNEUROSCI.4764-14.2015.

Briley, D.A., Tucker-Drob, E.M., 2014. Genetic and environmental continuity in personality development: a meta-analysis. Psychol. Bull. 140 (5), 1303–1331, http://dx.doi.org/10.1037/a0037091.

Buelow, M.T., Blaine, A.L., 2015. The assessment of risky decision making: a factor analysis of performance on the iowa gambling task, balloon analogue risk task, and columbia card task. Psychol. Assess. 27 (3), 777–785, http://dx.doi.org/10.1037/a0038622.

Campbell, D.T., Fiske, D.W., 1959. Convergent and divergent validation by the multitrait-mulitmethod matrix. Psychol. Bull. 56, 81–105, http://dx.doi.org/10.1037/h0046016.

Carrol, E.N., Zuckerman, M., 1977. Psychopathology and sensation seeking in downers, speeders: and trippers: a study of the relationship between personality and drug choice. Int. J. Addict. 12, 591–601.

Carver, C.S., White, T.L., 1994. Behavioral inhibition, behavioral activation, and affective responses to impending reward and punishment: the BIS/BAS scales. J. Pers. Soc. Psychol. 67 (2), 319–333, http://dx.doi.org/10.1037/0022-3514.67.2.319.

Carver, C.S., Johnson, S.L., Joormann, J., 2008. Serotonergic function, two-mode models of self-regulation, and vulnerability to depression: what depression has in common with impulsive aggression. Psychol. Bull. 134 (6), 912.

Carver, C.S., Johnson, S.L., Joormann, J., 2009. Two-mode models of self-regulation as a tool for conceptualizing effects of the serotonin system in normal behavior and diverse disorders. Curr. Direct. Psychol. Sci. 18 (4), 195–199.

Casey, B.J., 2015. Beyond simple models of self-control to circuit-based accounts of adolescent behavior. Annu. Rev. Psychol. 66, 295–319.

Casey, B.J., Jones, R.M., Somerville, L.H., 2011. Braking and accelerating of the adolescent brain. J. Res. Adolesc. 21 (1), 21–33, http://dx.doi.org/10.1111/j.1532-7795.2010.00712.x.

Casey, B., Galván, A., Somerville, L.H., 2016. Beyond simple models of adolescence to an integrated circuit-based account: a commentary. Dev. Cogn. Neurosci. 17, 128–130, http://dx.doi.org/10.1016/j.dcn.2015.12.006.

Cauffman, E., Shulman, E.P., Steinberg, L., Claus, E., Banich, M.T., Graham, S., Woolard, J., 2010. Age differences in affective decision making as indexed by performance on the Iowa Gambling Task. Dev. Psychol. 46 (1), 193–207, http://dx.doi.org/10.1037/a0016128.

Centers for Disease Control and Prevention, 2015. Adolescent Health: Leading causes of death. http://www.cdc.gov/nchs/fastats/adolescent-health.htm.

Cohen, L., Dingemans, P.M.A.J., Lesnik-Oberstein, M., van der Vlugt, H., 1983. Report of the findings with the sensation seeking scale in Holland. Poster Session Presented at the 91st Annual Convention of the American Psychological Association, Anaheim, CA.

Conley, D., Rauscher, E., Dawes, C., Magnusson, P.K., Siegal, M.L., 2013. Heritability and the equal environments assumption: evidence from multiple samples of misclassified twins. Behav. Genet. 43 (5), 415–426.

Cramer, A.O., Waldorp, L.J., van der Maas, H.L., Borsboom, D., 2010. Complex realities require complex theories: refining and extending the network approach to mental disorders. Behav. Brain Sci. 33 (2–3), 178–193.

Cronbach, L.J., Meehl, P.E., 1955. Construct validity in psychological tests. Psychol. Bull. 52 (4), 281–302, http://dx.doi.org/10.1037/h0040957.

Cronbach, L.J., 1960. Essentials of Psychological Testing, 2nd ed. Harper and Row, New York.

Cyders, M.A., Coskunpinar, A., 2011. Measurement of constructs using self-report and behavioral lab tasks: is there overlap in nomothetic span and construct representation for impulsivity? Clin. Psychol. Rev. 31 (6), 965–982, http://dx.doi.org/10.1016/j.cpr.2011.06.001.

Davies, G., Armstrong, N., Bis, J.C., Bressler, J., Chouraki, V., Giddaluru, S., et al., 2015. Genetic contributions to variation in general cognitive function: a meta-analysis of genome-wide association studies in the CHARGE consortium (N = 53949). Mol. Psychiatry 20 (2), 183–192, http://dx.doi.org/10.1038/mp.2014.188.

Duckworth, A.L., Steinberg, L., 2015. Unpacking self-control. Child Dev. Perspect. 9 (1), 32–37.

Engelhardt, L.E., Briley, D.A., Mann, F.D., Harden, K.P., Tucker-Drob, E.M., 2015. Genes unite executive functions in childhood. Psychol. Sci. 26 (8), 1151–1163, http://dx.doi.org/10.1177/0956797615577209.

Ernst, M., 2014. The triadic model perspective for the study of adolescent motivated behavior. Brain Cogn. 89, 104–111.

Franken, R.E., Gibson, K.J., Rowland, G.L., 1992. Sensation seeking and the tendency to view the world as threatening. Personal. Individ. Differ. 13 (1), 31–38, http://dx.doi.org/10.1016/0191-8869(92)90214-A.

Galvan, 2010. Adolescent development of the reward system. Front. Hum. Neurosci., http://dx.doi.org/10.3389/neuro.09.006.2010.

Ge, T., Schumann, G., Feng, J., 2013. Imaging genetics—towards discovery neuroscience. Quant. Biol. 1 (4), 227–245.

Geier, C.F., 2013. Adolescent cognitive control and reward processing: implications for risk taking and substance use. Horm. Behav. 64 (2), 333–342, http://dx.doi.org/10.1016/j.yhbeh.2013.02.008.

Green, L., Myerson, J., 2004. A discounting framework for choice with delayed and probabilistic rewards. Psychol. Bull. 130 (5), 769–792, http://dx.doi.org/10.1037/0033-2909.130.5.769.

Harden, K.P., Mann, F.D., 2015. Biological risk for the development of problem behavior in adolescence: integrating insights from behavioral genetics and neuroscience. Child Dev. Perspect. 9 (4), 211–216.

Harden, K.P., Tucker-Drob, E.M., 2011. Individual differences in the development of sensation seeking and impulsivity during adolescence: further evidence for a dual systems model. Dev. Psychol. 47 (3), 739–746, http://dx.doi.org/10.1037/a0023279.

Harden, K.P., Quinn, P.D., Tucker-Drob, E.M., 2012a. Genetically influenced change in sensation seeking drives the rise of delinquent behavior during adolescence. Dev. Sci. 15 (1), 150–163, http://dx.doi.org/10.1111/j.1467-7687.2011.01115.x.

Harden, K.P., Quinn, P.D., Tucker-Drob, E.M., 2012b. Genetically influenced change in sensation seeking drives the rise of delinquent behavior during adolescence. Dev. Sci. 15 (1), 150–163.

Harden, K.P., Tucker-Drob, E.M., Tackett, J.L., 2013. The texas twin project. Twin Res. Hum. Genet. 16 (Special Issue 01), 385–390, http://dx.doi.org/10.1017/thg.2012.97.

Harris, P.A., Taylor, R., Thielke, R., Payne, J., Gonzalez, N., Conde, J.G., 2009. Research electronic data capture (REDCap)—a metadata-driven methodology and workflow process for providing translational research informatics support. J. Biomed. Inform. 42 (2), 377–381, http://dx.doi.org/10.1016/j.jbi.2008.08.010.

Heath, A.C., Nyholt, D.R., Neuman, R., Madden, P.A.F., Bucholz, K.K., Todd, R.D., et al., 2003. Zygosity diagnosis in the absence of genotypic data: an approach using latent class analysis. Twin Res. Hum. Genet. 6 (01), 22–26, http://dx.doi.org/10.1375/twin.6.1.22.

Heckman, J.J., Rubinstein, Y., 2001. The importance of noncognitive skills: lessons from the GED testing program. Am. Econ. Rev. 91 (2), 145–149.

Ibrahim-Verbaas, C.A., Bressler, J., Debette, S., Schuur, M., Smith, A.V., Bis, J.C., et al., 2016. GWAS for executive function and processing speed suggests involvement of the CADM2 gene. Mol. Psychiatry 21 (2), 189–197, http://dx.doi.org/10.1038/mp.2015.37.

Icenogle, G., Steinberg, L., Olino, T., Shulman, E., Chein, J., Alampay, L., 2016. Puberty predicts approach but not avoidance on the Iowa Gambling Task in a multinational sample. Child Dev. (in press).

Keller, M.C., Coventry, W.L., Heath, A.C., Martin, N.G., 2005. Widespread evidence for non-additive genetic variation in Cloninger's and Eysenck's personality dimensions using a twin plus sibling design. Behav. Genet. 35, 707–721.

Kim-Spoon, J., Kahn, R., Deater-Deckard, K., Chiu, P., Steinberg, L., King-Casas, B., 2016. Risky decision making in a laboratory driving task is associated with

health risk behaviors during late adolescence but not adulthood. Int. J. Behav. Dev. 40 (1), 58–63, http://dx.doi.org/10.1177/0165025415577825.
Kish, G.B., Busse, W., 1968. Correlates of stimulus-seeking: age, education, intelligence: and aptitudes. J. Consult. Clin. Psychol. 32, 633–637.
Krueger, R.F., Hicks, B.M., Patrick, C.J., Carlson, S.R., Iacono, W.G., McGue, M., 2002. Etiologic connections among substance dependence: antisocial behavior and personality: modeling the externalizing spectrum. J. Abnorm. Psychol. 111, 411–424.
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). J. Exp. Psychol. Appl. 8 (2), 75–84.
Lejuez, C.W., Aklin, W.M., Zvolensky, M.J., Pedulla, C.M., 2003. Evaluation of the Balloon Analogue Risk Task (BART) as a predictor of adolescent real-world risk-taking behaviours. J. Adolesc. 26 (4), 475–479, http://dx.doi.org/10.1016/S0140-1971(03)00036-8.
Lejuez, C.W., Aklin, W., Daughters, S., Zvolensky, M., Kahler, C., 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. J. Clin. Child Adolesc. Psychol. 36 (1), 106–111, http://dx.doi.org/10.1080/15374410709336573.
Lissek, S.P.M., Pine, D.S., Orme, K., Dvir, S., Rosenberger, E., Grillon, C., 2005. Sensation seeking and the aversive motivational system. Emotion 5 (4), 396–407, http://dx.doi.org/10.1037/1528-3542.5.4.396.
Luna, B., Marek, S., Larsen, B., Tervo-Clemmens, B., Chahal, R., 2015. An integrative model of the maturation of cognitive control. Annu. Rev. Neurosci. 38 (1), 151–170, http://dx.doi.org/10.1146/annurev-neuro-071714-034054.
MacKillop, J., Amlung, M.T., Few, L.R., Ray, L.A., Sweet, L.H., Munafò, M.R., 2011. Delayed reward discounting and addictive behavior: a meta-analysis. Psychopharmacology 216 (3), 305–321, http://dx.doi.org/10.1007/s00213-011-2229-0.
MacKillop, J., Weafer, J., Gray, J.C., Oshri, A., Palmer, A., de Wit, H., 2016. The latent structure of impulsivity: impulsive choice, impulsive action: and impulsive personality traits. Psychopharmacology 233, 3361–3370.
Mann, F.D., Patterson, M.W., Grotzinger, A.D., Kretsch, N., Tackett, J.L., Tucker-Drob, E.M., Harden, K.P., 2016. Sensation seeking, peer deviance: and genetic influences on adolescent delinquency: evidence for person-environment correlation and interaction. J. Abnorm. Psychol. 125, 679–691.
Manolio, T.A., Collins, F.S., Cox, N.J., Goldstein, D.B., Hindorff, L.A., Hunter, D.J., et al., 2009. Finding the missing heritability of complex diseases. Nature 461 (7265), 747–753.
Marioni, R.E., Ritchie, S.J., Joshi, P.K., Hagenaars, S.P., Okbay, A., Fischer, K., Amador, C., 2016. Genetic variants linked to education predict longevity. Proc. Nat. Acad. Sci. 113 (47), 13366–13371.
Marsh, H.W., Morin, A.J., Parker, P.D., Kaur, G., 2014. Exploratory structural equation modeling: an integration of the best features of exploratory and confirmatory factor analysis. Annu. Rev. Clin. Psychol. 10, 85–110.
McNeish, D., Stapleton, L.M., 2016. On the unnecessary ubiquity of hierarchical linear modeling. Psychol. Methods, http://dx.doi.org/10.1037/met0000078, Advance online publication.
Medland, S.E., Jahanshad, N., Neale, B.M., Thompson, P.M., 2014. Whole-genome analyses of whole-brain data: working within an expanded search space. Nat. Neurosci. 17, 791–800.
Molenaar, P.C., 2010. Latent variable models are network models. Behav. Brain Sci. 33 (2–3), 166–166.
Monfils, M.H., Cowansage, K.K., Klann, E., LeDoux, J.E., 2009. Extinction-reconsolidation boundaries: key to persistent attenuation of fear memories. Science 324 (5929), 951–955.
Muthén, L.K., Muthén, B.O. (1998–2015). Mplus User's Guide: Statistical Analysis with Latent Variables: User's Guide. Muthén & Muthén.
Neale, M.C., Maes, H.H., 2004. Methodology for Genetic Studies of Twins and Families. Virginia Commonwealth University, Department of Psychiatry, Richmond, VA.
Neuroskeptic (2016). Flexible measures are a problem for science. Discover. Available at http://blogs.discovermagazine.com/neuroskeptic/2016/04/05/flexible-measures-science/#.WG20j7YrK8U (April 5).
Neisser, U., Boodoo, G., Bouchard Jr, T.J., Boykin, A.W., Brody, N., Ceci, S.J., et al., 1996. Intelligence: knowns and unknowns. Am. Psychol. 51 (2), 77–101.
Norbury, A.E., 2015. Brain Mechanisms Underlying Sensation-seeking in Humans. UCL (University College London), Retrieved from http://discovery.ucl.ac.uk/1471571/.
Okbay, A., Baselmans, B.M., De Neve, J.E., Turley, P., Nivard, M.G., Fontana, M.A., et al., 2016a. Genetic variants associated with subjective well-being, depressive symptoms, and neuroticism identified through genome-wide analyses. Nat. Gene. 48, 624–633.
Okbay, A., Beauchamp, J.P., Fontana, M.A., Lee, J.J., Pers, T.H., Rietveld, C.A., et al., 2016b. Genome-wide association study identifies 74 loci associated with educational attainment. Nature 533 (7604), 539–542.
Op de Macks, Z.A., Moor, B.G., Overgaauw, S., Güroğlu, B., Dahl, R.E., Crone, E.A., 2011. Testosterone levels correspond with increased ventral striatum activation in response to monetary rewards in adolescents. Dev. Cogn. Neurosci. 1 (4), 506–516.
Peper, J.S., Koolschijn, P.C.M., Crone, E.A., 2013. Development of risk taking: contributions from adolescent testosterone and the orbito-frontal cortex. J. Cogn. Neurosci. 25 (12), 2141–2150.
Petersen, A.C., Crockett, L., Richards, M., Boxer, A., 1988. A self-report measure of pubertal status: reliability, validity: and initial norms. J. Youth Adolesc. 17, 117–133.
Pfeifer, J.H., Allen, N.B., 2016. The audacity of specificity: moving adolescent developmental neuroscience towards more powerful scientific paradigms and translatable models. Dev. Cogn. Neurosci. 17, 131–137, http://dx.doi.org/10.1016/j.dcn.2015.12.012.
Plomin, R., DeFries, J.C., Knopik, V.S., Neiderheiser, J., 2013. Behavioral Genetics. Palgrave Macmillan.
Quinn, P.D., Harden, K.P., 2013. Differential changes in impulsivity and sensation seeking and the escalation of substance use from adolescence to early adulthood. Dev. Psychopathol. 25 (Special Issue 01), 223–239, http://dx.doi.org/10.1017/S0954579412000284.
Raine, A., Reynolds, C., Venables, P.H., Mednick, S.A., 2002. Stimulation seeking and intelligence: a prospective longitudinal study. J. Pers. Soc. Psychol. 82 (4), 663–674.
Reynolds, B., Ortengren, A., Richards, J.B., de Wit, H., 2006. Dimensions of impulsive behavior: personality and behavioral measures. Personal. Individ. Differ. 40 (2), 305–315, http://dx.doi.org/10.1016/j.paid.2005.03.024.
Rietveld, C.A., Medland, S.E., Derringer, J., Yang, J., Esko, T., Martin, N.W., Albrecht, E., 2013. GWAS of 126,559 individuals identifies genetic variants associated with educational attainment. Science 340 (6139), 1467–1471.
Rietveld, C.A., Esko, T., Davies, G., Pers, T.H., Turley, P., Benyamin, B., et al., 2014. Common genetic variants associated with cognitive performance identified using the proxy-phenotype method. Proc. Natl. Acad. Sci. U. S. A. 111 (38), 13790–13794, http://dx.doi.org/10.1073/pnas.1404623111.
Russo, M.F., Stokes, G.S., Lahey, B.B., Christ, M.A.G., McBurnett, K., Loeber, R., et al., 1993. A sensation seeking scale for children: further refinement and psychometric development. J. Psychopathol. Behav. Assess. 15 (2), 69–86.
Selzam, S., Krapohl, E., von Stumm, S., O'Reilly, P.F., Rimfeld, K., Kovas, Y., Plomin, R., 2016. Predicting educational achievement from DNA. Mol. Psychiatry, Advanced Online Publication.
Shallice, T., 1982. Specific impairments of planning. Philos. Trans R. Soc. Lond. B Biol. Sci. 298 (1089), 199.
Shulman, E.P., Harden, K.P., Chein, J.M., Steinberg, L., 2015a. Sex differences in the developmental trajectories of impulse control and sensation-seeking from early adolescence to early adulthood. J. Youth Adolesc. 44 (1), 1–17, http://dx.doi.org/10.1007/s10964-014-0116-9.
Shulman, E.P., Smith, A.R., Silva, K., Icenogle, G., Duell, N., Chein, J., Steinberg, L., 2015b. The dual systems model: review, reappraisal: and reaffirmation. Dev. Cogn. Neurosci. 17, 103–177.
Shulman, E.P., Harden, K.P., Chein, J.M., Steinberg, L., 2016. The development of impulse control and sensation-seeking in adolescence: independent or interdependent processes? J. Res. Adolesc. 26, 37–44, http://dx.doi.org/10.1111/jora.12181.
Simmons, J.P., Nelson, L.D., Simonsohn, U., 2011. False-positive psychology undisclosed flexibility in data collection and analysis allows presenting anything as significant. Psychol. Sci. 22, 1359–1366.
Skeel, R.L., Neudecker, J., Pilarski, C., Pytlak, K., 2007. The utility of personality variables and behaviorally-based measures in the prediction of risk-taking behavior. Personal. Individ. Differ. 43 (1), 203–214, http://dx.doi.org/10.1016/j.paid.2006.11.025.
Smith, A.R., Chein, J., Steinberg, L., 2013. Impact of socio-emotional context, brain development, and pubertal maturation on adolescent risk-taking. Horm. Behav. 64 (2), 323–332.
Steiger, J.H., 1990. Structural model evaluation and modification: an interval estimation approach. Multivar. Behav. Res. 25 (2), 173–180, http://dx.doi.org/10.1207/s15327906mbr2502_4.
Steinberg, L., Chein, J.M., 2015. Multiple accounts of adolescent impulsivity. Proc. Natl. Acad. Sci. U. S. A. 112 (29), 8807–8808, http://dx.doi.org/10.1073/pnas.1509732112.
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 self-report: evidence for a dual systems model. Dev. Psychol. 44 (6), 1764–1778, http://dx.doi.org/10.1037/a0012955.
Steinberg, L., Graham, S., O'Brien, L., Woolard, J., Cauffman, E., Banich, M., 2009. Age differences in future orientation and delay discounting. Child Dev. 80 (1), 28–44, http://dx.doi.org/10.1111/j.1467-8624.2008.01244.x.
Steinberg, L., 2010. A dual systems model of adolescent risk-taking. Dev. Psychobiol. 52 (3), 216–224, http://dx.doi.org/10.1002/dev.20445.
Tough, P., 2012. How Children Succeed: Grit, Curiosity, and the Hidden Power of Character. Houghton Mifflin Harcourt, Boston.
Tucker-Drob, E.M., Briley, D.A., 2014. Continuity of genetic and environmental influences on cognition across the life span: a meta-analysis of longitudinal twin and adoption studies. Psychol. Bull. 140 (4), 949–979, http://dx.doi.org/10.1037/a0035893.
Tucker-Drob, E.M., Briley, D.A., Harden, K.P., 2013. Genetic and environmental influences on cognition across development and context. Curr. Direct. Psychol. Sci. 22, 349–355.
Tucker-Drob, E.M., Briley, D.A., Engelhardt, L.E., Mann, F.D., 2016. Genetically-mediated associations between measures of childhood character and academic achievement. J. Pers. Soc. Psychol., http://dx.doi.org/10.1037/pspp0000098, Advance online publication.
van Duijvenvoorde, A.C.K., Op de Macks, Z.A., Overgaauw, S., Gunther Moor, B., Dahl, R.E., Crone, E.A., 2014. A cross-sectional and longitudinal analysis of reward-related brain activation: effects of age, pubertal stage, and reward

sensitivity. Brain Cogn. 89, 3–14, http://dx.doi.org/10.1016/j.bandc.2013.10.005.
van Ravenzwaaij, D., Dutilh, G., Wagenmakers, E.J., 2011. Cognitive model decomposition of the BART: assessment and application. J. Math. Psychol. 55 (1), 94–105.
van den Bos, W., Eppinger, B., 2016. Developing developmental cognitive neuroscience: from agenda setting to hypothesis testing. Dev. Cogn. Neurosci. 17, 138–144.
van den Bos, W., Rodriguez, C.A., Schweitzer, J.B., McClure, S.M., 2014. Connectivity strength of dissociable striatal tracts predict individual differences in temporal discounting. J. Neurosci. 34 (31), 10298–10310, http://dx.doi.org/10.1523/JNEUROSCI.4105-13.2014.
van den Bos, W., Rodriguez, C.A., Schweitzer, J.B., McClure, S.M., 2015. Adolescent impatience decreases with increased frontostriatal connectivity. Proc. Natl. Acad. Sci. U. S. A. 112 (29), E3765–E3774, http://dx.doi.org/10.1073/pnas.1423095112.
Wallsten, T.S., Pleskac, T.J., Lejuez, C.W., 2005. Modeling behavior in a clinically diagnostic sequential risk-taking task. Psychol. Rev. 112 (4), 862–880.
Wechsler, D., 1999. Wechsler abbreviated scale of intelligence. Psychological Corporation.
Weigard, A., Chein, J., Albert, D., Smith, A., Steinberg, L., 2014. Effects of anonymous peer observation on adolescents' preference for immediate rewards. Dev. Sci. 17 (1), 71–78, http://dx.doi.org/10.1111/desc.12099.
Whiteside, S.P., Lynam, D.R., 2001. The Five Factor Model and impulsivity: using a structural model of personality to understand impulsivity. Personal. Individ. Differ. 30 (4), 669–689, http://dx.doi.org/10.1016/S0191-8869(00)00064-7.
Zimbardo, P.G., Boyd, J.N., 1999. Putting time in perspective: a valid, reliable: individual-differences metric. J. Pers. Soc. Psychol. 77, 1271–1288.