PNAS RESEARCH ARTICLE | PSYCHOLOGICAL AND COGNITIVE SCIENCES OPEN ACCESS

Impulsivity is a stable, measurable, and predictive psychological trait

Yuqi Huang[a,b], Shenghua Luan[a,b,1], Baizhou Wu[a,b], Yugang Li[a,b], Junhui Wu[a,b], Wenfeng Chen[c], and Ralph Hertwig[d]

Edited by Elke Weber, Princeton University, Princeton, NJ; received December 19, 2023; accepted April 12, 2024

Impulsivity is a personality construct frequently employed to explain and predict important human behaviors. Major inconsistencies in its definition and measurement, however, have led some researchers to call for an outright rejection of impulsivity as a psychological construct. We address this highly unsatisfactory state with a large-scale, preregistered study (N = 1,676) in which each participant completed 48 measures of impulsivity derived from 10 self-report scales and 10 behavioral tasks and reported frequencies of seven impulsivity-related behaviors (e.g., impulsive buying and social media usage); a subsample (N = 196) then completed a retest session 3 mo later. We found that correlations between self-report measures were substantially higher than those between behavioral tasks and between self-report measures and behavioral tasks. Bifactor analysis of these measures exacted one general factor of impulsivity *I*, akin to the general intelligence factor *g*, and six specific factors. Factor *I* was related mainly to self-report measures, had high test–retest reliability, and could predict impulsivity-related behaviors better than existing measures. We further developed a scale named the adjustable impulsivity scale (AIMS) to measure *I*. AIMS possesses excellent psychometric properties that are largely retained in shorter versions and could predict impulsivity-related behaviors equally well as *I*. These findings collectively support impulsivity as a stable, measurable, and predictive trait, indicating that it may be too early to reject it as a valid and useful psychological construct. The bifactorial structure of impulsivity and AIMS, meanwhile, significantly advance the conceptualization and measurement of construct impulsivity.

trait impulsivity | bifactor model | adjustable impulsivity scale | machine learning

Significance

Impulsivity is a personality trait associated with many behaviors in clinical and nonclinical contexts. Serious doubts, however, have been raised on impulsivity as a valid psychological construct, let alone a personality trait. In this large-scale study (N = 1,676), each participant completed 48 measures of impulsivity, and we extracted one general factor of impulsivity *I*, akin to the general intelligence factor *g*, and six specific factors from these measures. Besides being temporally stable, factor *I* could predict self-reported impulsivity-related behaviors (e.g., impulsive buying and social media usage) better than existing measures and be measured with a psychometrically well-performing scale. These findings show that individuals do differ in trait impulsivity, and such differences are stable, measurable, and predictive of real-world behaviors.

Impulsivity is a personality construct considered to affect a wide range of human behaviors. It is associated with behaviors detrimental to the self and others, such as violence, binge eating, and excessive use of social media (1–3), and is a key diagnostic feature for an array of psychological disorders, such as bipolar disorder, antisocial personality disorder, and substance use disorder (4). Impulsivity also plays an important role in decision-making, leading individuals to overlook important information prior to making a decision, take unreasonably high risks, and opt for immediate rather than delayed payoffs in intertemporal choices (5–7).

In the 1930s, J. P. and Ruth Guilford first brought impulsivity to researchers' attention by naming rhathymia as a personality trait, characterizing it as "freedom from care and concern; a lack of serious-mindedness and an impulsiveness" (p. 28) (8). Over the years, impulsivity has evolved from being treated as a component of a major personality trait to an independently studied trait of its own. The study of impulsivity has focused on three main areas: 1) identifying the underlying conceptual structure of impulsivity based on established theories of personality (9, 10) and psychometric modeling of empirical data (11, 12); 2) developing proper measurements of impulsivity that consist of both self-report scales, such as the Barratt Impulsiveness Scale (13) and Eysenck's I-7 scale (9), and behavioral tasks that are supposed to reflect facets of impulsivity, such as (in)ability to inhibit responses in the stop signal task (14) and risk-taking propensity in the balloon analogue risk task (15); and 3) accessing the associations of impulsivity with various behaviors in both the clinical and the nonclinical contexts to gauge the predictive and diagnostic usefulness of impulsivity (16, 17). Yet, despite this long history of research and the frequent use of impulsivity to explain and predict behaviors, there are serious challenges to its validity as a stable, measurable, and predictive psychological construct.

The greatest challenge is the lack of a clear definition of construct impulsivity. Impulsivity is now commonly viewed as a multidimensional construct that comprises distinct factors. What these factors are, however, has been hotly debated. Barratt and colleagues suggested that there are three main factors of impulsivity: motor impulsiveness, nonplanning impulsiveness, and attentional impulsiveness (13, 18). Others have proposed two-factor structures that differ in what they thought the two factors ought to be; examples include

Author contributions: Y.H., S.L., J.W., W.C., and R.H. designed research; Y.H., B.W., and Y.L. performed research; Y.H. and W.C. contributed new reagents/analytic tools; Y.H., B.W., and Y.L. analyzed data; and Y.H., S.L., and R.H. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

[1]To whom correspondence may be addressed. Email: luansh@psych.ac.cn.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2321758121/-/DCSupplemental.

Published June 3, 2024.

venturesomeness and Impulsiveness (9), functional and dysfunctional impulsivity (19), and behavioral activation and inhibition systems (20). In an attempt to synthesize the existing measures of impulsivity, Whiteside and Lynam (12) created a four-factor solution, which was later extended to five (21): negative urgency, positive urgency, (lack of) premeditation, (lack of) perseverance, and sensation seeking. So far, there is no consensus among researchers on which of these dimensional conceptualizations and definitions is best, in theory or in practice.

Impulsivity is often measured by scales either specifically designed for it or by a relevant subscale in a general personality inventory. Alternatively, researchers who think that "talk is cheap" and that preferences and propensities are best revealed through behavior (22) have attempted to measure impulsivity using behavioral tasks. These tasks tap into psychological processes that are supposed to reflect different aspects of impulsivity, such as response inhibition, delay discounting, risk taking, and (in)attention. Behavioral measures are particularly useful when the target sample has limited or no abilities of self-reflection or language comprehension (e.g., children and nonhuman animals). Their use, however, further complicates the definition of impulsivity, because they add even more factors to an already crowded field, and it is unclear how to integrate them with factors identified in the scales under the same framework (17, 23, 24).

The lack of a clear definition and the existence of a large number of impulsivity measures have created chaos in impulsivity measurement. Unable to agree on which mode of measurement (i.e., self-report scale or behavioral task) and which specific measure in each mode would be best to use, researchers and practitioners have applied a diverse set of measures while rarely providing justification for why one measure was selected over others. This state of affairs makes it difficult to compare and integrate findings of disparate studies, putting a significant roadblock on the progress in impulsivity research.

Another challenge in impulsivity research concerns the temporal stability of impulsivity. Temporal stability is critical for conceptualizing constructs as either traits or states. A trait is a personality characteristic that remains relatively stable across time and situations, while a state refers to a transient emotional or behavioral reaction to a specific situation or event, and is thus relatively unstable. Despite the vast amount of research on impulsivity, not much is known about its temporal stability. Most studies examining the issue focus on one or a few impulsivity measures, keep the time interval between test and retest relatively short (e.g., under 2 wk), or rely on small samples (25–27). These studies therefore offer only limited evidence on the temporal stability or instability of impulsivity. Nonetheless, a general finding in this research is that behavioral measures tend to be less stable than self-report scales—a finding that resonates with those in other personality domains (22, 28, 29).

A personality trait should be predictive of relevant future behaviors. Findings on the predictiveness of impulsivity, however, have been inconsistent. For example, Sharma and colleagues (17) conducted a meta-analysis of the relationships between impulsivity scales, behavioral measures of impulsivity, and some impulsivity-related behaviors (e.g., aggression, substance use, and pathological gambling). They found that the strength of the relationship depends on the impulsivity measure used and the behavior evaluated, ranging from 0 (between the Iowa gambling task and pathological gambling) to .66 (between Eysenck's I-7 scale and delinquency behavior). In a more recent study, Creswell et al. (16) administered a battery of impulsivity scales and four behavioral tasks on a large sample of individuals and examined how these measures were related to substance use and aggression. They found that whereas statistically extracted factors from scales could generally predict behaviors well, the predictive power of the behavioral tasks was almost zero.

In sum, impulsivity research suffers from four major problems: a lack of agreement on its definition, unprincipled use of diverse measurement tools, inconclusive findings regarding its temporal stability, and inconsistent results concerning its predictiveness. On top of that, a recent review by Strickland and Johnson also highlighted problems in the applications of impulsivity in neuroscience and clinical research, and how the everyday usage of the word impulsivity has confused and hampered its scientific investigation (30). In the face of these problems, the authors proclaimed that "impulsivity fails to satisfy even the basic requirements of a psychological construct and should be rejected as such" (p. 337).

We sympathize with the authors' frustration, but think that it may be too early to reject impulsivity as a valid and useful psychological construct. There have been similarly discouraging situations in the studies of other psychological constructs, such as intelligence, risk preference, and psychopathology (31–33). One common means to resolving the issue is to administer a large battery of measures on a large number of people in order to gather better and more comprehensive evidence for or against a construct's validity and to further develop a better measurement of the construct. Despite the ongoing controversy, no such studies have been conducted in research of impulsivity. The present study is designed to fill this gap.

Measuring a large sample of participants (N = 1,676) with 10 self-report scales and 10 behavioral tasks of impulsivity, our study focuses on addressing four main issues in impulsivity research. First, we explore how well a bifactorial structure of impulsivity—that is, one general factor I, akin to the general intelligence factor g, plus some specific factors—can explain measurement data. This conceptualization of impulsivity departs from the predominant view of impulsivity as a congregate of distinct factors and has not been seriously investigated in previous research. A good fit of the bifactor model and the presence of a general factor would support impulsivity as a psychological construct, though with an unconventional structure.

Second, we determine whether impulsivity meets the temporal stability requirement of a psychological trait, calculating the test–retest reliabilities of the administered measures and the extracted factors in a second round of measurement completed by a subsample of participants. Third, we assess the predictiveness of the impulsivity measures and the extracted factors for seven self-reported impulsivity-related behaviors, such as impulsive eating, social media usage, and Internet gaming, using machine-learning algorithms. Last, we come up with a measure of impulsivity in light of our analyses and evaluate the measure's psychometric properties and its predictiveness for impulsivity-related behaviors.

The results of our study can provide valuable theoretical insights and practical guidance to impulsivity researchers and practitioners. To paraphrase Mark Twain, reports of impulsivity's death may have been exaggerated, and the construct may still be salvaged.

Results

The entire study was conducted online. Participants completed 10 self-report scales and 10 behavioral tasks of impulsivity within two weeks at their own pace. The scales and tasks produced a total of 48 impulsivity measures, including subscales and dependent variables (Table 1). Participants also reported how frequently they had engaged in seven impulsivity-related behaviors in the previous 3 mo. A valid sample of 1,676 participants completed all measures, and a subsample of 196 participants completed a retest session

Table 1. Measures of impulsivity and impulsivity-related behaviors

Measure	Subscale/Dependent variable	Abbreviation
Self-report scale		
Barratt Impulsiveness Scale-11	Attentional impulsiveness	BIS11a
	Motor impulsiveness	BIS11m
	Nonplanning impulsiveness	BIS11n
Behavioral inhibition system and behavioral activation system scales (BIS/BAS)	BIS	BIS
	BAS-reward responsiveness	BASr
	BAS-drive	BASd
	BAS-fun seeking	BASf
Dickman Impulsivity Inventory	Functional impulsivity	DIIf
	Dysfunctional impulsivity	DIId
Eysenck's I-7	Impulsiveness	I7i
	Venturesomeness	I7v
Impulsive Sensation Seeking Scale	Sensation seeking	IMPSSs
	Impulsiveness	IMPSSi
Sensitivity to Punishment and Sensitivity to Reward Questionnaire	Sensitivity to punishment	SPSRQsp
	Sensitivity to reward	SPSRQsr
UPPS-P Impulsive Behavior Scale	Premeditation	UPPSPpr
	Negative urgency	UPPSPnu
	Sensation seeking	UPPSPss
	Perseverance	UPPSPpe
	Positive urgency	UPPSPpu
Multidimensional Personality Questionnaire—Control	Deliberation	MPQCd
	Planning	MPQCp
	Remaining items	MPQCr
NEO PI-R	Impulsiveness	NEOi
	Self-discipline	NEOsd
	Deliberation	NEOd
	Excitement seeking	NEOes
Tridimensional Personality Questionnaire	Novelty seeking	TPQ
Behavioral task		
Balloon Analogue Risk Task	Number of pumps	BART
Decision from experience	Number of samplings	DFE
Delay discounting	Discounting rate of 50 RMB	DD1
	Discounting rate of 500 RMB	DD2
	Discounting rate of 5,000 RMB	DD3
Go/no-go	Commission error rate	GNG
Passive avoidance with loss of reward	Commission error rate	PALR
Stop signal task	Stop signal reaction time	SST
Information Sampling Task with fixed and decreased rewards (ISFR and ISDR)	Decision accuracy in ISFR	ISFRda
	Number of samples in ISFR	ISFRns
	Decision accuracy in ISDR	ISDRda
	Number of samples in ISDR	ISDRns
Time estimation (TIME)	Estimation bias of 5 s	TIME5
	Estimation bias of 10 s	TIME10
	Estimation bias of 30 s	TIME30
	Estimation bias of 60 s	TIME60
Immediate and Delayed Memory Task (IMT and DMT)	Ratio of commission error rate to correct detection rate in IMT	IMT
	Ratio of commission error rate to correct detection rate in DMT	DMT
Synthetic face identification task	Response time for top face	SFITtf
	Response time for bottom face	SFITbf

Table 1. (Continued)

Measure	Subscale/Dependent variable	Abbreviation
Frequency of impulsivity-related behavior		
Alcohol Use Disorders Identification Test	Total score	Drinking
Fagerström test for nicotine dependence	Total score	Smoking
Buying impulsiveness scale	Total score	Buying
Three-Factor Eating Questionnaire	Total score	Eating
Nine-item Internet Gaming Disorder Scale	Total score	Gaming
Short video app addiction test, adopted from the Internet Addiction Test	Total score	Short Video
Social media addiction scale, adopted from the Bergen Facebook Addiction Scale	Total score	Social Media

3 mo later in the same manner as in the first measurement round (see details in *Methods*).

Correlation Analysis: Convergent Validity

We calculated pairwise correlations between all impulsivity measures after controlling for age, sex, education, and occupation, and generated a network plot to visualize the results (Fig. 1). The plot shows that self-report measures were clustered together, indicating a high level of convergence among them. That said, some self-report measures, such as UPPSPss, I7v, NEOes, and BIS, were less correlated with the others. These measures were developed primarily to measure sensation seeking or sensitivity to punishment and reward.

Measures from behavioral tasks were only weakly correlated with each other, suggesting that these tasks may measure distinct states, constructs, or aspects of impulsivity. In contrast, the measures from the same task (e.g., delay discounting) were highly

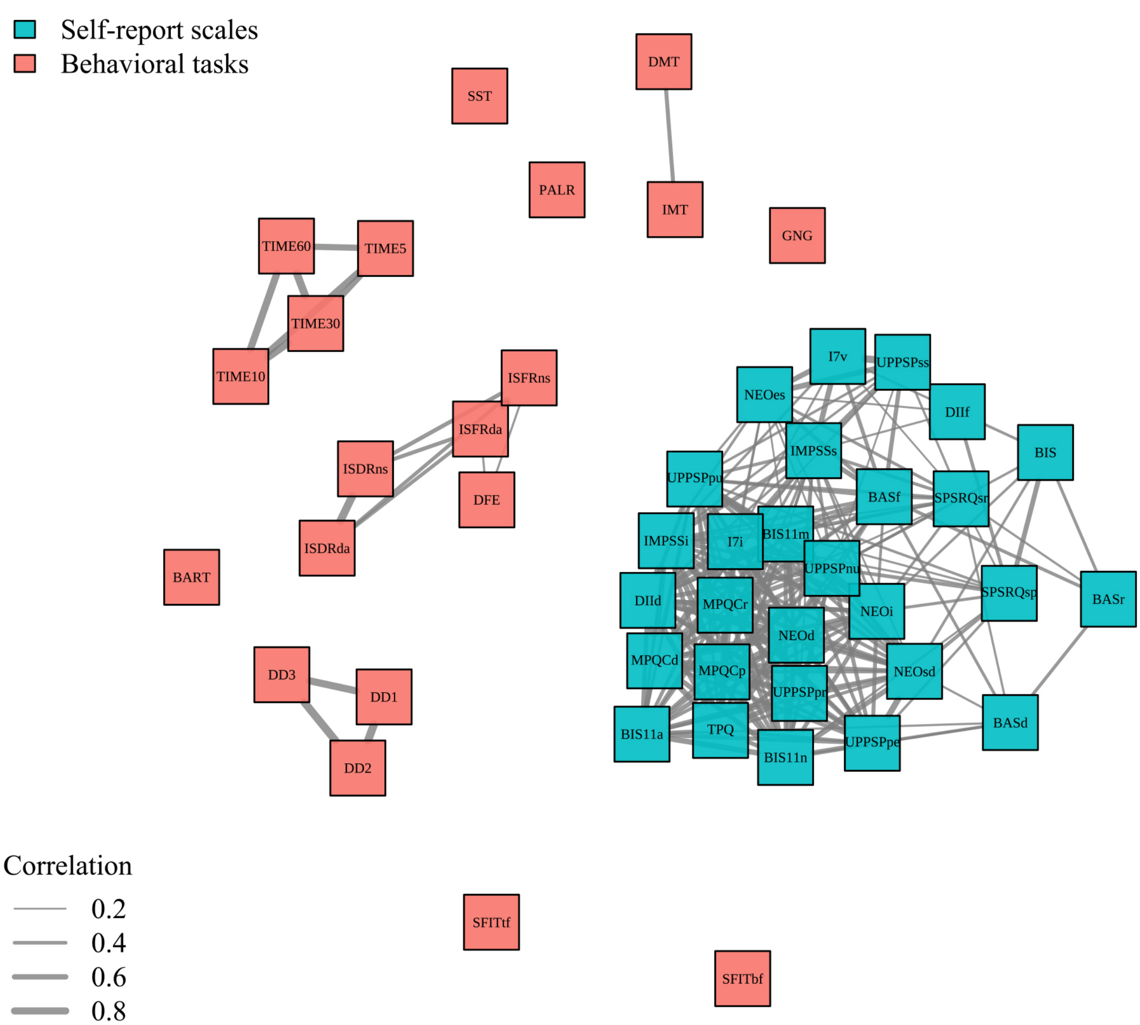

Fig. 1. Pairwise correlations between impulsivity measures. Each node represents a measure of impulsivity. Only absolute correlations larger than 0.20 are shown. The closer the distance and the thicker the edge between two nodes, the higher the correlation.

correlated with each other and demonstrated high convergence. Moreover, the correlations between self-report measures and the measures derived from behavioral tasks were weak, and many were close to zero (see specific values in *SI Appendix*, Table S4).

Overall, the correlation analysis shows that behavioral measures of impulsivity diverge not only from the self-report measures but also largely from each other. These results align with findings of prior research and suggest that the lack of convergent validity in the behavioral measures may be a crucial factor in the unsatisfactory state of impulsivity research (17, 30).

Psychometric Modeling

We conducted an exploratory factor analysis with bifactor rotation to obtain a bifactor model of impulsivity. In a bifactor model, the general factor directly accounts for the shared variance across all included measures, while the residual variance is captured by orthogonal specific factors (34, 35). Previous studies largely discarded the view that impulsivity is a unitary construct, instead assuming impulsivity to be a multidimensional construct with distinct factors and analyzing the data accordingly (17, 36, 37). With a large battery of measures administered on a large sample, we took an alternative view on construct impulsivity and examined how well, if at all, this general-plus-specific bifactorial structure could explain the data.

Fig. 2 displays the results of the bifactor model with all 48 measures. It shows a general factor *I* and six specific factors. Similar to the *g* factor in intelligence, factor *I* represents the common construct underlying the impulsivity measures and can be construed as the construct reflecting individual differences in trait impulsivity; a specific factor, meanwhile, captures unique variances shared by only some of the measures and reflects individual differences in a specific domain or aspect of impulsivity (38).

In general, factor *I* could account for a substantial portion of the variance in the self-report measures, but for little to no variance in the behavioral measures. Among the six specific factors, three corresponded to three behavioral tasks: F1 corresponded to time estimation, F3 to delay discounting, and F4 to information sampling. The other three were extracted from self-report measures: F2 pertained mostly to measures supposed to capture sensation seeking (e.g., UPPSPss and I7v), F5 to sensitivity to punishment (e.g., SPSRQsp and BIS), and F6 to sensitivity to reward (e.g., SPSRQsr and BAS). These three concepts were first proposed independently from impulsivity (10, 39), but were gradually incorporated as factors in some multidimensional framework of impulsivity (12). Our bifactor model suggests a different view on these concepts: each may be treated as a measure of impulsivity in a specific domain, akin to the specific factors, such as mathematics and memory, in Spearman's two-factor framework of intelligence (32).

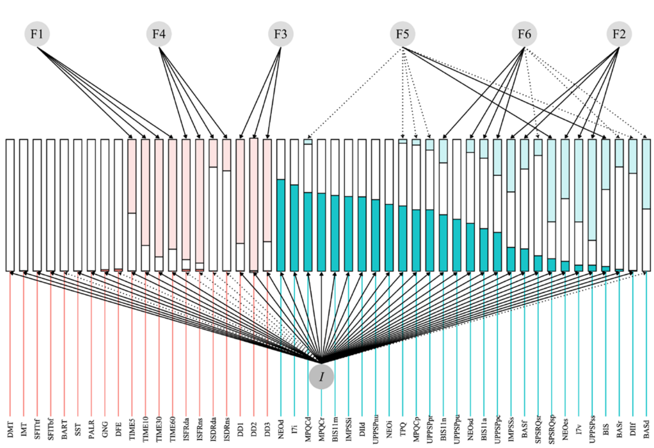

Fig. 2. Results of a bifactor model with all 48 impulsivity measures. Self-report measures are shown in blue, and behavioral measures are shown in red. *I* represents the extracted general factor; F1 to F6 are the specific factors. The light-colored portion in the upper part of each bar represents the variance that can be explained by the specific factors; the dark-colored portion in the lower part represents the variance that can be explained by the general factor *I*; the white portion represents the unexplained variance. Negative loadings are indicated by dashed lines.

The bifactor model explained 53% of the total variance, and factor *I* explained 20% of the total variance. In other words, factor *I* accounted for 38% (i.e., 20/53) of the explained variance by the bifactor model. A confirmatory factor analysis (CFA) indicated a satisfactory fit of the bifactor model: standardized RMS residual (SRMR) = 0.06, RMSE of approximation (RMSEA) = 0.06, comparative fit index (CFI) = 0.93, and Tucker–Lewis index (TLI) = 0.93. Given that the bifactor model was extracted from the entire sample (N = 1,696), we additionally conducted a CFA on the retest sample (N = 196) to avoid overfitting, and it also showed a good fit: SRMR = 0.08, RMSEA = 0.04, CFI = 0.97, and TLI = 0.97. Moreover, we compared the bifactor model with three other models: unidimensional, nonhierarchical multidimensional, and second-order. *SI Appendix*, Table S7 summarizes the results, which show consistently better fits of the bifactor model to our data than those of other models.

The 20% of total variance explained by *I* is not high but also not particularly low. In comparison, an extracted general factor was found to explain 18% of the total variance in risk preference (31), 35% in intelligence (32), and 41% in psychopathology (33). One possible reason for this result is that the bifactor model includes too many unrelated measures, primarily the behavioral ones. To address this, we conducted a bifactor analysis with only the 28 self-report measures. Based on this more closely related set of measures, we obtained a model (*SI Appendix*, Fig. S2) that explains 62% of the total variance, with factor *I* explaining 34% of the total variance. The CFA showed a good fit: SRMR = 0.09, RMSEA = 0.09, CFI = 0.93, and TLI = 0.92.

The results of psychometric modeling are consistent with those of the correlation analysis in that both indicate the problematic role of behavioral tasks in defining and measuring the impulsivity construct: most of them appear to measure different things from the self-report scales, are hardly connected with each other, and do not load on a general impulsivity factor. As a result, the fit of the bifactor model improved substantially once the behavioral measures were removed. That said, behavioral measures do make up several specific factors in the broad bifactorial structure of impulsivity, and some of them, including those from go/no-go, decision from experience, and information sampling, correlate with factor *I* to some degree.

Temporal Stability

Fig. 3 shows the test–retest reliabilities of all 48 impulsivity measures and the extracted factors with these measures, based on a retest sample (N = 196) that went through the second round of measurement 3 mo after the first. In general, self-report measures exhibited greater temporal stability (M = 0.66) than behavioral measures (M = 0.44), consistent with findings of previous research on impulsivity (40). The temporal stability of the extracted factors (M = 0.64) was on par with that of self-report measures, and the factors that were mainly related to self-report measures (i.e., *I*, F2, F5, and F6) were generally more temporally stable than the other factors. Notably, the general impulsivity

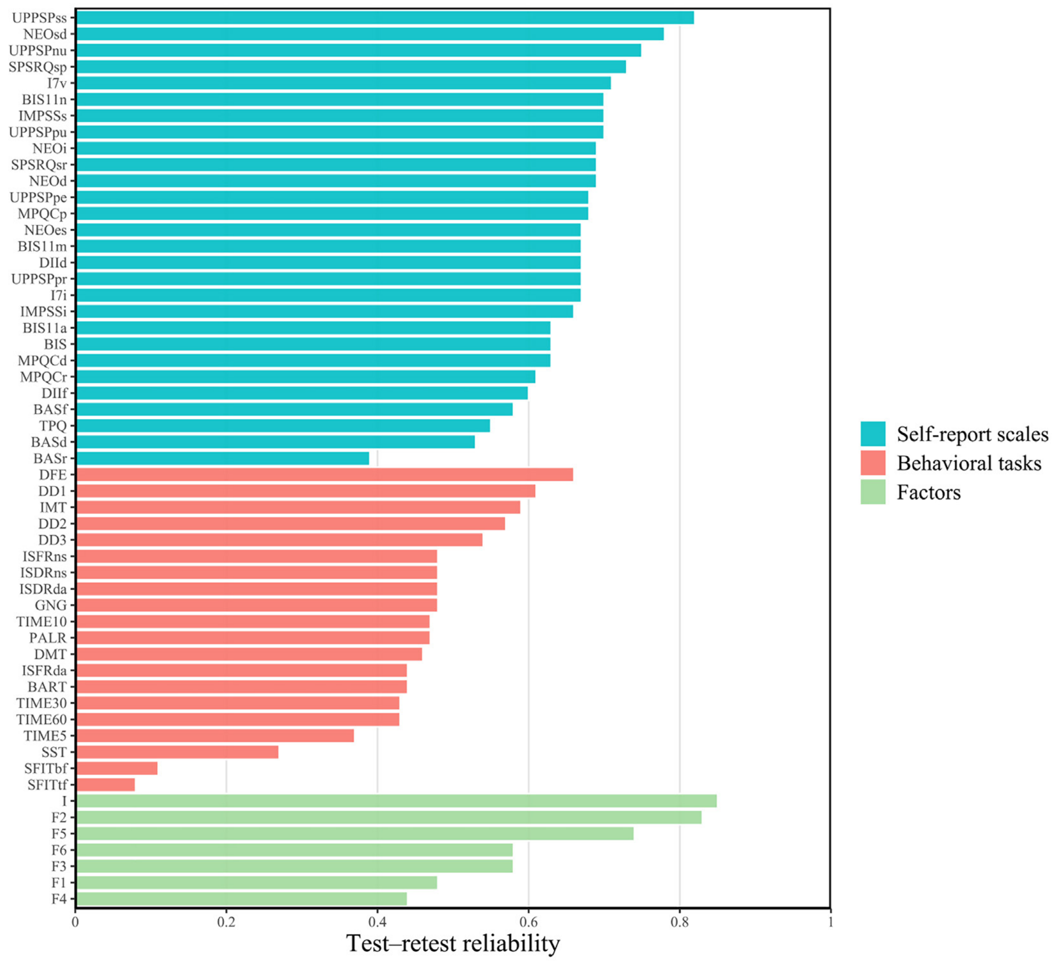

Fig. 3. Test–retest reliabilities of 48 impulsivity measures and the extracted factors with these measures. Test–retest reliability was measured by Spearman correlation.

factor *I* had the highest test–retest reliability of all factors and measures at 0.85.

One possible reason for the generally low test–retest reliability of behavioral measures is that these measures may be more susceptible to contextual and situational influences (31); thus, they may assess mainly momentary impulsive behaviors at the time of testing rather than prototypical behaviors over a long period of time (41). In addition, some cognitive tasks, such as go/no-go and stop signal, were explicitly designed to maximize between-condition variation at the cost of reduced levels of between-person variation (42, 43); this property can also attenuate measurement reliability.

Predicting Impulsivity-Related Behaviors

To examine the predictiveness of impulsivity measures on impulsivity-related behaviors, we first calculated the correlations between each of the measures, as well as the extracted factors, and the seven self-reported impulsivity-related behaviors ("Frequency" in Table 1). The results show that factor *I* was similarly or more highly correlated with each of the behaviors—with the exceptions of drinking and smoking—than the existing measures and other extracted factors. In addition, it had the highest correlation averaged over the seven behaviors (*SI Appendix*, Table S9). Although correlation is an indicator of the relationship between two variables, it is less useful in predicting behaviors (e.g., who will be a more impulsive buyer?). To address this issue, we applied seven machine-learning algorithms to predict participants who reported a relatively high level of frequency (i.e., 10%) in an impulsivity-related behavior, based on four different sets of predictors (*Methods*). For each behavior and each predictor set, the algorithm with the highest predictive performance was selected as the best prediction model.

Fig. 4A shows the performance of the best prediction model for each of the seven impulsivity-related behaviors with each predictor set. By comparing performances using different sets of predictors, we could gauge the predictive powers of factor *I*, the specific factors, and the behavioral measures. Specifically, the performance difference between using predictor set 1 (demographics only) and set 2 (demographics plus factor *I*) indicates the predictive power of factor *I*; the difference between set 2 and set 3 indicates the additional predictive power of the specific factors (i.e., set 3 includes set 2 predictors plus specific factors); and the difference between set 2 and set 4 (i.e., set 4 includes set 2 predictors plus behavioral measures) indicates the additional predictive power of the behavioral measures.

Using a 0.10 increment in d' as the criterion to judge whether an added predictor had good or limited predictive power, we found that factor *I* was good at predicting impulsive buying, impulsive eating, short video app usage, and social media usage, but less adept at predicting drinking, smoking, and internet gaming. Factor *I*'s limited predictiveness of the latter three behaviors might be caused by a ceiling effect—namely, that predicting them based solely on demographic variables already worked well. Adding the specific factors and behavioral measures to the predictor set did not improve performance for most behaviors: specific factors only had some predictive powers of social media usage and internet gaming, while behavioral measures only had a small predictive power of social media usage.

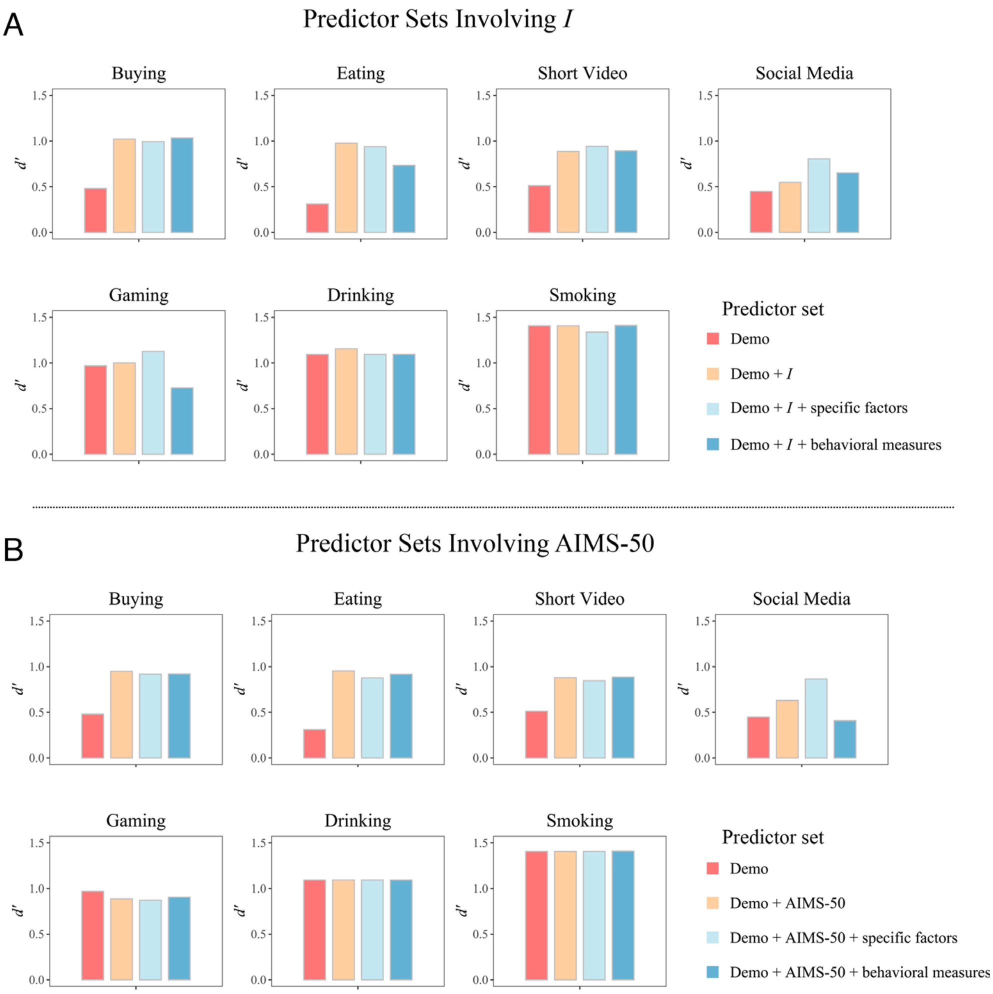

Fig. 4. Performances of the best machine-learning models in predicting impulsivity-related behaviors. (*A*) Results with the extracted general factor *I* in the predictor sets. (*B*) Results with the adjustable impulsivity scale with 50 items (AIMS-50) in the predictor sets. In predicting each behavior, there were four sets of predictors: set 1 included only demographic information (Demo); set 2 included demographic information plus factor *I* (or AIMS-50); set 3 included set 2 predictors and the six extracted specific factors; and set 4 included set 2 predictors and the behavioral measures. Performance was evaluated using *d'*, with a higher *d'* indicating better performance and a *d'* of 1.0 roughly equal to an overall accuracy of 70%.

We additionally applied the best models trained on the first measurement data to predict impulsivity-related behaviors reported at the retest round (3 mo later). *SI Appendix*, Fig. S5*A* illustrates the results of these cross-time predictions, which show that factor *I* was generally predictive in this more difficult prediction task, and its predictive power did not decrease much in comparison to that for the same-time predictions. These results suggest that factor *I* could be a useful prognostic indicator for impulsivity-related behaviors.

In sum, the correlation and machine-learning prediction analyses demonstrate that factor *I* not only was generally more highly correlated with self-reported impulsivity-related behaviors than existing measures but also could improve predictions for a majority of these behaviors. The behavioral measures, on the other hand, had only limited predictive powers, consistent with findings from other studies (16).

The Adjustable Impulsivity Scale (AIMS)

Having identified a general impulsivity factor *I* that is both temporally stable and generally predictive of related behaviors, we next explored ways to measure it. Because self-report measures were overall more stable and had much higher loadings on factor *I* than behavioral measures, it is more appropriate to use a scale than a behavioral task to measure *I*. To develop a scale that can measure *I* better than extant tools, we employed a data-driven approach (*Methods*). In a nutshell, participants were first randomly divided into a training group and a testing group; we then drew random samples of 50 items from the unique items in the 10 scales included in the present study and tested two aspects of a scale composed of these items: the correlation of its total score with *I* and the Cronbach's alpha (internal consistency) of the scale, in both the training and the testing participant groups. Samples with the correlation exceeding 0.90 and Cronbach's alpha exceeding 0.80 in both groups were selected, and an item's frequency of appearance in these quality samples was recorded. The 50 items that appeared most after a redundancy check were selected to form the scale measuring *I* (see the items in *SI Appendix*, Tables S10 and S11).

The full scale consists of all 50 items. However, because the items were ranked according to their frequencies in the quality samples, the length of the scale can be adjusted to include only the top-*m* items. Fig. 5 shows values of three key psychometric properties—correlation with *I*, Cronbach's alpha, and test–retest reliability—of the scale with length ranging from one item to 50. Shorter scales can be practically more useful when there is limited time for testing. We therefore named the scale the adjustable impulsivity scale (AIMS).

With more items, AIMS tends to have better psychometric properties. However, the property values are already quite good with 25 items in the scale: Both the correlation with factor *I* and Cronbach's alpha were around 0.90, and test–retest reliability was above 0.80. Even with only 10 items, the correlation with factor *I* and Cronbach's alpha were above 0.80 and test–retest reliability above 0.75. With all 50 items (i.e., AIMS-50), correlation with

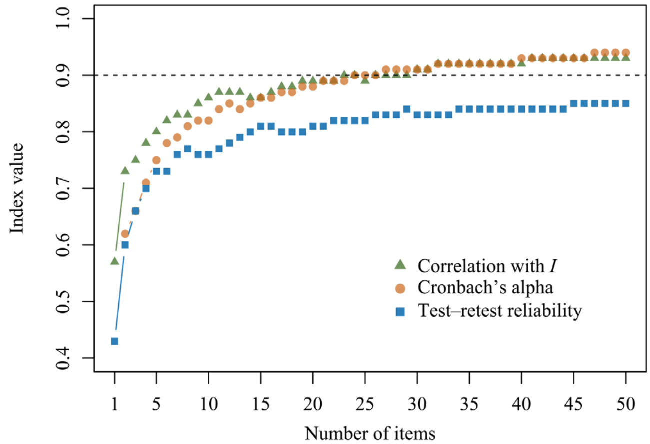

Fig. 5. Psychometric properties of the AIMS with different lengths. In AIMS that includes *m* items (i.e., AIMS-*m*), the items were selected according to their rankings produced by an item sampling and testing approach.

factor *I*, Cronbach's alpha, and test–retest reliability were 0.93, 0.94, and 0.85, respectively. We also ran an EFA with AIMS-50 to examine its factor structure. It suggested a one-factor solution, and a subsequent CFA showed a good fit of this model: SRMR = 0.06, RMSEA = 0.05, CFI = 0.96, and TLI = 0.96.

We next examined the predictive power of AIMS-50 on the self-reported impulsivity-related behaviors by running the same analyses that examined the predictive power of factor *I* but replacing factor *I* with score of AIMS-50. On average, AIMS-50 had a slightly higher correlation with impulsivity-related behaviors than did factor *I* (*SI Appendix*, Table S9). Fig. 4*B* shows performances of the best machine-learning models with AIMS-50 in the predictor sets. The results are similar to those obtained using factor *I*, in that AIMS-50 also had good predictive powers for impulsive buying, impulsive eating, short video app usage, and social media usage, but not for smoking, drinking, and internet gaming. Moreover, adding specific factors and behavioral measures to the predictor set did not improve predictions for most behaviors: specific factors were only predictive of social media usage, and behavioral measures were not predictive of any of the seven behaviors. Finally, *SI Appendix*, Fig. S5*B* illustrates the cross-time predictive performances of AIMS-50 and shows that AIMS-50 was generally as predictive as factor *I* and could predict some behaviors, such as impulsive buying and short video app usage, even better than factor *I*.

The most prominent characteristic of AIMS is that even at a reduced length, it does not suffer much in its psychometric quality. The full-length AIMS (i.e., AIMS-50) has excellent psychometric properties and similar predictive powers to factor *I* for impulsivity-related behaviors. Furthermore, to make AIMS more practically useful, we unified the response format of its items, which were selected from existing scales with different response formats. We tested this reformatted AIMS (*SI Appendix*, Table S11) in an independent sample of participants (N = 236). It had high internal consistency (Cronbach's alpha = 0.96), and a CFA showed good fit of a single-factor model: SRMR = 0.07, RMSEA = 0.03, CFI = 0.99, and TLI = 0.99. Details of this additional validation study can be found in *SI Appendix*.

Discussion

Our study was designed to address some major conceptual and measurement issues that have plagued research on impulsivity and challenged the validity of impulsivity as a psychological construct. Modeling impulsivity as a bifactorial construct, we found that a general impulsivity factor *I* could be extracted from a large battery of measures administered on a large sample of participants. Measures derived from self-report scales loaded much higher on factor *I* than did those based on behavioral tasks, suggesting a substantial gap between stated and behavioral measures of impulsivity. We also found that factor *I* was temporally stable, having a high test–retest reliability of 0.85 with a 3-mo interval, and predictive of a majority of seven self-reported impulsivity-related behaviors. Furthermore, we developed a scale, AIMS, to measure the general impulsivity factor *I*. AIMS has excellent psychometric properties, is similarly predictive of impulsivity-related behaviors as factor *I*, and, importantly, its length can be tailored to research or practical needs. Taken together, these results support impulsivity as a stable, measurable, and predictive psychological trait.

We acknowledge that our findings and conclusions are at odds with a recent call to reject impulsivity as a psychological construct (30). Several important reasons motivated this call, of which the most important is perhaps the lack of a unitary construct underlying a variety of impulsivity measures. Most studies examining construct impulsivity have been dedicated to demonstrating impulsivity as a multidimensional rather than unidimensional construct, without considering the possibility of the coexistence of a general factor and a few specific impulsivity factors. These studies either applied exploratory factor analysis or principal component analysis to determine the multidimensional structure of impulsivity (24, 36, 44) or employed CFA to show that the fit of a multidimensional model was superior to a unidimensional one (11, 37). There are also a few studies attempting to extract a common factor from different impulsivity measures; however, they predominantly use higher-order models, which first extract first-order factors and then higher-order factors from the first-order factors (45). This method cannot directly extract common components at the level of measures and thus tends to overlook the possible associations among them. The bifactor model we tested avoids such problems and is better suited to explore the possible presence of a general factor.

A bifactor model encompasses both the broad and the narrow concepts of an underlying construct by establishing a general factor and several specific factors (46). In our case, the relatively large variance explained by factor *I* supports the existence of construct impulsivity in a broad sense, and the six specific factors highlight domains of impulsivity that can facilitate a more fine-grained understanding of individual differences in impulsivity (e.g., between two similarly impulsive persons, one may be more sensation seeking, while the other more sensitive to reward), and may be useful for predicting particular behaviors (47). For instance, we found that on top of factor *I*, F2 added predictive power for drinking, and F5 for impulsive eating, short video app usage, and social media usage. Furthermore, within the bifactor model, we can examine to what extents the factors in a previous multidimensional framework of impulsivity tap on the general factor *I* and to what extents on a specific factor. For example, all three factors in BIS-11 and four of the five factors in UPPS-P had high loadings on factor *I*, suggesting that their variance was primarily attributable to the general factor, while some factors of UPPS-P (e.g., negative urgency) were also related to the specific factors. Viewing these factors through the lens of our bifactor model, therefore, provides a unique way to connect and integrate previous findings in impulsivity research.

Besides exploring the bifactorial structure of impulsivity, we tested the largest collection of measures in impulsivity research so far. The large number of measures, paradoxically, is the likely reason why the amount of total variance explained by the exacted general factor *I* was not particularly high (i.e., 20%). An inspection of the extracted factors shows that measures based on behavioral tasks loaded little on factor *I*, which captures mainly variances in the self-report measures. There has been ample evidence showing a divide between self-report and behavioral measures of impulsivity (11, 17). Our results confirm this divide: behavioral measures not only converged little with self-report measures but also were quite divergent among themselves; in addition, in comparison to self-report measures, behavioral measures were both temporally less stable and less predictive of impulsivity-related behaviors. A bifactor model with only the self-report measures shows a substantial improvement of total variance explained, jumping from 20% to 34%. In light of these results, are behavioral measures still of value in impulsivity measurement?

We think so. First, there are some behavioral measures that were related to factor *I*, albeit only weakly: the measures from information sampling, decision from experience, and go/no-go (Fig. 2). In situations where it is not possible or feasible to apply self-report scales or where the measured population has incentives to not report truthfully (e.g., prisoners and drug addicts), these tasks may be used as alternative—albeit not the best—measures of factor *I*. Second, having realized the problems with behavioral tasks that are mostly administered in laboratory settings, some researchers have tried to redesign traditional tasks with game-like interfaces that are more attractive to participants and more externally valid. These tasks performed better than traditional tasks, in terms of both key psychometric properties and predictiveness of impulsivity-related behaviors (48). Such work is promising and may be able to alleviate flaws of behavioral tasks and increase their utility in impulsivity research. Third, besides the general factor *I*, our bifactor analysis also returned six specific factors, three of which were based on behavioral tasks (i.e., information sampling, delayed discounting, and time estimation). This suggests that even if some behavioral measures do not directly tap into general impulsivity, they still capture distinct components of impulsivity and can therefore be valid tools for measuring and differentiating people in these domains.

The other three specific factors were related to self-report measures, and one of them is sensation seeking. Zuckerman et al. (39) developed the first scale of sensation seeking, which they conceptualized as the "optimal stimulation level" experienced by an individual and did not make any association with impulsivity. Early follow-up studies also treated and measured the two concepts separately—for example, Eysenck and Eysenck (9) distinguished impulsivity from venturesomeness, a concept similar to sensation seeking. In subsequent research, however, sensation seeking became more closely associated with impulsivity and was eventually integrated as one of the five factors in the widely used UPPS-P impulsive behavior scale (12, 21). This development trajectory is representative of how various concepts have gradually become parts of construct impulsivity, turning it into a hodgepodge of incompatible components. To overcome this state of affairs, we attempted to psychometrically consolidate the existing measures and test for the possibility of a more unified construct of impulsivity.

Our approach was facilitated by two recent methodological innovations in behavioral research. The first is internet-based data collection, which made it easier for people to participate in a study (our study was conducted during the COVID-19 pandemic). Although steps must be taken to control data quality (see ours in *Methods*), carefully designed internet-based studies are generally beneficial to behavioral research (48–50). The second is the use of machine learning for psychological and behavioral research. With the availability of large datasets and the development of AI, machine-learning algorithms have been increasingly applied to help researchers find patterns, test hypotheses, and even build theories (51). We used machine-learning algorithms to examine the predictiveness of factor *I* and of AIMS for impulsivity-related behaviors both statically (i.e., within the same time frame) and dynamically (i.e., across time). The results show that the predictive powers of factor *I* and AIMS were similar and similarly long-lasting, indicating their diagnostic potential.

The development of AIMS was also inspired by the data-driven approach underlying machine learning. The classic approach to developing a scale starts with judging the face validity of possible items. We bypassed this step because all items were already included in established scales and presumedly enjoy high face validity. The next step is item selection, which is usually done by running statistical analyses, such as exploratory factor analysis and principal component analysis, on data collected from one or a few samples. The goal is to form a scale that has good psychometric properties, such as high internal consistency and good fit of the underlying model. We adopted a different approach: drawing a huge number of item samples (i.e., 10 million) and forming a scale with items that have potentially the highest positive impacts on the scale's psychometric properties. Although such an approach has rarely been applied in scale development, the outcome of our attempt was promising. An additional benefit of this approach is that it allows researchers to form a scale with adjustable length, giving it much more flexibility than the binary "standard–short" versions commonly seen in scales.

AIMS is the most tangible output of our study. Compared to the self-report measures we examined, it has higher correlation with the general impulsivity factor *I*, higher internal consistency, and higher test–retest reliability, and it correlates more highly with impulsivity-related behaviors in general. Importantly, it was developed based on a large sample of participants with diverse backgrounds, rather than only college students, and after a careful analysis of the underlying construct structure of impulsivity. Although its external validity needs to be further tested, we anticipate that AIMS—facilitated by the flexibility of tailoring its length to practical needs—is likely to prove highly valuable in measuring trait impulsivity, understanding the impacts of impulsivity on daily behavior, and predicting and diagnosing abnormal behaviors.

Methods

Participants, Measures, and Study Procedure.

Participants. We recruited participants via social media and online advertisement. A total of 1,797 participants completed all measures within a required two-week period in the first measurement round. After excluding 121 participants who did not meet the data inclusion criteria (see details in *Data Processing*), the final sample consisted of 1,676 participants, M_{age} = 28.87 y, age range 17 to 65 y, 59.7% female (see detailed demographic statistics in *SI Appendix*, Table S1). Three months after the first measurement round, we retested a subsample of 211 participants. In this second measurement round, stratified random sampling was used for participant selection, in which participants were divided into strata based on sex, education, occupation, and age (under 30 y, 30 to 44 y, and 45 y and above). Among the retest subsample, 196 participants were included after data processing and imputation.

The study was approved by the Ethics Committee of the Institute of Psychology, Chinese Academy of Sciences (Approval #H20031).

Measures. In each measurement round, we administered 10 self-report scales and 10 behavioral tasks and asked participants to report their frequencies of engagement in seven impulsivity-related behaviors in the past 3 mo. We divided the seven impulsivity-related behaviors into three questionnaires: one on smoking and drinking, one on impulsive buying and impulsive eating, and one on internet gaming, short video app usage, and social media usage. Table 1 lists all measures; detailed descriptions of these measures are provided in *SI Appendix*.

Study Procedure. The study was conducted on an internet platform designed specifically for the study. All participants received general instructions and gave their consent prior to the start of the study. The study was divided into two parts: a survey session and an impulsivity measurement session. Participants completed surveys assessing their frequencies of seven impulsivity-related behaviors before completing the impulsivity measures. Among the impulsivity measures, self-report scales and behavioral tasks were presented in alternate orders, and the orders of the measures in each category were randomized for each participant. Upon completing a round of measurement, participants who met the data-inclusion criterion were given a fixed participation fee of 200 RMB and an additional bonus contingent on their performance in four incentivized behavioral tasks: passive avoidance with loss of reward, information sampling, the Balloon Analogue Risk Task, and decision from experience. Three months later, 211 participants who met the data-inclusion criterion in the first round of measurement were invited to take part in the second round. On average, participants earned 237 RMB (roughly $35) in the first round and 241 RMB in the second round.

The study was preregistered (https://aspredicted.org/blind.php?x=BFS_57R), and the data are open to access at Open Science Framework (52).

Data Preprocessing

Data Quality Control. Table 2 lists the criteria we used for data quality control in determining whether a participant paid sufficient attention in the study. We first implemented some control criteria in four behavioral tasks. In each task, if a participant did not meet the criterion or criteria on the first try, they were asked to redo the task for a maximum of three times; after three tries, the participant was allowed to proceed in the study regardless of whether their performance met the criterion or criteria. When participants tried a task multiple times, their best performance was recorded as their performance in the task.

For each of the 10 self-report scales and three questionnaires, we added a check item ("When you see this item, please choose the left option"), resulting in a total of 13 check items. For each self-report scale, we also added three to five lie items, depending on the length of the scale, from the lie subscale of either the Eysenck Personality Questionnaire (53) or the Revised Eysenck Personality Questionnaire Short Scale for Chinese (54). There were 21 unique lie items and 34 items administered in total; participants who gave disguised answers to 20 or more of these items were judged to have intended to lie. Furthermore, we inserted one probe question in the delay discounting task, in which one option was dominant over the other one (e.g., getting 50 RMB today versus getting 5 RMB in 1 wk), after every 50 questions. Participants who did not choose the dominant option in half or more of the probe questions were judged to not have paid sufficient attention to the task. Overall, there were a total of 15 data inclusion criteria, including the 13 check items, one lie tendency judgment, and one attention test in delay discounting. Participants who met 10 or more of these criteria were included and eligible for payment.

Of the 1,797 participants who completed all measures in the first round, 1,676 met the data-inclusion criteria, as did all 211 participants in the second round. The retest sample was important in calculating the test–retest reliabilities of measures and the development of an impulsivity scale. Because invalid data (*Data Imputation*) in this sample would have large adverse impacts on these analyses, we removed 15 participants from the retest sample due to invalid data, leaving a total of 196 participants in this sample for data analyses.

Data Transformation. For slightly right-skewed measures (i.e., BART, SST, I7i, I7v, DIIf, and DIId), we did square root transformations of the original values. For heavily right-skewed measures (i.e., DD1, DD2, and DD3), we did log transformations. We did not transform values of other measures, because they either were close to being normally distributed or could not be easily transformed to normal distributions (see *SI Appendix*, Fig. S1 for the distributions of the original values of all measures). Of the seven impulsivity-related behaviors, drinking, smoking, and internet gaming had heavily skewed distributions. For each, we transformed the original values into ordinary bins. There were at least 50 participants in each bin, and the number of bins was maximized (31).

Data Imputation. We identified invalid results and treated them as missing values. *SI Appendix*, Table S2 shows how invalid results were defined in each measure. Of the 80,448 data points in the first round (1,676 participants × 48 DVs), 472 were deemed invalid (0.59%); of the 10,128 data points in the second round (211 participants × 48 DVs), 39 were deemed invalid (0.39%). To avoid convergence issues for the main analyses in the first round, particularly in latent variable modeling, we imputed missing data values by the means obtained from Gibbs sampling using the R package mice (55). The imputation of the missing data affected the correlations between different measures only negligibly, with the absolute values of the correlations changing on average by 0.0017 and the most affected correlation changing by an absolute value of 0.016.

Table 2. Criteria for data quality control

Task	Indicator	Criterion
Quality control in certain behavioral task		
Go/no-go	Correct response rate	Go trials: ≥ 0.75 No-go trials: ≥ 0.25
Stop signal task	Correct response rate	Go trials: ≥ 0.75
Immediate and Delayed Memory Task (IMT and DMT)	Correct response rate in IMT	Target trials: ≥ 0.60 Filler trials: ≥ 0.90
	Correct response rate in DMT	Target trials: ≥ 0.75 Filler trials: ≥ 0.90
Synthetic face identification task	Correct response rate	> 0.90
Data inclusion criteria		
Self-report scales	Lie items	Disguised answers < 20
	One check item in each	Choose the leftmost option
Three frequency questionnaires pertaining to impulsivity-related behaviors	One check item in each	Choose the leftmost option
Delay discounting task	Two probe questions	Pass rate ≥ 0.5
	More than two probe questions	Pass rate > 0.5

Linear Regression. To reduce the effects of demographic variables (i.e., age, sex, occupation, and education) on the correlations between impulsivity measures and between measures and impulsivity-related behaviors, we first ran linear regression models with the four demographic variables as predictors for each of the impulsivity measures and impulsivity-related behaviors, then used the resulting residuals for the main analyses.

Main Analyses

Correlation Analysis. Because some impulsivity measures were not normally distributed, we computed the Spearman rank correlations between the measures. To visualize the results, we used a force-directed algorithm to generate a network plot such that correlated measures attracted each other and uncorrelated ones repulsed each other (Fig. 1).

Latent Variable Modeling. We applied bifactor models to examine whether there was a general factor among the impulsivity measures (Fig. 2 and *SI Appendix*, Fig. S2). In a bifactor model, the extracted general factor directly accounts for shared variance at the level of measures, leaving the residual variance to be captured by specific, orthogonal factors. Thus, compared to a hierarchical model, a bifactor model is a more direct test for the presence of a general factor.

We first standardized the data to avoid convergence issues and then used Kaiser–Meyer–Olkin and Bartlett's tests to check the suitability of the data for bifactor modeling. After that, we used parallel analysis, variable selection strategy (VSS), and measurement and assessment program (MAP) to determine the number of factors. Although the parallel analysis suggested a nine-factor solution, both the VSS and the MAP suggested an eight-factor solution, and with eight factors, the Bayesian information criterion was minimized. Therefore, we ran a bifactor exploratory factor analysis (EFA) in the R package GPArotation (34) across all 48 measures with an eight-factor solution and maximum likelihood estimation. To ensure that each factor contained more than three measures so that the model could be recognized, we combined the four measures of information sampling into one factor, resulting in a final bifactor model of seven factors (i.e., one general factor plus six specific factors).

Next, to determine the fit of the resulting factor structure, we implemented a bifactor CFA, estimating the factor loadings of all measures on the general impulsivity factor, as well as the loadings of measures that loaded above 0.30 on any of the six specific factors in the preceding EFA (see results in *SI Appendix*, Table S5). This model was estimated using the R package lavaan (56) with the weighted least-squares mean and variance estimator, using diagonally weighted least squares and computing robust SE, and all factors forced to be orthogonal, as defined by the standard bifactor model. There was one measure, BASf, that had loadings on two specific factors. Because the bifactor model generally does not permit cross-loadings (38), we retained the loading of BASf on one factor on which it had a higher loading and set the loading on the other factor to zero.

Finally, because the self-report measures loaded much higher on the general factor than the behavioral measures, we ran the same bifactor model analyses with only the 28 self-report measures (*SI Appendix*, Fig. S2).

Test–Retest Reliability. To assess the temporal stability of the impulsivity measures and of the extracted factors, we calculated the test–retest Spearman correlations on the retest sample of 196 participants.

Predicting Behaviors With Machine-Learning Algorithms. We applied seven popular machine-learning algorithms to predict impulsivity-related behaviors with four sets of predictors as the input. We first converted each impulsivity-related behavior to two categories, high and low, in terms of frequency of engagement. Specifically, we labeled participants who scored in the top 10% as high and the remaining as low. In the first and second measurement rounds, 167 and 16 participants, respectively, were in the high category for impulsive buying, 200 and 24 for impulsive eating, 134 and 9 for smoking, 194 and 20 for drinking, 203 and 21 for social media usage, 172 and 20 for short video app usage, and 187 and 21 for internet gaming. The goal of the machine-learning algorithms was to predict participants' categories based on demographic and measurement data.

Four sets of predictors were fed to the algorithms. Set 1 served as the baseline, including only demographic information (i.e., age, sex, education, and occupation); set 2 included demographic information and the general factor *I*; set 3 added the specific factors on top of the set 2 features; and set 4 added the behavioral measures on top of the set 2 features. By comparing the prediction performances using predictor sets 1 and 2, we could determine the predictive power of factor *I*, and by comparing the performances using predictor set 2 and set 3 (or set 4), we could determine the added predictive power of the specific factors (or the behavioral measures).

Each behavior was predicted by the following seven machine-learning algorithms: logistic regression with regularization, decision tree, support vector machine, Gaussian Naïve Bayes, random forest, AdaBoost, and multilayer perceptron neural network. For each algorithm using each predictor set, we first performed a grid search via tenfold cross-validation to identify the best hyperparameters. Algorithm performance was evaluated by the metric d' because it is more suitable than accuracy rate for unbalanced category distributions. Next, we applied the best hyperparameters to train and test an algorithm for 1,000 times (i.e., 100 iterations of 10-fold cross-validation), and the average performance over the 1,000 testing sets was taken as the algorithm's final performance. We also applied the algorithm to the retest sample to assess its cross-time prediction performance, using information collected in the first measurement round to predict impulsivity-related behaviors reported in the second round (see *SI Appendix*, Fig. S5 for the results).

For a certain behavior predicted with a specific set of predictors, we compared the seven machine-learning algorithms and selected the one with the highest prediction performance as the best model. For algorithms with similar d's, the best model was selected based on the hit (or true positive) rate, because the consequence of a false negative is usually more severe than the consequence of a false positive in diagnostic settings. *SI Appendix*, Figs. S3 and S4 show the detailed performances of each machine-learning algorithm.

Developing an Impulsivity Scale. We developed an impulsivity scale, the AIMS, using a data-driven two-step process. First, we divided participants randomly into a training group and a testing group with a ratio of 4:1 and repeated the split 100 times. Within each split, we randomly sampled 50 items from the item pool that consisted of 263 unique items from all administered scales, and then computed the correlation between the total score of these items and the score of factor *I*, as well as the Cronbach's alpha of these items, for both the training group and the testing group. If the correlation with factor *I* was above 0.90 and Cronbach's alpha above 0.80, the sample of items was treated as a quality sample; this item sampling process was repeated 10,000 times for each participant split. Thus, a total of 10 million item samples (i.e., 100 splits × 10,000 samplings) were evaluated in this step.

Second, we calculated an item's frequency in the quality samples and ranked all items according to this frequency. After compiling a ranking list of items, we manually checked item similarities and removed items that were highly similar to a more highly ranked item. The top 50 items in this processed list were then included in AIMS-50. *SI Appendix*, Table S10 shows the items with their original response formats and reports the psychometric properties of AIMS-50, as well as those of AIMS of any length (e.g., AIMS-25). *SI Appendix*, Table S11 shows the 50 items of AIMS with a uniform four-point response scale. This reformatted AIMS should make the scale easier and more convenient to use in practice.

Data, Materials, and Software Availability. Anonymized data (Measurement data) have been deposited in Open Science Framework (52).

ACKNOWLEDGMENTS. We thank Bo Xu and his programming team for help in designing and managing the internet-based data collection platform; Deborah Ain for editing the article; and members of the Risk and Uncertainty Management lab for discussions and help on various aspects of the study.

Author affiliations: [a]Key Laboratory for Behavioral Science, Institute of Psychology, Chinese Academy of Sciences, Beijing 100101, China; [b]Department of Psychology, University of the Chinese Academy of Sciences, Beijing 101408, China; [c]Department of Psychology, Renmin University of China, Beijing 100872, China; and [d]Center for Adaptive Rationality, Max Planck Institute for Human Development, Berlin 14195, Germany

1. T. Fahy, I. Eisler, Impulsivity and eating disorders. *Br. J. Psychiatry* **162**, 193–197 (1993).
2. A. Meyer-Lindenberg *et al.*, Neural mechanisms of genetic risk for impulsivity and violence in humans. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 6269–6274 (2006).
3. X. Wang, S. Zhao, M. X. Zhang, F. Chen, L. Chang, Life history strategies and problematic use of short-form video applications. *Evol. Psychol. Sci.* **7**, 39–44 (2021).
4. American Psychiatric Association, *Diagnostic and Statistical Manual of Mental Disorders* (American Psychiatric Association, ed. 5, 2013).
5. D. Bennett *et al.*, Systematic overestimation of reflection impulsivity in the information sampling task. *Biol. Psychiatry* **82**, E29–E30 (2017).
6. K. Ioannidis, R. Hook, K. Wickham, J. E. Grant, S. R. Chamberlain, Impulsivity in gambling disorder and problem gambling: A meta-analysis. *Neuropsychopharmacology* **44**, 1354–1361 (2019).
7. J. B. Richards, L. Zhang, S. H. Mitchell, H. de Wit, Delay or probability discounting in a model of impulsive behavior: Effect of alcohol. *J. Exp. Anal. Behav.* **71**, 121–143 (1999).
8. J. P. Guilford, R. B. Guilford, Personality factors D, R, T, and A. *J. Abnorm. Soc. Psychol.* **34**, 21–36 (1939).
9. S. B. G. Eysenck, H. J. Eysenck, Impulsiveness and venturesomeness: Their position in a dimensional system of personality description. *Psychol. Rep.* **43**, 1247–1255 (1978).
10. J. A. Gray, "A critique of Eysenck's theory of personality" in *A Model for Personality*, H. J. Eysenck, Ed. (Springer, 1981).
11. J. MacKillop *et al.*, The latent structure of impulsivity: Impulsive choice, impulsive action, and impulsive personality traits. *Psychopharmacology* **233**, 3361–3370 (2016).
12. S. P. Whiteside, D. R. Lynam, The Five Factor Model and impulsivity: Using a structural model of personality to understand impulsivity. *Pers. Individ. Differ.* **30**, 669–689 (2001).
13. J. H. Patton, M. S. Stanford, E. S. Barratt, Factor structure of the Barratt impulsiveness scale. *J. Clin. Psychol.* **51**, 768–774 (1995).
14. F. Verbruggen *et al.*, A consensus guide to capturing the ability to inhibit actions and impulsive behaviors in the stop-signal task. *eLife* **8**, e46323 (2019).
15. C. W. Lejuez *et al.*, Evaluation of a behavioral measure of risk taking: The Balloon Analogue Risk Task (BART). *J. Exp. Psychol. Appl.* **8**, 75–84 (2002).
16. K. G. Creswell, A. G. C. Wright, J. D. Flory, C. J. Skrzynski, S. B. Manuck, Multidimensional assessment of impulsivity-related measures in relation to externalizing behaviors. *Psychol. Med.* **49**, 1678–1690 (2019).
17. L. Sharma, K. E. Markon, L. A. Clark, Toward a theory of distinct types of "impulsive" behaviors: A meta-analysis of self-report and behavioral measures. *Psychol. Bull.* **140**, 374–408 (2014).
18. E. S. Barratt, Anxiety and impulsiveness related to psychomotor efficiency. *Percept. Motor Skills* **9**, 191–198 (1959).
19. S. J. Dickman, Functional and dysfunctional impulsivity: Personality and cognitive correlates. *J. Pers. Soc. Psychol.* **58**, 95–102 (1990).
20. C. S. Carver, T. L. White, Behavioral inhibition, behavioral activation, and affective responses to impending reward and punishment: The BIS/BAS Scales. *J. Pers. Soc. Psychol.* **67**, 319–333 (1994).
21. M. A. Cyders *et al.*, Integration of impulsivity and positive mood to predict risky behavior: Development and validation of a measure of positive urgency. *Psychol. Assess.* **19**, 107–118 (2007).
22. R. Mata, R. Frey, D. Richter, J. Schupp, R. Hertwig, Risk preference: A view from psychology. *J. Econ. Perspect.* **32**, 155–172 (2018).
23. S. D. Lane, D. R. Cherek, H. M. Rhodes, C. J. Pietras, O. V. Tcheremissine, Relationships among laboratory and psychometric measures of impulsivity: Implications in substance abuse and dependence. *Addict. Disord. Their Treat.* **2**, 33–40 (2003).
24. B. Reynolds, A. Ortengren, J. B. Richards, H. de Wit, Dimensions of impulsive behavior: Personality and behavioral measures. *Pers. Individ. Differ.* **40**, 305–315 (2006).
25. F. Baker, M. W. Johnson, W. K. Bickel, Delay discounting in current and never-before cigarette smokers: Similarities and differences across commodity, sign, and magnitude. *J. Abnorm. Psychol.* **112**, 382–392 (2003).
26. M. T. Fillmore, J. Weafer, "Behavioral inhibition and addiction" in *The Wiley-Blackwell Handbook of Addiction Psychopharmacology*, J. MacKillop, H. de Wit, Eds. (John Wiley & Sons, 2013), pp. 135–164.
27. T. L. White, C. W. Lejuez, H. de Wit, Test-retest characteristics of the Balloon Analogue Risk Task (BART). *Exp. Clin. Psychopharmacol.* **16**, 565–570 (2008).
28. Y. Chuang, L. Schechter, Stability of experimental and survey measures of risk, time, and social preferences: A review and new results. *J. Dev. Econ.* **117**, 151–170 (2015).
29. A. Z. Enkavi *et al.*, Large-scale analysis of test–retest reliabilities of self-regulation measures. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 5472–5477 (2019).
30. J. C. Strickland, M. W. Johnson, Rejecting impulsivity as a psychological construct: A theoretical, empirical, and sociocultural argument. *Psychol. Rev.* **128**, 336–361 (2021).
31. R. Frey, A. Pedroni, R. Mata, J. Rieskamp, R. Hertwig, Risk preference shares the psychometric structure of major psychological traits. *Sci. Adv.* **3**, e1701381 (2017).
32. C. Spearman, "General intelligence" objectively determined and measured. *Am. J. Psychol.* **15**, 201–292 (1904).
33. A. Caspi *et al.*, The p factor: One general psychopathology factor in the structure of psychiatric disorders? *Clin. Psychol. Sci.* **2**, 119–137 (2014).
34. C. A. Bernaards, R. I. Jennrich, Gradient projection algorithms and software for arbitrary rotation criteria in factor analysis. *Educ. Psychol. Meas.* **65**, 676–696 (2005).
35. S. P. Reise, J. Morizot, R. D. Hays, The role of the bifactor model in resolving dimensionality issues in health outcomes measures. *Qual. Life Res.* **16**, 19–31 (2007).
36. J. MacKillop *et al.*, Multidimensional examination of impulsivity in relation to disordered gambling. *Exp. Clin. Psychopharmacol.* **22**, 176–185 (2014).
37. C. Stahl *et al.*, Behavioral components of impulsivity. *J. Exp. Psychol.: Gen.* **143**, 850–886 (2014).
38. S. P. Reise, T. M. Moore, M. G. Haviland, Bifactor models and rotations: Exploring the extent to which multidimensional data yield univocal scale scores. *J. Pers. Assess.* **92**, 544–559 (2010).
39. M. Zuckerman, E. A. Kolin, L. Price, I. Zoob, Development of a sensation-seeking scale. *J. Consult. Psychol.* **28**, 477–482 (1964).
40. J. Weafer, M. J. Baggott, H. de Wit, Test–retest reliability of behavioral measures of impulsive choice, impulsive action, and inattention. *Exp. Clin. Psychopharmacol.* **21**, 475–481 (2013).
41. S. Palminteri, C. Chevallier, Can we infer inter-individual differences in risk-taking from behavioral tasks? *Front. Psychol.* **9**, 2307 (2018).
42. J. Dang, K. M. King, M. Inzlicht, Why are self-report and behavioral measures weakly correlated? *Trends Cogn. Sci.* **24**, 267–269 (2020).
43. C. Hedge, G. Powell, P. Sumner, The reliability paradox: Why robust cognitive tasks do not produce reliable individual differences. *Behav. Res. Methods* **50**, 1166–1186 (2018).
44. A. J. Caswell, R. Bond, T. Duka, M. J. Morgan, Further evidence of the heterogeneous nature of impulsivity. *Pers. Individ. Differ.* **76**, 68–74 (2015).
45. G. T. Smith *et al.*, On the validity and utility of discriminating among impulsivity-like traits. *Assessment* **14**, 155–170 (2007).
46. S. P. Reise, The rediscovery of bifactor measurement models. *Multivar. Behav. Res.* **47**, 667–696 (2012).
47. F. F. Chen, S. G. West, K. H. Sousa, A comparison of bifactor and second-order models of quality of life. *Multivar. Behav. Res.* **41**, 189–225 (2006).
48. A. Verdejo-Garcia *et al.*, A unified online test battery for cognitive impulsivity reveals relationships with real-world impulsive behaviors. *Nat. Hum. Behav.* **5**, 1562–1577 (2021).
49. I. Adjerid, K. Kelley, Big data in psychology: A framework for research advancement. *Am. Psychol.* **73**, 899–917 (2018).
50. K. Casler, L. Bickel, E. Hackett, Separate but equal? A comparison of participants and data gathered via Amazon's MTurk, social media, and face-to-face behavioral testing. *Comput. Hum. Behav.* **29**, 2156–2160 (2013).
51. J. C. Peterson, D. D. Bourgin, M. Agrawal, D. Reichman, T. L. Griffiths, Using large-scale experiments and machine learning to discover theories of human decision-making. *Science* **372**, 1209–1214 (2021).
52. Y. Huang *et al.*, Data from "Impulsivity is a stable, measurable, and predictive psychological trait." ImpulsivityData. https://osf.io/f8g5s/files/osfstorage. Deposited 12 May 2024.
53. H. J. Eysenck, S. B. G. Eysenck, *Manual of the Eysenck Personality Questionnaire* (Educational and Industrial Testing Service, 1975).
54. M. Qian, G. Wu, R. Zhu, S. Zhang, Development of the revised Eysenck personality questionnaire short scale for Chinese (EPQ-RSC). *Acta Psychol. Sin.* **32**, 317–323 (2000).
55. S. van Buuren, K. Groothuis-Oudshoorn, mice: Multivariate imputation by chained equations in R. *J. Stat. Software* **45**, 1–67 (2011).
56. Y. Rosseel, lavaan: An R package for structural equation modeling. *J. Stat. Software* **48**, 1–36 (2012).