Rapid-Response Impulsivity: Definitions, Measurement Issues, and Clinical Implications

Kristen R. Hamilton
University of Maryland

Andrew K. Littlefield
Texas Tech University

Noelle C. Anastasio, Kathryn A. Cunningham, and Latham H. L. Fink
University of Texas Medical Branch

Victoria C. Wing
Centre for Addiction and Mental Health, Toronto, Canada

Charles W. Mathias
University of Texas Health Science Center at San Antonio

Scott D. Lane
University of Texas at Houston Medical School

Christian G. Schütz
University of British Columbia

Alan C. Swann
Baylor College of Medicine

C. W. Lejuez
University of Maryland

Luke Clark
University of British Columbia

F. Gerard Moeller
Virginia Commonwealth University School of Medicine

Marc N. Potenza
Yale University School of Medicine and Connecticut Mental Health Center, New Haven, Connecticut

Impulsivity is a multifaceted construct that is a core feature of multiple psychiatric conditions and personality disorders. However, progress in understanding and treating impulsivity is limited by a lack of precision and consistency in its definition and assessment. Rapid-response impulsivity (RRI) represents a tendency toward immediate action that occurs with diminished forethought and is out of context with the present demands of the environment. Experts from the International Society for Research on Impulsivity (InSRI) met to discuss and evaluate RRI measures in terms of reliability, sensitivity, and validity, with the goal of helping researchers and clinicians make informed decisions about the use and interpretation of findings from RRI measures. Their recommendations are described in this article. Commonly used clinical and preclinical RRI tasks are described, and considerations are provided to guide task selection. Tasks measuring two conceptually and neurobiologically distinct types of RRI, “refraining from action initiation” (RAI) and “stopping an ongoing action” (SOA) are described. RAI and SOA tasks capture distinct aspects of RRI that may relate to distinct clinical outcomes. The InSRI group recommends that (a) selection of RRI measures should be informed by careful consideration of the strengths, limitations, and practical considerations of the available measures; (b) researchers use both RAI and SOA

Kristen R. Hamilton, Department of Psychology, Maryland Neuroimaging Center, Center for Addictions, Personality, and Emotion Research, University of Maryland; Andrew K. Littlefield, Department of Psychology, Texas Tech University; Noelle C. Anastasio, Center for Addiction Research, Department of Pharmacology and Toxicology, University of Texas Medical Branch; Kathryn A. Cunningham, Center for Addiction Research, Department of Pharmacology and Toxicology, University of Texas Medical Branch; Latham H. L. Fink, Center for Addiction Research, University of Texas Medical Branch; Victoria C. Wing, Schizophrenia Division, Complex Mental Illness, Centre for Addiction and Mental Health, Toronto, Ontario, Canada; Charles W. Mathias, Department of Psychiatry, Division of Neuropsychiatric Research, University of Texas Health Science Center at San Antonio; Scott D. Lane, Department of Psychiatry and Behavioral Sciences, University of Texas at Houston Medical School; Christian G. Schütz, Department of Psychiatry, University of British Columbia; Alan C. Swann, Menninger Department of Psychiatry and Behavioral Sciences, Baylor College of Medicine; C. W. Lejuez, Department of Psychology, Maryland Neuroimaging Center, Center for Addictions, Personality, and Emotion Research, University of Maryland; Luke Clark, Centre for Gambling Research at UBC, Department of Psychology, University of British Columbia; F. Gerard Moeller, Departments of Psychiatry and Pharmacology and Toxicology, Virginia Commonwealth University School of Medicine; Marc N. Potenza, Departments of Psychiatry, Neurobiology, and Child Study Center, Yale University School of Medicine, and Connecticut Mental Health Center, New Haven, Connecticut.

continued
Impulsive people have tendencies to act rapidly with diminished consideration of future consequences, often to their detriment. Impulsivity has been associated with a range of psychiatric conditions and represents a hallmark feature of multiple personality disorders (PDs; Swann, Bjork, Moeller, & Dougherty, 2002), with a focus on Cluster B PDs. Impulsivity has been associated with problem behaviors including substance use (de Wit, 2009; Lejuez et al., 2010), problem gambling (Grant, Chamberlain, O'Callaghan, Potenza, & Kim, 2010; Lawrence, Luty, Bogdan, Sahakian, & Clark, 2009), aggression against others (Mouloso, Calhoun, & Rosenbloom, 2013), deliberate self-harm (Di Pierro, Sarno, Perego, Gallucci, & Madeddu, 2012), and suicidality (Swann et al., 2005).

As a construct, impulsivity can be measured either as a relatively stable characteristic using self-report questionnaires (K. R. Hamilton, Sinha, & Potenza, 2012; Littlefield, Sher, & Steinley, 2010) or as a characteristic sensitive to contexts or states, which may be assessed by behavioral tasks and/or self-report assessments (Fillmore & Weaver, 2013). Different forms of impulsivity have been proposed, and factor analyses indicate the presence of two or more types of impulsivity (Meda et al., 2009; Reynolds, Ortengren, Richards, & de Wit, 2006). The number and types of impulsivity factors (or impulsivity-related factors) have been discussed and debated (Gullo, Loxton, & Dawe, 2014). For example, a recent review identified and described, on the basis of theoretical, behavioral, and biological findings, four domains of impulsivity relating to response, choice, reflection, and decision-making (Fineberg et al., 2014). Other studies have used factor analysis to identify separable constructs related to impulsivity, with up to five domains or factors identified depending on the study (Meda et al., 2009; Reynolds et al., 2006). There has been discussion regarding the boundaries of impulsivity, with some researchers calling for careful consideration of the number and types of domains contributing to impulsivity with a harkening for parsimony (Gullo et al., 2014). Although multiple impulsivity-related domains have been identified, the constructs link back to definitions of impulsivity based on tendencies relating to acting rapidly and/or with diminished forethought or consideration of negative consequences to oneself or others (Fineberg et al., 2014; Moeller, Barratt, Dougherty, Schmitz, & Swann, 2001). Two types of impulsivity identified in multiple studies are delayed reward, or “choice,” impulsivity and rapid-response impulsivity (RRI; Winstanley, Dalley, Theobald, & Robbins, 2004). Choice impulsivity is characterized by a diminished ability or willingness to tolerate delay. RRI reflects a tendency toward immediate action that is out of context with the present demands of the environment and that occurs with diminished forethought; RRI also has been described as a diminished ability to inhibit prepotent responses (Moeller et al., 2001).

Choice impulsivity and RRI are distinct constructs that, although they link back to the core theoretical definitions of impulsivity, they correlate weakly or not at all (Broos, Diergaard, Schoffelmeeer, Pattij, & De Vries, 2012; Lane, Cherek, Rhoades, Pietras, & Thermenissine, 2003; Reynolds et al., 2006), perhaps reflecting their differences in underlying neurobiology (van Gaalen, Bruegge, Bronius, Schoffelmeeer, & Vanderschuren, 2006, van Gaalen, van Koten, Schoffelmeeer, & Vanderschuren, 2006; Winstanley, Theobald, Dalley, Glennon, & Robbins, 2004). Each type of impulsivity may contribute uniquely to specific phases of psychiatric disorders, such as addictions (de Wit, 2009). However, conflation of types of impulsivity has led to inconsistencies across research domains and disciplines, slowing scientific progress (Cyders & Coskunpinar, 2011; Smith et al., 2007).

Keywords: impulsivity, response, behavioral control, personality, assessment
Standardized RRI assessments are needed to inform research and clinical practice and to promote public health. Assessment of participants from the general population is required to examine the normative distribution of RRI, to study its association with risky behaviors, and to provide comparison data for groups diagnosed with specific disorders. Developmental changes in RRI throughout the life span (during childhood, adolescence, adulthood, and senium) also need to be examined. Within psychiatric samples, determination of associations with symptom severity, prognosis, and treatment outcome is critical. Measures of RRI can be used to assess changes over time resulting from pharmacological and behavioral manipulations and changes in disorder states, for example, incremental fluctuations in functioning over time in PDs to more discrete episode changes in mood disorders.

To address the existing situation, the International Society for Research on Impulsivity (InSRI) convened at the 2012 annual meeting to discuss the definition and assessment of RRI across species and in special populations (e.g., developmentally and in groups with psychiatric illness). RRI measures were considered in terms of reliability, sensitivity, and validity (see Table 1), with the goal of helping researchers and clinicians make informed decisions about the use and interpretation of findings from RRI measures. Differences in types of RRI were considered. Specifically, RRI measurement and theory identifies two basic types of conceptually and neurobiologically distinct inhibitory errors: (a) failure to refrain from action initiation (RAI; such as a No-Go response commission error on a go/no-go (GNG) task versus (b) failure to stop an ongoing or prepotent action (SOA; such as a stop error on a stop-signal task [SST]) (Rubia et al., 2001; Swick, Ashley, & Turken, 2011).

### Rapid-Response Impulsivity Neurocircuitry

Neuroimaging studies are valuable for identifying areas of activation implicated in RRI, and lesion studies (as well as studies involving temporary activation/inactivation of neural regions) provide critical confirmatory evidence of neuroimaging results (see Bari & Robbins, 2013, for a review). Response inhibition requires the activation of a complex circuit that includes the inferior frontal cortex and presupplementary motor area (pre-SMA) as major components (Bari & Robbins, 2013; Isoda & Hikosaka, 2007). Additional regions important for response inhibition that have been identified in fMRI and lesion studies include the supplementary motor area (SMA) (Mostofsky et al., 2003; Simmonds, Pekar, & Mostofsky, 2008), premotor cortex (Picton et al., 2007; Watanabe et al., 2002), parietal cortex (Menon, Adleman, White, Glover, & Reiss, 2001; Rubia et al., 2001), ventrolateral prefrontal cortex (PFC) and insula (Bari & Robbins, 2013; Boehler, Appelbaum, Krebs, Hopf, & Woldorff, 2010; Swick, Ashley, & Turken, 2008). Lesion studies with rodents implicate the dorsomedial PFC in RRI on the SST (Bari et al., 2011) and five-choice serial reaction time task (5-CSRTT) (Muir, Everitt, & Robbins, 1996; Paine, Slipp, & Carlezon, 2011).

Although there is some overlap in the neural regions activated during performance on the SST and GNG tasks, there also are regions of activation that are specific to each task (Fineberg et al., 2014; Swick et al., 2011). The neural correlates of RRI have been examined in meta-analyses (Buchsbaum, Greer, Chang, & Berman, 2005; Simmonds et al., 2008), with some of these studies having examined SST performance together with GNG performance. The findings of these neuroimaging studies suggest that RRI is associated with a large-scale distributed system of bilateral

<table>
<thead>
<tr>
<th>RRI Measures</th>
<th>Human lab task</th>
<th>Type of RRI</th>
<th>Internal validity</th>
<th>External validity</th>
<th>Construct validity</th>
<th>Discriminant validity</th>
<th>Reliability</th>
</tr>
</thead>
<tbody>
<tr>
<td>GNG</td>
<td>RAI</td>
<td>Modest (owing to many different procedural variations)</td>
<td>Strong</td>
<td>Strong</td>
<td>Modest (may be correlated with other domains, e.g., short term memory)</td>
<td>Modest (generally stable but some versions may be subject to practice effects)</td>
<td></td>
</tr>
<tr>
<td>CPT</td>
<td>RAI</td>
<td>Strong</td>
<td>Strong</td>
<td>Modest</td>
<td>Modest (may be correlated with other domains such as attention)</td>
<td>Good (generally stable and limited practice effects)</td>
<td></td>
</tr>
<tr>
<td>SST</td>
<td>SOA</td>
<td>Strong</td>
<td>Strong</td>
<td>Strong</td>
<td>Strong</td>
<td>Strong</td>
<td></td>
</tr>
<tr>
<td>Antisaccade</td>
<td>RAI</td>
<td>Modest</td>
<td>Modest</td>
<td>Modest</td>
<td>Poor</td>
<td>Strong</td>
<td></td>
</tr>
</tbody>
</table>

Note. GNG = go/no-go; CPT = continuous performance test; SST = stop signal task; RAI = refraining from action initiation; SOA = stopping and ongoing action; 5-CSRTT = five-choice serial reaction time task; SSRT = stop-signal reaction time; est = established.
cortical and subcortical regions, with right hemisphere dominance (Swick et al., 2011).

It has been proposed that if distinct patterns of neural activation on the SST and GNG tasks exist, then it follows that the two tasks engage different cognitive processes (Lenartowicz, Kalar, Congdon, & Poldrack, 2010). In a meta-analysis in which SST-related activation was compared with GNG-related activation, key differences in the activation associated with each task were revealed (Swick et al., 2011), supporting a conceptual distinction between RAI and SOA. In the GNG, right-lateralized clusters were activated to a greater extent in the middle and superior frontal gyri, the inferior parietal lobule, and the precuneus when compared with the SST. By contrast, two foci were activated to a greater extent in the SST than in the GNG: the thalamus and the left insula (Swick et al., 2011).

Although there were important differences in the neural correlates of GNG and SST performance in the meta-analysis examining the two tasks, there also were two primary areas of overlap: bilateral anterior insular regions and the SMA/pre-SMA (Swick et al., 2011). These regions have been characterized as part of a “salience network” that is activated by personally relevant stimuli (Seeley et al., 2007). Hypotheses regarding insular involvement have been proposed that in addition to interoceptive awareness (Craig, 2009; Swick et al., 2011) include responding to salient events and initiating cognitive control (Menon & Uddin, 2010), capturing focal attention and task maintenance (Nelson et al., 2010), and coordinating appropriate responses to internal and external events (Medford & Critchley, 2010). Based on their results, Swick et al. (2011) suggest that the insula is important for maintenance of task rules and readiness, rather than response inhibition per se. The overlapping activation observed in the SMA/pre-SMA during GNG and SST performance is consistent with previous suggestions that these regions are a critical part of the circuit that executes response inhibition (Chao, Luo, Chang, & Li, 2009; Duann, Ide, Luo, & Li, 2009; Mostofsky & Simmonds, 2008). The combination of overlapping and distinct areas of activation during GNG and SST performance provides support for different subtypes of RRI.

A description of the proceedings of and recommendations from the meeting follows. SOA and RAI tasks used to assess RRI in research with animal models and human participants are described, and the premises, characteristics, strengths and limitations of each task are considered (see Table 1).

### Measurements of RRI in Animal Models

Preclinical measures of RRI are highly translational in that the tasks are employed across species (human, nonhuman primate, and rodent) with only minor alterations in design (Eagle, Bari, & Robbins, 2008; Robbins, 2002; Winstanley, 2011). Thus, these tasks facilitate the investigation across species of the neurobiological underpinnings of behaviors and disorders characterized by RRI. Multiple tasks have been developed to assess RRI in animal models.

#### Go/No-Go Task

The animal version of the GNG task requires the subject to learn to discriminate between two visual (or auditory) signals, one requiring a “go” response and the other requiring that the response be withheld, a “no-go” response, similar to the human versions of this task (Harrison, Everitt, & Robbins, 1999; Hogg & Evans, 1975). Trials are presented in random order, with stimuli ranging from presentation of a constant light and/or flashing light to

<table>
<thead>
<tr>
<th>Primary outcome measures</th>
<th>Translational nonhuman analog</th>
<th>Availability?</th>
<th>Overall strengths</th>
<th>Overall weaknesses</th>
<th>InSRI recommendation</th>
</tr>
</thead>
<tbody>
<tr>
<td>False alarms; d’</td>
<td>Direct: Rat GNG, monkey GNG</td>
<td>Public domain; many versions</td>
<td>Widely used; est links to neurobiology and clinical outcomes; OK for repeated measures; sensitive to manipulations (drug)</td>
<td>Many different procedural variants and measurement techniques; potential influence by attentional factors?</td>
<td>YES</td>
</tr>
<tr>
<td>Commission errors/false alarms/catch trials; d’</td>
<td>5-CSRTT</td>
<td>Public and private domain; a limited number of widely used versions.</td>
<td>Widely used; est links with clinical outcomes; OK for repeated measures; sensitivity to manipulations; has been used in adolescents; norms available for Conners</td>
<td>High attentional load; may be influenced by working memory; few commission error trials makes less sensitive.</td>
<td>YES</td>
</tr>
<tr>
<td>SSRT</td>
<td>Yes, has rodent and nonhuman primate analogs</td>
<td>Many versions; most versions are in the public domain</td>
<td>Widely used; est links to neurobiology and clinical outcomes; OK for repeated measures; sensitive to manipulations; variations used in children</td>
<td>Motivational considerations, problems with data distribution; task is lengthy</td>
<td>YES</td>
</tr>
<tr>
<td>Reaction time; failure to antisaccade</td>
<td>Yes, has nonhuman primate analogs</td>
<td>Public domain</td>
<td>Sensitive to manipulations (drug); can be administered throughout life span; less susceptible to motivational effects</td>
<td>Costly; lack of standardization; susceptibility to confounds (e.g., differences in vision)</td>
<td>YES</td>
</tr>
</tbody>
</table>
variations in the light frequency (Harrison et al., 1999). Stimulus dynamics may affect task acquisition and performance, complicating interpretation of results. Accordingly, careful consideration of the type of visual stimulus is warranted (Harrison et al., 1999; Jakubowska & Gray, 1982). The proportion of go to no-go trials is essential, and symmetrical reinforcement of the no-go trials is necessary to ensure that animals do not adopt a go response bias (Harrison et al., 1999; Perry & Carroll, 2008). The primary measure of RRI in this task is the number of inappropriate responses on no-go trials (errors of commission, or false alarms). Additional measures include accuracy of responding, latency to respond correctly, and latency to retrieve the reinforcer.

The GNG task is highly translational. However, the GNG task has a decision-making element that can confound interpretation of RRI manipulations (Harrison et al., 1999; Schachar et al., 2007). Although promising, the translational utility of the GNG task following neurochemical or pharmacological manipulations on action restraint (false alarms) is difficult to assess as insufficient work in this regard has been performed preclinically (Winstanley, 2011). Nonetheless, the GNG task in animals has proven useful in the dissection of the complex neurocircuitry underlying RAI (Eagle et al., 2008).

Choice Serial Reaction Time Task

The CSRT tasks of attention and impulsive action are most closely modeled after the human continuous performance task (CPT). These tasks include the widely used 5-CSRT task (Carl, Robbins, Evenden, & Everitt, 1983; K. R. Hamilton, Potenza, & Grunberg, 2014), as well as the one-choice (1-CSRT, or “fixed-choice”) and two-choice (2-CSRT) variants (Anastasio et al., 2011, 2013; Cunningham et al., 2013; Dalley, Theobald, Eagle, Passetti, & Robbins, 2002; Dillon et al., 2009; Winstanley, Dalley, et al., 2004; Winstanley, Theobald, et al., 2004). These operant tasks entail a series of trials in which animals respond to a visual stimulus for delivery of a reinforcer. After completion of a trial, the animal must inhibit the acquired prepotent response during an intertrial interval (ITI); a response during the ITI, termed a premature response, is not reinforced and bears the further negative consequence of increased delay until the next trial. The principal measure of RRI is the premature response, which reflects a failure of RAI. Additional data include accuracy, omissions, and latency to collect the reinforcer, all of which can be used to delineate RRI from attentional and motivational processes.

Premature responses typically follow a normal distribution across populations and remain reliably stable over time (Dalley et al., 2007). The CSRT tasks demonstrate excellent sensitivity to experimental manipulation with genetic, pharmacological, and neuroanatomical approaches, as well as to alterations of task parameters (Pattij & Vanderschuren, 2008; Robbins, 2002). Baseline premature response rates are low for typical employment of the 5-CSRT task relative to the 1- or 2-CSRT task variants; however, modification to a lengthened or variable ITI substantially increases premature responding (Carl et al., 1983; Cole & Robbins, 1989; Dalley, Theobald, Eagle, Passetti, & Robbins, 2002). The high sensitivity of premature responding to ITI manipulations suggests a strong association between interval timing and this measure of RRI. Although it may appear that the involvement of interval timing may challenge construct validity of the CSRT tasks, clinical and preclinical studies have demonstrated a close, if not fundamental, relationship between interval timing and impulsivity (Rubia, Halari, Christakou, & Taylor, 2009).

The 1- and 2-CSRT tasks are less commonly employed but offer distinct advantages when assessing RRI. Decreased visual and attentional demands in these tasks may improve discriminant validity by reducing potential influence of these variables on task performance (Anastasio et al., 2011, 2013; Cunningham et al., 2013; Dalley et al., 2002; Winstanley, Dalley, et al., 2004). The elevated baseline premature responding in the 1- and 2-CSRT variants make these tasks amenable to experimental manipulations that reduce RRI, whereas a floor effect can hinder interpretation of such experiments in the 5-CSRT task. However, the neural mechanisms underlying premature responses in the 1-, 2-, and 5-CSRT tasks may not be identical; each variant may rely more heavily on particular neural circuits or neurotransmitters, differentially yielding certain neural correlates of RRI. These possibilities require further experimental confirmation.

Stop-Signal Task

The SST is similar to the GNG task except that the stop signal is presented after the go signal, thereby emphasizing the cancellation of a probable or ongoing motor response (Logan et al., 1984; Verbruggen & Logan, 2009). Performance on the SST has been described using a horse-race model, in which the stopping process and the reaction process (to the initial stimulus) compete for the first finishing time (Logan & Cowan, 1984). Following from this model, a response is inhibited when the stopping process finishes before the reaction process. In rodent models, animals are trained to respond (lever press) rapidly and accurately to first one then a second target following the go signal; the time to execute this sequence is the mean reaction time (mRT). On a subset of trials, the stop signal (e.g., auditory tone) is presented and the animal must cancel its prepotent response to obtain a reinforcer. Stop trials account for 20% of the trials in a test session and are randomly signaled after the rat responds on the first lever, but before the rat responds on the second lever. The delay to the stop signal varies across trials; prolonging the delay between the go and stop signals (i.e., presenting the stop signal closer to the mRT) increases the difficulty to inhibit the response (i.e., increases stop errors). The primary measure of the SST is the stop-signal reaction time (SSRT), which is inferred from a subject’s mRT and inhibition of responding at different stop signal delays (Logan et al., 1984).

The SST effectively measures SOA in both preclinical and clinical environments, a facet of RRI that is neuroanatomically and pharmacologically distinct from RAI (Eagle et al., 2008; Eagle & Baunez, 2010; Eagle et al., 2009; Rubia et al., 2001; Winstanley, 2011). Further, inherent levels of premature responding in the 5-CSRT do not correlate to individual SSRTs (Robinson et al., 2009), supporting the hypothesis that these tasks are elucidating independent measures of the RRI construct. One limitation of the SST task is the critical requirement that animals respond to the go signal as quickly as possible and cancel responding on all trials in which the stop signal is delivered to accurately estimate the SSRT; failure to do so can result in the exclusion of subjects from final statistical analyses. The SST is highly amenable to experimental
manipulations and exhibits very high cross-species comparability (Aron, Fletcher, Bullmore, Sahakian, & Robbins, 2003; Wiskerke et al., 2011); however, its pharmacological predictive validity is dependent on the class of drugs under investigation (Eagle et al., 2008; Winstanley, 2011).

Measurements of RRI in Humans

In addition to its harmful impact on behavior in normative populations, RRI is a principal component of a wide range of psychiatric conditions (Lipszyc & Schachar, 2010; Moeller et al., 2001; Wright, Lipszyc, Dupuis, Thayapararajah, & Schachar, 2014). As in preclinical RRI measurement, clinical measurement of RRI typically involves either RAI or SOA.

Go/No-Go

The GNG task was designed to assess ability to inhibit inappropriate responses, and was originally adapted from a rodent measure (Iversen & Mishkin, 1970). The GNG instructions ask participants to make motor responses as rapidly as possible to visual presentations of stimuli designated as “go,” and to withhold motor responses to stimuli with a “no-go” designation. Go events are typically more frequent than no-go events to establish the go response as dominant. (By contrast, the establishment of a dominant response is avoided in preclinical models, a procedural difference that should be considered when comparing clinical and preclinical GNG research.) Errors of omission (withholding a response when a go stimulus is presented) and errors of commission/false alarms (responding to a no-go stimulus) are recorded during the task, with the latter indexing RRI. The basic task can be modified in various ways. In reinforced versions of the GNG, participants are rewarded for correct responses and/or penalized for incorrect responses (Avila, 2001; Crockett, Clark, & Robbins, 2009; Newman, Wallace, Schmitt, & Arnett, 1997). Affective versions of the GNG (e.g., Cambridge Neuropsychological Test Automated Battery; CANTAB; Fray, Robbins, & Sahakian, 1996) allow comparison of inhibitory control to emotional distractors of different valences (positive vs. negative) (Murphy et al., 1999).

Construct validity, or the degree to which an instrument measures the intended underlying construct, is evaluated by examining factors including the instrument’s content validity, internal structure, and relations to other variables (Cook & Beckman, 2006). By precisely assessing RRI, or the ability/willingness to withhold responses, the GNG demonstrates strong content validity (Moeller et al., 2001). The GNG has a strong internal structure, which can be evaluated by examining the instrument’s reliability and factor structure (Cook & Beckman, 2006). With an r of 0.65, the GNG had a moderate to high level of test–retest reliability when participants were tested with a mean intersession interval of 8.6 days (Weafer, Baggott, & de Wit, 2013). The measure also has a strong factor structure, as it was part of a factor labeled Impulsive Disinhibition in a principal component analysis (Reynolds et al., 2006). The relations of the GNG to other variables, which are evidenced by discriminant validity and concurrent validity, also provide support for the strong construct validity of the GNG. The discriminant validity of the GNG is evidenced by its exclusive assessment of RRI relative to choice impulsivity and behavioral risk-taking (Reynolds et al., 2006). The GNG also has strong concurrent validity, correlating with other RRI measures, such as the SST (Reynolds, 2006). Taken together, these sources of validity evidence support the construct validity of the GNG.

Additionally, the GNG is sensitive to manipulations such as drug administration (Sofuoglu, Herman, Li, & Waters, 2012) and associates with clinical outcomes, such as smoking cessation (Berkman, Falk, & Lieberman, 2011). The neural substrates of the GNG have been characterized in fMRI studies (Simmonds et al., 2008). Additional strengths of the GNG include its high degree of translatability to preclinical models and its widespread use. From a practical perspective, GNG assessment is relatively brief and does not require extensive training. However, the different parameters of the many GNG variants may limit its internal validity, as multiple task variants may introduce confounding variables (Bodnar, Prahme, Cutting, Dencikl, & Mahone, 2007). Furthermore, attention, vigilance, and working memory also may affect GNG performance, thereby increasing the complexity of data interpretation. The GNG has been used to examine response impulsivity in a wide range of psychiatric conditions, including attention-deficit/hyperactivity disorder (ADHD), bipolar disorder, depression, and schizophrenia (Wright et al., 2014). In a meta-analysis of psychiatric research, the mean effect size for higher response impulsivity on the GNG was 0.56 in bipolar disorder, 0.48 in ADHD, 0.29 in schizophrenia, and 0.32 in addiction (Wright et al., 2014). In some research, participants with major depressive disorder (MDD) had higher levels of response impulsivity on the GNG than did healthy control participants (Kaiser et al., 2003; Katz et al., 2010).

Even in fMRI studies in which there were no differences in GNG performance, there were differences in the patterns of neural activation associated with RRI in some psychiatric conditions. For example, in a study in which participants had similar GNG performance, patients with remitted MDD had hypoactivity in the right dorsomedial PFC and right anterior cingulate cortex during response inhibition when compared with healthy control participants (Nixon, Liddle, Worwood, Liotti, & Dixon, 2013). In a study of adolescents, there were no group differences in GNG performance among adolescents who were depressed and who had attempted suicide, adolescents who were depressed and who had not attempted suicide, and healthy control participants (Pan et al., 2011). However, adolescents who were depressed and who had not attempted suicide had greater activation during response inhibition in the right anterior cingulate cortex than adolescents who were depressed and who had attempted suicide. Individuals with obsessive–compulsive disorder (OCD) displayed hypoactivity in the right anterior cingulate cortex, which patients with MDD had similar levels of response inhibition as healthy controls. For example, in a study in which patients with MDD had similar levels of response inhibition to healthy controls, the patients had greater activation in frontal, limbic, and temporal regions during response inhibition than healthy controls (Langenecker et al., 2007). Furthermore, greater activation in many of these regions was predictive of treatment response in the MDD patients. In a study of cocaine-dependent participants with clean urine screens, increased activation to no-go cues in the bilateral postcentral gyri was prospectively associated with drug use. This suggests that the GNG may be a useful tool for assessing treatment response in patients with cocaine dependence.
with having used cocaine at an assessment that took place one week later (Priciandaro, Myrick, Henderson, McRae-Clark, & Brady, 2013). Therefore, patterns of neural activation during response inhibition on the GNG may have value for characterizing different diagnostic groups and predicting behavioral outcomes.

**Continuous Performance Test**

The CPT is a GNG with unique attributes. In the CPT, participants are instructed to respond to target stimuli and to inhibit responses to incorrect stimuli that are similar to the target. Responses to incorrect stimuli, or commission errors, index RRI. Although there are many versions of the CPT, all involve the maintenance of focus throughout the duration of a repetitive task to respond to targets or inhibit responses. Therefore, in addition to measuring RRI as indexed by errors of commission, the CPT also measures sustained and selective attention (indexed by errors of omission).

The content validity of the CPT is strong, as the task measures ability/willingness to inhibit responses (Moeller et al., 2001). The CPT has high test–retest reliability with an r of 0.73 when the mean time between assessments was 8.6 days (Weafer et al., 2013). In addition, the convergent validity of the CPT is high, as CPT performance is correlated with GNG performance (Weafer et al., 2013). However, although high levels of content validity, reliability, and convergent validity support the construct validity of the CPT, reduced discriminant validity slightly diminishes the construct validity of the measure. The discriminant validity of the CPT is limited, as the task has commonly been used to measure attention (Harmell et al., 2014; Posada et al., 2012), vigilance (Bubnik, Hawk, Pelham, Waxmonsky, & Rosch, 2015), and working memory (Bartés-Serrallonga et al., 2014); impairments in these domains may impact task performance and reduce the number of trials that probe RRI via commission errors. Therefore, it is reasonable to conclude that although there is evidence to support the construct validity of the CPT, the evidence is not as strong as the evidence for the GNG task’s construct validity.

Strengths of the CPT include high external validity (Dougherty et al., 2003; Schepis, McFetridge, Chaplin, Sinha, & Krishnan-Sarin, 2011; Strakovski et al., 2010) and utility in adolescents (Schepis et al., 2011). The Conners’ CPT-II, in particular, is widely used clinically and has a set of well-validated norms that facilitate comparisons across studies (Conners, Epstein, Angold, & Klaric, 2003). However, different CPTs vary in task difficulty and response characteristics, and these differences warrant consideration in RRI studies. The CPT can be used during neuroimaging to provide information about the neural substrates underlying task performance (Moeller et al., 2005; Ogg et al., 2008; Sepede et al., 2010), although the neural features associated with CPT performance have been studied less extensively than those associated with performance of the GNG and SST.

The CPT has been widely used in clinical samples. In the previously discussed meta-analysis of psychiatric research (Wright et al., 2014), CPT effect sizes were comparable to effect sizes calculated from GNG research. The effect size for elevated scores on the CPT was 0.48 in bipolar disorder, 0.45 in ADHD, 0.37 in schizophrenia, and 0.30 in addiction (Wright et al., 2014).

The CPT has been used to evaluate the effects of a pharmacological treatment on prefrontal activity in children with ADHD (Araki et al., 2014). In an age-matched control group, children had increased oxygenated hemoglobin (oxy-HB) concentration in the bilateral dorsolateral PFC (dlPFC). Children with ADHD, by contrast, did not have an increase in oxy-HB concentration in the dlPFC during CPT performance; instead they had a decrease in oxy-HB concentration in the ventrolateral PFC (vlPFC). When the ADHD children were reevaluated during CPT performance after six months of atomoxetine treatment, they had the same activation that had occurred in control children in the dlPFC, and there was no longer a decrease in oxy-HB in the vlPFC (Araki et al., 2014).

**Stop-Signal Task**

In the SST (Logan et al., 1984), participants are trained to execute an action, such as pressing a button, in response to a visually presented stimulus. However, on some trials, participants are signaled to withhold this response by an auditory or visual signal that occurs unpredictably. The main RRI outcome measure of the task, the SSRT, is an estimate of the amount of time a participant takes to halt the ongoing action (Logan & Cowan, 1984). Similar to the GNG, the SST is widely used and translational between clinical and preclinical models.

The SST has a high level of content validity because the task measures willingness/ability to withhold a response (Moeller et al., 2001). The SST has moderately high reliability, with test–retest reliability coefficients that range from 0.61 when the first and second assessments were approximately 28 days apart (Wöstmann et al., 2013) to 0.65 when the mean time between assessments was 8.6 days (Weafer et al., 2013). In addition, the SST has strong concurrent validity, as performance on the SST correlates with that on the GNG, and strong discriminant validity, as it does not correlate with measures of choice impulsivity or risk taking (Reynolds et al., 2006). Taken together, the SST’s reliability and content, discriminant, and concurrent validity provide support for strong construct validity.

SST performance is sensitive to pharmacological and contextual manipulations such as smoking abstinence and sleep deprivation (Ashare & Hawk, 2012; Sagaspe, Philip, & Schwartz, 2007). Although the SST and GNG both measure RRI, the SST is differentiated from the GNG by a long history of basic parametric manipulations (Alderson, Rapport, & Kofler, 2007; Huizenga, van Bers, Plat, van den Wildenberg, & van der Molen, 2009). Variations of the SST are well validated in children (Devaney et al., 2012; Nederkoorn, Coelho, Guerrieri, Houben, & Jansen, 2012), and the neural substrates of SST processes have been characterized in fMRI studies (Bednarški et al., 2012; Spunt, Lieberman, Cohen, & Eisenberger, 2012).

The SST is limited by possible strategic compensations participants may adopt (e.g., delaying the Go response), and this can impact the reaction time (RT) distribution in unintended ways and produce unwanted noise or bias in the calculation of the SSRT. This necessitates careful training of experimenters and participants, as well as screening of data prior to analysis.

In a meta-analysis of psychiatric research using the SST, the effect sizes for higher response impulsivity on the SST were 0.62 in ADHD, 0.69 in schizophrenia, 0.39 in addiction (Lipszyc & Schachar, 2010). Taking these results together with the results of the GNG and CPT meta-analysis by the same group (Lipszyc & Schachar, 2010; Wright et al., 2014), the authors concluded that...
ADHD is characterized by a more pervasive deficit in response inhibition, with high levels of both SOA and RAI. By contrast, schizophrenia was characterized by a greater deficit in SOA than RAI, and bipolar disorder was characterized by a deficit in RAI only (Wright et al., 2014).

Although several studies support the conclusion that individuals with bipolar disorder do not have a deficit in SOA when compared with healthy control subjects (Deveney et al., 2012; Pavuluri, Ellis, Wegbreit, Passarotti, & Stevens, 2012; Strakowski et al., 2008; Weathers et al., 2012), there were differences in neural activation during SST performance in each of these studies. For example, adults and children with bipolar disorder were characterized by less activation than healthy control participants in the right nucleus accumbens and left ventral PFC during successful inhibition (Weathers et al., 2012). In another study, children with bipolar disorder had less activation than healthy controls in the right nucleus accumbens during inhibition failures (Deveney et al., 2012). When comparing neural activation in adults and children with bipolar disorder, there was an interaction during inhibition failures in the anterior cingulate cortex, with children having less activation and adults having greater activation compared with age-matched healthy controls (Weathers et al., 2012). Therefore, even in the absence of behavioral deficits, differences in the neural activation underlying SST performance may be valuable for understanding bipolar disorder.

**Antisaccade**

The antisaccade task measures eye movements while participants follow instructions to look away from a target. RT and antisaccade errors (i.e., failure to look away from the target or resist distracter interference) are the primary outcome measures, with antisaccade errors indexing RRI. The task has translational value given its use in humans and nonhuman primates (Valero-Cabre et al., 2012), and has little error variance. The antisaccade task does not elicit an anxiety response and is less susceptible to motivation effects than the conventional GNG. The task has established sensitivity to manipulations, including methylphenidate and nicotine consumption (Dawkins, Powell, Pickering, Powell, & West, 2009) and can be employed throughout the life span.

The antisaccade task has strong concurrent validity (Spinella, 2004) but relatively poor discriminant validity, with task scores correlating with self-report measures of related constructs (e.g., Boredom Susceptibility on the Sensation Seeking Scale) (Pettiford et al., 2007) and symptom severity in psychiatric conditions, including autism (Mosconi et al., 2009) and schizophrenia (Turetsky et al., 2007). Test–retest reliability when assessed in a sample of children, adolescents, and young adults with a mean of 18.9 months between assessments was moderate, with a reliability coefficient of 0.48 (Klein & Fischer, 2005). When considered together, the poor discriminant validity and moderate test–retest reliability of the antisaccade indicate that the task has limited construct validity.

Neural regions subserving antisaccade performance are more regionally localized than those involved in GNG or SST performance, which can be considered both an attribute and a limitation of the task with regard to basic processes and work with special populations. In addition to its reduced construct validity, the antisaccade is limited by high cost of the required apparatus, lack of standardization, and susceptibility to individual differences in vision.

In addition, it should be noted that the antisaccade task involves eye movements rather than hand movements, eliciting patterns of neural activation in regions that are specific to the task. In neuroimaging research, the frontal eye field and the supplementary eye field interacted with the ventrolateral PFC during response inhibition on the antisaccade task (Heinen, Rowland, Lee, & Wade, 2006), and a lesion study with frontal lobe patients provided support for these findings (Hodgson et al., 2007). Because the areas of activation elicited by the antisaccade task do not overlap with other response inhibition tasks, caution is warranted when comparing results across tasks.

Of all psychiatric disorders, antisaccades have been most widely studied among people with schizophrenia, who have higher levels of anticipatory antisaccade errors than do healthy control subjects (Hutton & Ettinger, 2006; Turetsky et al., 2007). Compared with the large body of antisaccade research in schizophrenia, there are relatively few studies in which antisaccade performance is examined in other disorders. Some research has reported higher levels of antisaccade errors in individuals with bipolar disorder compared with healthy control participants (Gooding & Tallent, 2001; Katsanis, Kortenkamp, Iacono, & Grove, 1997). However, research also has indicated that performance is not temporally stable in individuals with bipolar disorder, suggesting that antisaccade deficits may be a state, rather than a trait, marker of bipolar disorder (Gooding, Mohapatra, & Shea, 2004; Hutton & Ettinger, 2006).

**Other Tasks**

Other domains may be associated with impulsivity or aspects thereof. Tasks assessing such domains include risk-taking on the Balloon Analogue Risk Task; cognitive control on the Simon, Flanker, and Stroop tasks; decision making on the Iowa gambling task; and perseveration on the intradimensional/extradimensional set-shifting task. These tasks may involve assessments of RTs that might link to or correlate with measures of RRI. However, these tasks were designed to measure processes other than RRI and should not be interpreted as assessing RRI.

**Practical Considerations**

Characteristics and capabilities of research populations in which RRI measures are employed can differ greatly from the population in which the RRI measures were developed, and cognitive capacity and levels of RRI can vary significantly across different age groups and research populations (Butler & Zacks, 2006; Luna, Padmanabhan, & O’Hearn, 2010; Williams, Ponesse, Schachar, Logan, & Tannock, 1999), raising questions about interpretability of RRI task performance. Therefore, careful consideration of the characteristics of the study population and the features of a particular RRI measure is warranted. Using a RRI measure in a new population requires consideration and/or evaluation of the performance range of RRI scores and consideration of other relevant performance variables (e.g., attention span, motivation) that may influence the interpretability of a given performance.
Distributional Issues

Relative to other types of behavioral impulsivity measures (e.g., choice), RRI measures are fairly sensitive to bias relating to a mismatch between task difficulty and participant capability, which can result in nonnormal (skewed) distributions of scores that may lead to floor/ceiling effects. This effect can be observed in studies that report a very low mean impulsivity score (near zero) or large variance in scores. Typically, such studies fail to find group or treatment differences except in cases in which the standard deviation of scores exceeds that of the mean. Difficulties introduced by floor effects and positively skewed data include (a) compromise in the robustness of parametric tests occurring with unequal variance and sample size, as well as outliers; (b) nonnormality in the residuals of general linear model tests; and (c) problems with nonnormality in the residuals of general linear model tests; and (c) problems with unequal sample size, unequal variance between groups, and leverage (Sawilowsky & Blair, 1992; Stonehouse & Forrester, 1998). As a result, both Type I and Type II errors, decreased power, and threats to assumptions of parametric tests are introduced when task parameters unsuitable to the population produce aberrant data distributions.

Determining whether a task has been used previously in comparable populations is important when selecting RRI tasks. If it has been used previously, researchers should determine the performance range of RRI scores and whether the central tendency nears a floor-level effect. Specifically, whether the scoring range allows for the detection of divergence between groups and/or improvement or worsening of performance in response to treatment should be determined. If researchers plan an intervention requiring repeated-measures testing, determination of whether the test in question provides acceptable test–retest reliability is important.

Avoiding problems with performance range may be achieved by selecting RRI procedures that allow for adjustment of task parameters to match participants’ capabilities. Some procedures offer manual manipulation of task parameters (e.g., interstimulus presentation interval, onset of stop signal), although such changes may require pilot testing. An alternative approach is to use a trial-by-trial adjustment procedure to titrate the task difficulty to individuals’ level of performance. This approach enables the same basic task to be used with a range of populations varying in their inhibitory capacity. It should be noted, however, that the adjustment procedure requires careful screening of individual participants for convergence. Further, adjusting task parameters to participants’ capabilities may limit comparability of scores across and within experiments, introducing measurement complexity and reducing generalizability. In summary, thoughtful selection of RRI tasks or task parameters can minimize the threat of obtaining nonnormal, skewed distributions in RRI performance.

With respect to preclinical studies, similar considerations should be taken into account, particularly with respect to animal models of human conditions. Through careful consideration of the clinical characteristics of individuals with the disorders being modeled and the manipulations employed to mimic the conditions, one might select appropriate tasks and employ relevant modifications. Such efforts should aim toward harmonizing assessments across human and nonhuman investigations in order to facilitate comparisons and integration of findings.

Evaluating Interpretability

RRI measures usually record additional data that are useful to evaluate the interpretability of test performance, such as latency to initiate a response and an overall measure of discriminability and accuracy, such as the signal detection parameter \( d' \) (Gescheider, 1985). Prior to analyses of impulsive performance, there should be an evaluation of interpretability of performance. This should involve inspection of other response variables, beyond RRI scores, to determine whether they fall within a range of performance that reflects effortful performance. This requires establishing exclusion criteria for outlier performance. There are several approaches to establishing these criteria: (a) excluding based on below-chance target responses; (b) excluding based on cutoff thresholds reported in previous reports; or (c) developing local norms that are specific to the target population, which typically requires a large number of cases to allow for the evaluation of the distribution of performance and identification of outliers. Examination of response data should take place prior to data analysis, and exclusion criteria and rates should be reported in the publication of these findings.

Summary and Discussion

RRI has been implicated in major public health problems including PDs, substance and nonsubstance addictions, impulsive aggression, ADHD, and suicide. The multiple models of RRI and tools for its assessment may lead to conceptual confusion and difficulties in making comparisons across studies. A consilience of RRI concepts and methods would enhance understanding of the construct, improve collaboration among RRI researchers from diverse disciplines, and move the field forward. Tasks measuring two conceptually and neurobiologically distinct types of RRI, RAI (e.g., GNG, CPT, 5-CSRTT), and SOA (e.g., SST), may capture distinct aspects of the construct, each of which is encompassed by the biopsychosocial definition of RRI (Moeller et al., 2001), and may relate to distinct clinical outcomes. Important differences between SOA and RAI warrant consideration in the selection of the most appropriate assessment for addressing specific research questions. The InSRI group recommends that researchers use both types of tasks (RAI, SOA) in each RRI study. Use of both types of measures in the same study will allow for direct comparisons between the two types of RRI, and will allow associations among each type of RRI with various outcome measures to be examined. In clinical research, this might be achieved by using SST with CPT or GNG tasks to assess RRI. In preclinical research, training considerations may limit the feasibility of administering more than one task to the same animal subject, although this has been accomplished by some groups (Broos et al., 2012). Generally, preclinical researchers can include both types of tasks by conducting multiple-group experiments with subjects from the same species that each use one of the types of tasks to assess RRI. Additionally, preclinical research should be mindful of the conditions being modeled in animals, take into account any task modifications that might relate importantly to the condition being modeled or biological manipulations being employed, and aim toward harmonizing measures across human and nonhuman studies in manners consistent with research domain criteria (Insel et al., 2010), PhenX (C. M. Hamilton et al., 2011; https://www.phenxtoolkit.org), and other initiatives.
In addition to its enhancement of research capabilities, the InSRI construct will have important clinical implications. This is certainly the case with aspects of psychiatric conditions (e.g., bipolar disorder, addictive disorders) and also is true for PDs such as borderline and antisocial PDs, in which aspects of impulsivity play a core role in diagnosis. The high rates of comorbidity across these psychiatric disorders emphasize RRI as a central endophenotype in several models of psychiatric disease.

Despite clear consensus that impulsivity (broadly defined) is a key feature of several psychiatric disorders, the subtle distinctions among facets of the multidimensional construct, as detailed herein, often are not addressed. Given the current prevailing perspective that PDs are not immutable conditions with treatment and even change can be facilitated with a clear and measurable definition of impulsivity.

References


