Disentangling cognitive processes in externalizing psychopathology using drift diffusion modeling: Antagonism, but not Disinhibition, is associated with poor cognitive control

Nathan T. Hall[1,2], Alison M. Schreiber[1,2], Timothy A. Allen[3], Michael N. Hallquist[1,2]

[1] Department of Psychology and Neuroscience, University of North Carolina at Chapel Hill

[2] Department of Psychology, The Pennsylvania State University

[3] Department of Psychiatry, University of Pittsburgh

Accepted for publication in the *Journal of Personality.*

Disclosures. The authors have no financial interests to disclose.

Acknowledgments. This work was funded by the National Institutes of Mental Health (R01 MH119399 to MNH; T32-MH016804 and K01-MH123915 to TAA). The funding agency had no role in the design and conduct of the study; the collection, management, analysis, and interpretation of the data; the preparation, review, and approval of the manuscript; or the decision to submit the manuscript for publication.

Corresponding author. Michael Hallquist, Department of Psychology and Neuroscience, University of North Carolina at Chapel Hill. 235 E. Cameron Avenue, Chapel Hill, NC 27599. michael.hallquist@unc.edu.

Abstract

Although externalizing psychopathology has been linked to deficits in cognitive control, the cognitive processes underlying this association are unclear. Here, we provide a theoretical account of how research on cognitive processes can help to integrate and distinguish personality and psychopathology. We then apply this account to connect the two major subcomponents of externalizing, Antagonism and Disinhibition, with specific control processes using a battery of inhibitory control tasks and corresponding computational modeling. Participants (final $N = 104$) completed the flanker, go/no-go, and recent probes tasks, as well as normal and maladaptive personality inventories and measures of psychological distress. We fit participants' task behavior using a hierarchical drift diffusion model (DDM) to decompose their responses into specific cognitive processes. Using multilevel structural equation models, we found that Antagonism was associated with faster RTs on the flanker task and lower accuracy on flanker and go/no-go tasks. These results were complemented by DDM parameter associations: Antagonism was linked to decreased threshold and drift rate parameter estimates in the flanker task and a decreased drift rate on no-go trials. Altogether, our findings indicate that Antagonism is associated with specific impairments in fast (sub-second) inhibitory control processes involved in withholding prepared/prepotent responses and filtering distracting information. Disinhibition and momentary distress, however, were not associated with task performance.

Keywords: antagonism, externalizing, inhibition, cognitive control, drift diffusion model

Major constructs in personality and psychopathology research share a hierarchical structure; elaborating and refining this structure can improve the diagnosis and treatment of mental illness (Kotov et al., 2017). The discovery of a shared nomological network that spans personality and psychopathology has led to increasing interest in how personality can be distinguished from psychopathology and maladaptive personality features (e.g., Bastiaansen et al., 2016). Such questions have long been central to theoretical and empirical work. For example, Theodore Millon argued that personality serves as a unifying framework for understanding the context in which symptoms of psychopathology are expressed (Pincus & Krueger, 2015).

The majority of structural research on personality and psychopathology has examined individual differences using self-report measures that build on a person's subjective experience and semantic self-knowledge (Robinson & Clore, 2002). A smaller, but nevertheless salient, thread in the personality literature has considered the function of traits and the momentary processes by which dispositions influence behavior and experience (e.g. Wright & Kaurin, 2020). Here, we argue that studying cognitive processes involved in personality and psychopathology is an essential component of understanding their shared structure and potentially differentiating them (Robinson, 2004).

The Value of Cognitive Process Research in Distinguishing Personality and Psychopathology

DeYoung (2015) defines personality traits as, "probabilistic descriptions of relatively stable patterns of emotion, motivation, cognition, and behavior, in response to classes of stimuli" (p. 35). Extending this notion, traits describe the central tendency and variability of trait-related states in daily life (Fleeson, 2001). For example, an individual who endorses worry, tension, and

moodiness on a self-report trait measure of Neuroticism likely does so because they frequently experience these states. While this view has received broad support, an integrated account of how personality traits and momentary environmental demands interact to guide behavior is not fully fledged, echoing the age-old person-situation debate (Fleeson, 2004).

A contemporary *cognitive process* account seeks to understand how traits are associated with core mental operations and how individual differences in these mental operations shape the distributions of states that become conceptualized as traits. Relative to a traditional trait perspective, the cognitive process account draws explicitly on lower-level constructs from cognitive neuroscience that describe basic information processing systems such as working memory or cognitive control (Poldrack & Yarkoni, 2016). We propose that the relationship between traits and momentary experiences can be better understood by interrogating processes that unfold relatively quickly (from seconds to minutes) using methods from experimental and cognitive psychology (Allen et al., 2020). Resolving momentary cognitive processes is particularly important for understanding experiences that reflect rapid and potentially problematic reactivity to certain stimuli.

The cognitive process account relies on the use of experimental paradigms to measure targeted cognitive systems, coupled with formal models of decision-making. Relative to traditional data analyses, cognitive models predict trial-to-trial performance within each person by mapping between latent states and behavioral outputs using a set of quantitative parameters (Maia et al., 2017). The latent states represent hypothesized cognitive processes that govern performance on the task, while model parameters control how different processes are combined to produce the observed pattern of behavior. Within this more dynamical view, individual

differences in distinct cognitive operations combine to give rise to trait-relevant behavior, which, when aggregated over time, form a pillar of personality.

How can a cognitive process account help to distinguish personality from psychopathology? Although there are probably many answers to this question, we focus here on one. We assert that adaptive personality functioning depends on the capacity to integrate and flexibly arbitrate among trait-related goals in the moment in order to select actions that are sensitive to the current context and lead to goal-consistent long-term outcomes. We propose that this capacity is closely related to *cognitive control*, a construct rooted in cognitive psychology that refers to processes that support flexible adaptation based on situational factors and goals (Alexander & Brown, 2010).

A personality trait can be conceptualized as modulating components of a self-regulating system that supports the attainment of specific goals by promoting certain actions and evaluating the results of these actions vis-à-vis a goal (DeYoung, 2015). DeYoung and Krueger (2018) argue that psychopathology is characterized by persistent failures to make progress toward important goals, whereas adaptive personality functioning depends on generating strategies that facilitate goal attainment or developing new goals when existing ones are blocked. Extending this idea to cognitive processes, we propose that cognitive control facilitates adaptation to situations in which more automatic, reactive responding is at odds with one's broader goals. In other words, adaptive personality functioning depends on the ability to simulate the outcomes of alternative choices[1] and to adjudicate among those outcomes in ways that optimize hierarchically superordinate goals (i.e., those that organize longer periods of time and may have greater

[1] Although such simulations are likely to be computationally costly, we do not claim that this is necessarily a conscious process. Indeed, goal-directed learning has been observed in many non-human studies (Dolan & Dayan, 2013).

importance in overall subjective wellbeing). In contrast, psychopathology is characterized by a tendency to respond inflexibly in ways that are often poorly adapted to the current situation or that optimize short-term over long-term goals. This distinction aligns with broader neuroscience research on the tendency to shift from flexible goal-directed strategies to habitual and Pavlovian-congruent responses in the face of stress (Arnsten, 2009)[2]. By interrogating decisions that rely on cognitive control, researchers can explore why some individuals successfully adapt their actions to meet superordinate goals while others struggle to attain these goals.

Considering the Role of Cognitive Control in Externalizing Symptoms

Externalizing psychopathology encompasses problems such as substance abuse, antisocial behavior, and oppositionality that are united by a tendency to 'act out' problematic behaviors (Krueger et al., 2005). These behaviors often reflect the execution of automatic responses that may be poorly adapted to the current situation and reveal a preference for immediate over future goals. For externalizing features, the distinction between adaptive and maladaptive functioning often hinges on whether problematic behaviors are enacted or suppressed (cf. nonaffective constraint; Depue & Collins, 1999). For example, physical aggression toward a rival is an unconditioned response (Domjan, 2005), yet athletes routinely refrain (though not always) from attacking one another during sporting events to avoid being ejected from the game. Such decisions reflect the ability to pursue a superordinate goal (trying to win the game) in part by suppressing goal-incongruent actions that would be more immediately reinforcing. This perspective is agnostic about the substantive content of one's goals, which relate to the configuration of personality traits and the goals those traits promote. That is,

[2] We further note that our perspective aligns with broader thinking in behavioral economics (e.g., system 1 versus system 2 in Kahneman, 2013), cognitive neuroscience (e.g., dual-systems models; Shulman et al., 2016) and learning theory (i.e., goal-directed learning versus habitual responding; Dolan & Dayan, 2013), but a detailed treatment of these convergences is beyond the scope of this paper.

although cognitive control has particular relevance to externalizing symptoms, we view it as a broader resource that supports adaptation in situations where effortfully remapping the value of alternative actions is needed to achieve long-term goals.

Cognitive control emerges in the first few years of life and matures gradually from childhood through early adulthood (Luna et al., 2015; Rothbart et al., 2000). In the temperament literature, cognitive control maps closely to individual differences in Effortful Control, a dispositional feature that reflects the ability to self-regulate via volitional control (Nigg, 2017). Over development, Effortful Control becomes increasingly differentiated, giving rise to the two adult personality traits most commonly implicated in cognitive control: Agreeableness and Conscientiousness/Constraint (Rothbart & Ahadi, 1994). Importantly, the low poles of these traits, often labeled *Antagonism* and *Disinhibition*, respectively, comprise the major subcomponents of externalizing psychopathology, differentially predicting antisocial behavior and substance use (Kotov et al., 2017).

Alterations in cognitive control have long been considered a hallmark of the externalizing spectrum (Hall et al., 2007; Meehan et al., 2013). For instance, externalizing problems have been linked to a reduction in the amplitude of several event-related potentials including the P3 and event-related negativity (ERN) during cognitive control tasks (Hall et al., 2007). Likewise, externalizing proneness is negatively associated with task-based indices of cognitive control both concurrently and prospectively, though the relationship is stronger for measures of response inhibition than set-shifting or working memory (Young et al., 2009).

Some studies have found preliminary evidence of specific relationships between cognitive control and subcomponents of externalizing psychopathology. For instance, Antagonism and Disinhibition are both associated with task-based cognitive control performance

(Fossati et al., 2018; Jensen-Campbell et al., 2002), but the pattern of effects differs between the two traits. In one study, Agreeableness alone was positively associated with performance on a response inhibition task, whereas Conscientiousness was more closely linked to resisting distractions and staying on task (Jensen-Campbell et al., 2002). Others have found that, independent of general externalizing problems, Antagonism-related traits were associated with greater behavioral adjustment following an error, as indicated by post-error slowing in reaction times (McDonald et al., 2019; Bresin et al., 2014).

Overall, both developmental and experimental research implicate cognitive control in Agreeableness/Antagonism and Conscientiousness/Disinhibition, with individual differences in cognitive control likely playing a role in where individuals rank on each dimension. At the same time, we propose that deficits in cognitive control are a necessary, but not sufficient, condition for externalizing psychopathology. Some individuals with low cognitive control may select or be selected into environments that allow them to thrive despite their tendencies to heed short-term impulses. These individuals are best characterized as low on Agreeableness and/or Conscientiousness as opposed to Antagonistic or Disinhibited, as these latter maladaptive terms should be accompanied by evidence of broad goal dysfunction. Thus, to understand psychopathology, we must understand both the mechanisms that drive variation in personality dimensions (DeYoung & Krueger, 2018), as well as the contextual variables that connect such variation with goal attainment and failure. In this paper, we focus primarily on the former topic, though we highlight potential ways for addressing the latter in our discussion.

Computational cognitive models of inhibitory control and the drift diffusion model

While concepts such as executive function and cognitive control are often used monolithically, it is important to understand that different aspects of cognitive control unfold on

different time scales, involve different cognitive systems, and may even show discriminant relations with personality and psychopathology (Nigg, 2017). In the current study, we focus on inhibitory control, which is a narrower form of cognitive control that is immediately deployed (usually at a sub-second time resolution) when making decisions in the face of conflicting information. Inhibitory control involves suppressing specific sources of information or dominant action tendencies in order to complete a goal (Nigg, 2000, 2017). Extending on Nigg (2000), our study focused on three forms of inhibitory control: interference control ("preventing interference due to stimulus competition"), cognitive inhibition ("suppressing nonpertinent ideation to protect working memory/attention"), and behavioral inhibition ("suppressing a prepotent [automatic/prepared/cued] response"). By focusing on three distinct forms of inhibitory control, we leave open the question of whether control deficits in externalizing psychopathology are specific or domain-general.

Studies of inhibitory control typically involve one or more tasks in which subjects make timed responses to stimuli that are presented on a computer monitor. For example, in a simple go/no-go task, participants press a button as quickly and accurately as possible when a range of stimuli are presented, yet on a small number of trials a special stimulus denotes that participants should *not* press the button (i.e. a no-go trial). In many inhibitory control tasks, there is a speed-accuracy tradeoff such that responding accurately comes at the expense of responding quickly and vice versa. Thus, in these speeded control tasks, reaction times (RTs) and accuracy statistics are fundamentally intermixed and likely arise from shared cognitive processes involved in accumulating information about the correct response on a given trial. Furthermore, trial-to-trial variability in responses may reflect experimental factors (e.g., go vs no-go trials) or individual difference variables (e.g., levels of Antagonism and Disinhibition). Aggregating performance to

average RT and accuracy per subject discards this variability and precludes an analysis that would consider how RT and accuracy are manifest outcomes of shared cognitive processes.

To overcome such difficulties, computational cognitive models provide a mathematical formalism that articulates both the hypothesized cognitive processes and their mapping to predicted behavior on relevant tasks. Often, multiple competing models are specified and fit to empirical data in order to examine which model is most consistent with the pattern of human behavior (Hallquist & Dombrovski, 2020). In the context of inhibitory control tasks, sequential sampling models (Ratcliff & Smith, 2004) have a rich history in cognitive and mathematical psychology. These models seek to describe the unfolding of accuracy and RTs on decision tasks involving many trials, in which participants are choosing between two or more alternatives.

The DDM is perhaps the best-established sequential sampling model (Ratcliff et al., 2016; Ratcliff & McKoon, 2008) and conceives of choices on cognitive tasks in terms of noisy accumulation of evidence in favor of one choice or another during a given trial. The rate at which one accumulates evidence towards deciding is referred to as the drift rate (v) and depends on the consistency and salience of information about the alternative choices, as well as the cognitive efficiency of processing this information. Another parameter, called decision-boundary or threshold (a) determines the amount of evidence that is needed to execute a response and is fundamentally related to the speed-accuracy tradeoff (i.e., a lower boundary leads to faster, more inaccurate responding; Bogacz et al., 2010). When the diffusion process crosses the threshold, an RT is produced. Additional parameters, including non-decision-time (t) and bias (z), control the amount of time devoted to decision-irrelevant processing (sensory, encoding, and motor components of choice) and the starting point of the diffusion process, respectively. Taken together, individual differences in these parameters interact with task demands to generate

plausible distributions of accuracy and RTs across a range of cognitive tasks, thus providing a predictive model of inhibitory control.

Current study: linking Antagonism and Disinhibition with inhibitory control

Here, we characterized how individual differences in inhibitory control processes may be relevant to disadvantageous decisions associated with the two traits most central to externalizing psychopathology, Disinhibition and Antagonism. To describe links between traits (Disinhibition and Antagonism) and performance variables (accuracy and RT) across a battery of cognitive control tasks, we first examined how traits moderated individual differences in trial-level effects (e.g., slower RTs on incongruent trials) and subject-level effects (e.g., average accuracy) using a multilevel structural equation modeling (MSEM) approach. We then used the DDM to identify specific cognitive processes that may be differentially associated with these traits. For example, do more disinhibited participants make more errors on inhibitory control tasks due to a lower decision boundary (a) or because of altered evidence accumulation during a noisy decision process (v)? Might highly antagonistic individuals also exhibit the same deficit, suggesting an association with generalized externalizing problems?

The overarching goal of our study was to examine the role cognitive control in externalizing traits, which are linked theoretically and empirically with psychopathology. By studying mechanisms of inhibitory control that unfold in the moment, this study examines how people effortfully adapt to environmental demands and prioritize competing interests to meet their goals. Insight into this process holds particular promise in understanding the functioning of externalizing psychopathology and its constituent traits, since externalizing regularly involves a failure to correctly prioritize one's goals in the moment. We were further interested in exploring the degree to which individual differences in cognitive processes (operationalized as dimensional

estimates of DDM parameters) were differentially predictive of maladaptive traits versus symptoms of psychopathology.

Method

Participants

Participants in the full sample were 112 undergraduate students (69 female, 43 male) enrolled in psychology courses at a northeastern university who completed the experiment in exchange for course credit. The average age of participants was 19.22 (SD = 1.82). Ethnic composition of the sample was 68% Caucasian, 17% Asian, 6% African- American, 6% Latino, and 3% Other. Participants provided informed consent prior to participation in this study.

Materials

Personality Assessment

Personality was assessed via self-reports on the Schedule for Nonadaptive and Adaptive Personality – Second Edition (SNAP-2) and the Multidimensional Personality Questionnaire – Short Form (MPQ-SF). The SNAP-2 (Clark et al., 2008) consists of 390 true-false items that tap a broad range of dysfunctional personality traits. The MPQ-SF (Patrick et al., 2002; Tellegen & Waller, 2008) is a 155-item variant of the MPQ that performs comparably well to the full inventory (Patrick et al., 2002). Scales assessing Antagonism and Disinhibition were selected from the SNAP-2 and MPQ-SF based on Markon et al., (2005). Specifically, we used the subscales of each measure that loaded highest onto what the Markon and colleagues referred to as “Disagreeable Disinhibition” (i.e. Antagonism) and “Unconscientious Disinhibition" (i.e. Disinhibition). For Antagonism, these included the Aggression, Harm Avoidance, and Alienation subscales from the MPQ, as well as Aggression, Manipulativeness, and Mistrust subscales from the SNAP. For Disinhibition, we used the Control, Traditionalism, and Achievement subscales

from the MPQ, as well as Impulsivity, Propriety, and Workaholism scales from the SNAP. Cronbach's alphas for all trait scales were in the acceptable range (all α's > .76).

Psychological Distress

Nonspecific psychological distress, particularly current anxiety and depression, was measured by the Kessler Psychological Distress Scale (K10), a 10-item self- report measure developed to screen for current distress (Kessler et al., 2002). Items on the K10 assess anxiety and depressed mood symptoms in the last 30 days and are rated on a 5-point Likert-type scale (α = .86). State anxiety was measured by the State Anxiety Scale from the State-Trait Anxiety Inventory (STAI-S; Spielberger, 1983). The STAI-S is a well-validated 20-item self-report instrument assessing current anxiety, rated on a 4-point Likert-type scale (α = .94).

Interference Control: Flanker Task

As a measure of interference control, we used the Eriksen flanker task (Eriksen & Eriksen, 1974; Posner et al., 2002). In each trial, participants saw five horizontal arrows and were told to press a key corresponding to the direction of the center arrow (left or right). Participants were instructed to respond as quickly and accurately as possible. Half of the trials presented arrows on either side of the center arrow that pointed in the same direction ("congruent" trials), whereas the flanking arrows on the other trials pointed in the opposite direction ("incongruent" trials). We used an adapted version of the flanker task study in which the frequency of incongruent trials was manipulated across conditions (Casey et al., 2000). Participants completed 160 trials in two conditions: mostly congruent (70% of trials congruent) and mostly incongruent (70% of trials incongruent). Each condition was split into blocks of 40 trials. The direction of the central arrow was counterbalanced across trials. Four blocks of forty trials were presented in ABBA order (Casey et al., 2000), where A is a mostly congruent block

and B is a mostly incongruent block. Stimuli were displayed for 1000ms each with a 500ms inter-trial interval (ITI). One subject in the final sample was missing data for this task due to experiment malfunction.

Behavioral Inhibition: Go/No-Go Task

Participants also completed a modified go/no-go task as a measure of behavioral inhibition (Durston et al., 2002). For this task, participants viewed single letters on each trial and were required to press a key (i.e., "go"; trial on which letters other than 'X' presented) or to withhold a key press (i.e., "no-go"; trial on which 'X' presented). Participants completed three blocks of 64 trials. Letters were presented for 500ms followed by a 1500ms ITI. The frequency of no-go trials was fixed at 20% within a block in order to promote a tendency to perform a key press. In line with Durston et al. (2002), the number of go trials preceding a no-go trial was manipulated to create varying levels of difficulty. No-go trials were preceded by one, three, five, or seven go trials, and inhibitory difficulty was thought to increase with a greater number of preceding go trials. Trials of varying difficulty were randomized within blocks.

Cognitive Inhibition: Recent Probes Task

Finally, participants completed the recent probes task (Nelson et al., 2003). This task measured the ability to inhibit the effects of proactive interference (i.e., difficulty remembering a set of information because of information retained previously) and interference due to making particular responses on previous trials. For each trial, participants viewed a set of four lower case letters for 1500ms, which they were instructed to remember. A 3000ms retention interval followed, then a 1500ms single letter probe was presented, followed by a 2000ms ITI before the next trial. The probe letter matched one of the four letters to be remembered on 50% of the trials (positive trials) and did not match any of the letters on 50% of the trials (negative trials).

Participants were asked to indicate as quickly as possible whether the probe letter was or was not part of the set of to-be-remembered letters. For negative trials, four types of single-letter probes were possible, each comprising 12.5% of the total task. Negative, unfamiliar probe letters were not present on either of the two preceding trials. Negative, familiar probes were present on the preceding trial, but not two trials prior. Negative, highly familiar probes were present on both of the two preceding trials. Negative, response-conflict probes were positive probes from the previous trial. Participants completed 72 trials total in two blocks.

Procedure

Participants were run individually by a single experimenter in one 120-minute session. Participants were asked to refrain from consumption of alcohol (verified via a breathlyzer test) and illicit drugs for at least 24 hours, as well as caffeine and nicotine for 8 hours. All tasks were administered using a 17-inch computer display and high-speed USB keyboard (1ms precision), and experimental tasks were controlled using the E-Prime 2.0 software suite. Participants first completed the STAI-S measure in order to record their baseline state anxiety. Self-report and experimental measures were interleaved with one another in four pseudorandomized orders. For each inhibitory control task, participants completed an untimed introduction and were given feedback about their performance in order to ensure that each task was understood properly.

Data Analysis

Latent Variable Modeling of Antagonism and Disinhibition

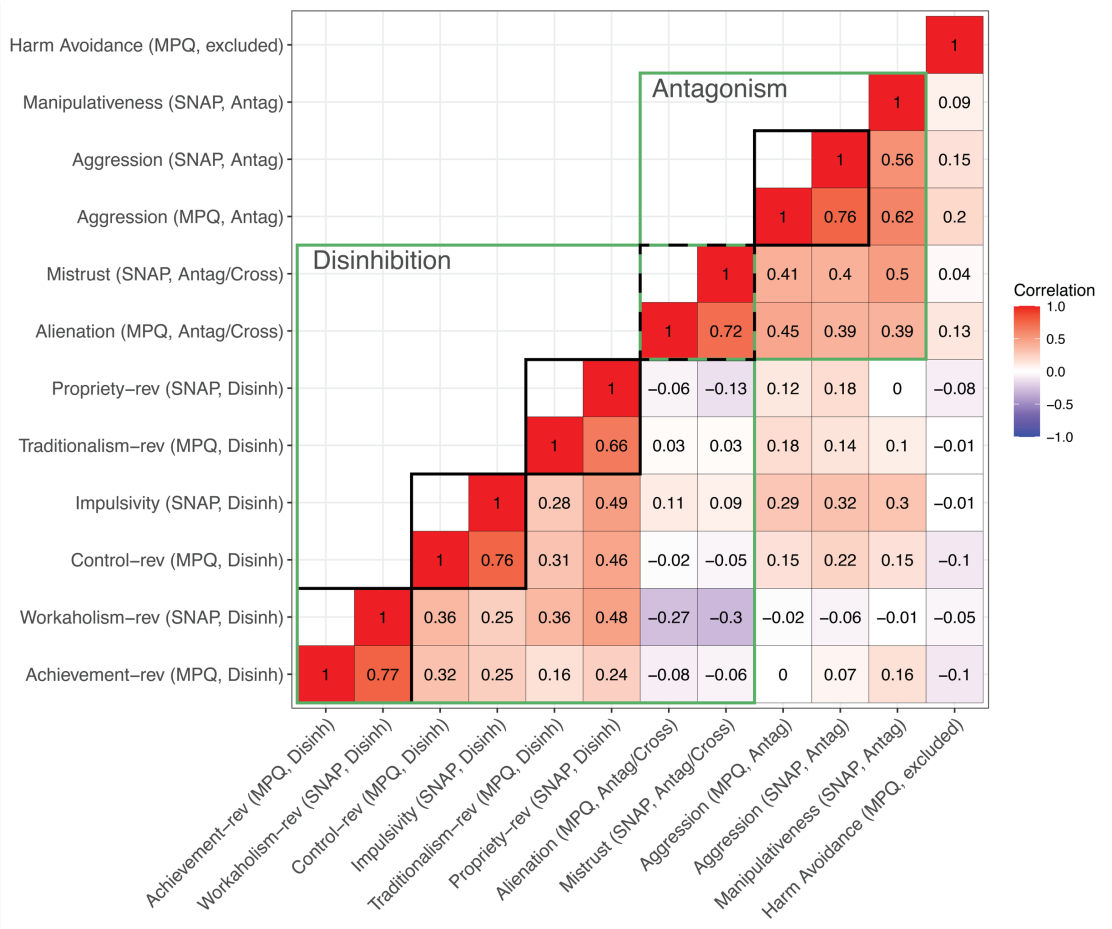

Figure 1 Caption: Bivariate correlations of SNAP and MPQ scales used as indicators of Antagonism and Disinhibition. Scales were selected from (Markon et al., 2005). Solid black boxes denote scales from the SNAP and MPQ that showed highly overlapping content and were thus averaged and fit as parcels. Green boxes denote indicators of latent Antagonism and Disinhibition factors, with modification indices suggesting a significant cross-loading between the Mistrust/Alienation parcel and the Disinhibition factor, to account for the negative association between Mistrust/Alienation and Workaholism.

Eight subjects were dropped from all analyses due to invalid responding on either the SNAP or MPQ, resulting in a final sample of 104 subjects (see supplemental materials for details). Confirmatory factor analyses were used to fit latent Antagonism and Disinhibition indicators. We used five parcels that combined overlapping SNAP and MPQ scale pairs to avoid conceptual redundancy and fit problems (see Figure 1 and supplemental materials), as well as the

Manipulativeness scale from the SNAP. The SNAP and MPQ share a conceptual and psychometric foundation, including at the facet level (Markon et al., 2005).

Model fit for the CFA was determined using established guidelines for the comparative fit index (CFI; close to .95 or above), the root mean square error of approximation (RMSEA; close to .06 or below) and the posterior predictive p-value (Hu & Bentler, 1999). Modification indices were used to assess sources of misfit and guide model adjustments.

Multilevel models of reaction and accuracy on cognitive tasks

Prior to analysis, we removed implausibly fast or slow RTs from consideration (see supplemental materials). We used generalized linear mixed models (GLMMs) to analyze reaction times (RTs) and accuracy. Models of transformed RTs assumed a Gaussian distribution, while accuracies were modeled using logistic GLMMs. We adopted a model-building strategy that progressively allowed for more complicated variance estimates at level 1 (L1; within-person, capturing experimental effects across the entire sample) and level 2 (L2; between-person, capturing individual differences). After identifying the best trial-level model of performance for every task, we incorporated latent variables representing Antagonism and Disinhibition in a multilevel structural equation modeling framework in Mplus 8.4 (Muthén & Muthén, 2017) using Bayesian parameter estimation. In MSEMs, we tested whether personality traits estimated at the between-person level moderated within-subject effects (i.e. cross-level moderation). For a detailed exposition of our RT and accuracy analytic approach, see the supplemental materials. Given the number of significance tests involved in the set of MSEMs, we report the Benjamini-Hochberg false discovery rate corrected *p*-values for each model in the supplemental tables, whereas the uncorrected *p*-values are reported in the text below. We note that FDR correction on the *p*-values did not qualitatively change any of the reported effects.

Hierarchical drift diffusion model estimation and fit evaluation

Drift diffusion models were estimated using the *hddm* Python package (Wiecki et al., 2013). This package leverages the statistical similarities between subjects by simultaneously fitting a group distribution from which individual subjects' parameters are drawn, resulting in more reliable parameter estimates (i.e. Bayesian shrinkage; Ratcliff & Childers, 2015; Wiecki et al., 2013). Models were fit using the *HDDMRegressor* function, allowing for drift rate and threshold to vary as a function of a set of linear predictors. All DDM models minimally allowed drift rate to be influenced by experimental conditions and we used results from the winning GLMMs to guide the addition of further predictors of potential interest in the DDM regression equations (e.g. previous error, trial number, see supplemental materials). HDDM models were fit separately for each task and were estimated with two MCMC chains run in parallel, each consisting of 10,000 draws from the joint posterior, including a 2,000 sample burn-in period. We assessed for model convergence via the Gelman-Rubin $\hat{R}$ statistic (Gelman & Rubin, 1992), and adjudicated among models with the deviance information criterion (DIC; Spiegelhalter et al., 2002). To summarize the direction and strength of relevant predictors, we report the maximum a posteriori probability (MAP) estimate and standard deviation of the distribution of the group posterior from which individual parameters are drawn.

Bayesian distributional regression analyses: relating model parameters to traits and symptoms

We linked inhibitory control processes (operationalized as DDM parameters) with personality traits in a Bayesian distributional multilevel regression framework using the *brms* package in *R* (Bürkner, 2017, 2018). Distributional models in a Bayesian multilevel regression framework allow for the direct incorporation of uncertainty in the single-subject estimates of cognitive processes derived from the DDM (Bürkner, 2018), thus overcoming limitations of the

standard two-step approach, where all subjects are weighted equally (see supplemental materials for more information). The goal of our multivariate Bayesian distributional models was to test if Antagonism and Disinhibition were associated with DDM parameters. These models also included DDM parameters from all three control tasks, allowing us to examine general versus task-specific effects as a function of traits. For each test of interest, we also report the 95% credible interval as an index of the model's uncertainty on the coefficient (Gelman et al., 2013). In secondary analyses, we ensured our results remained the same once accounting for distress (as measured by the K10 and STAI-S).

Results

Factor Analysis

Initial fit of the two-factor CFA of Antagonism and Disinhibition factors was mediocre (CFI = .89, RMSEA = .13, PPp = .06). Based on modification indices, the model was adjusted to allow the parcel composed of SNAP Mistrust and MPQ Alienation to cross-load on Disinhibition[3] (Figure 1). The subsequent model provided an improved fit to the data (CFI = .94, RMSEA = .11, PPp = .19). Plausible factor scores and per-subject factor score standard deviations (posterior SD) from this model were saved for use in subsequent DDM-trait analyses.

Trait Associations with Performance on Cognitive Tasks

Flanker Task

Initial GLMMs (Table S3) identified significant between-subject variation in the effects of trial type, trial number, previous RT, and previous error on the current RT, supporting tests of cross-level moderation of task effects by personality traits (Antagonism and Disinhibition; full

[3] To confirm that modeling this cross-loading did not significantly impact our results we ran identical MSEM analyses linking task performance with Antagonism and Disinhibition, omitting this cross-loading. Regression weights between performance indices and both personality traits were nearly identical across both sets of analyses, suggesting that the cross-loading did not alter the structural relationships between traits and cognitive measures.

model details in Table S6). We found that the effect of trial type on RT was significantly moderated by Antagonism, β = .48, p = .002, such that more antagonistic individuals had larger RT differences between incongruent and congruent trials. At the same time, antagonistic individuals responded more rapidly in general, β = -.31, p = .01 (see Figure 2). Antagonism did not moderate the influence of previous RT (p = .86) or previous error (p = .73) on current RT. Disinhibition was not related to average RT (p = .96) and did not significantly moderate the influences of trial type (p = .70), previous RT (p = .07) or previous error (p = .69) on current RT.

Adopting a similar approach to our RT-trait analyses, we adapted the best-fitting accuracy GLMM model into an MSEM framework to examine the association of personality traits with flanker accuracy. We found that Antagonism was negatively associated with accuracy, β = -.59, p = .002, whereas Disinhibition was not, β = .06, p = .41. Neither Antagonism nor Disinhibition moderated the relationship between trial type and accuracy (β = .16, p = .39, β = .10, p = .63, respectively), previous RT and accuracy (β = -.07, p = .70, β = .12, p = .58, respectively), or previous error and accuracy (β = -.40, p = .39, β = .10, p = .63, respectively).

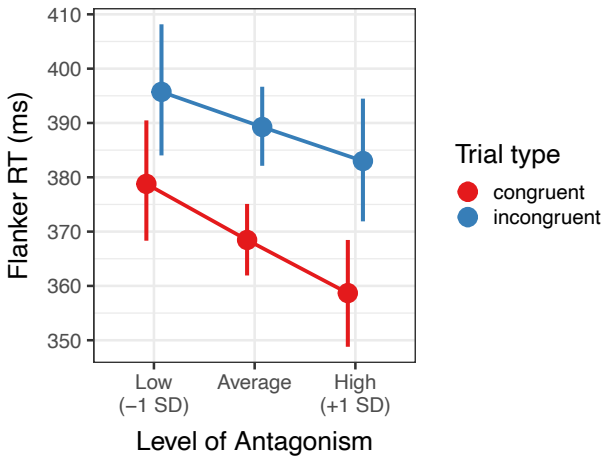

Figure 2 Caption. Association between level of Antagonism and reaction times on the flanker task. The RTs depicted reflect model-predicted RTs from the MSEM that includes effects of trial type, trial number, trial within block, block type, trial type x block type, previous reaction time, and previous error at the

within-person level. At the between-person level, Antagonism and Disinhibition were entered as simultaneous predictors of average RT and cross-level moderators of trial type, previous RT, and previous error. Predicted RTs were computed at low (-1 SD), average (mean), and high (+1 SD) levels of Antagonism, averaging over trial number, trial within block, block type, trial type x block type, and previous reaction time. Circles denote the model-predicted average RT and vertical lines denote the 95% credible interval.

Go/No-Go Task

In our MSEM analysis of RTs on go trials (details in Tables S12-13), Antagonism and Disinhibition did not significantly moderate the effect of the number of trials from last no-go (see supplemental methods; β = -.019, p = .894; β = -.146, p = .334, respectively), trial (β = -.073, p = .592; β = -.104, p = .478, respectively), or mean RT on go trials (β = -.162, p = .186; β = .020, p = .888). However, Antagonism was associated with worse accuracy (β = -.304, p = .034) while Disinhibition was not (β = -.275, p = .080).

Recent Probes Task

In MSEM analyses (details in Tables S18-19), Antagonism and Disinhibition did not significantly moderate the effect of trial (β = .117, p = .194; β = .147, p = .162, respectively) or positive condition (β = .057, p = .362; β = .021, p = .453, respectively) on RT. Further, Antagonism and Disinhibition were not associated with mean RTs (β = -.154, p = .107; β = .070, p = .315, respectively). Likewise, in our accuracy analysis, we did not find that Antagonism or Disinhibition was significantly associated with accuracy (β = -.032, p = .814; β = .108, p = .498, respectively) or trial (β = .095, p = .616; β = -.121, p = .546, respectively).

HDDM model selection

Model selection for each task was based on DIC values across all models tested (see Figures S8-S10). For the flanker task, the best-fitting model included stimulus type, trial number, and previous RT as predictors of drift rate. We found strong evidence that subjects accumulated

evidence quickly towards the correct option on congruent trials (drift rate B = 8.82, $SD[B]$ = 1.62) and that as the task progressed (indexed by the trial number regressor), the drift rate toward the correct option increased (B = 0.49, $SD[B]$ = 0.22). Drift rates toward the correct response (direction of central arrow) on incongruent trials were slower (B = -1.32, $SD[B]$ = 0.81) and slow RTs on one trial predicted reduced drift rates on the following trial (B = -6.66, $SD[B]$ = 1.44). For the go/no-go task, the best fitting model included separate drift rates for go and no-go trials and allowed for threshold to vary as a function of trials from no-go. As expected given the stimulus coding approach (see supplemental material), drift rate was high and positive on go trials (B = 7.12, $SD[B]$ = 0.62) and negative for no-go trials (B = -17.51, $SD[B]$ = 2.26). For the recent probes task, the only significant predictor of drift rate was the experimental condition. In this task, the drift rate for the positive condition (intercept) was M = 1.82, SD = 0.56. Drift rates for the inhibitory control conditions were not significantly different from the positive condition (all posterior parameter distributions included zero), but the drift rate for the negative unfamiliar condition was significantly higher (B = 0.68, $SD[B]$ = 0.11). We report correlations amongst drift rate and threshold parameters in Figure S11.

Relationships between personality traits and DDM parameters: Bayesian multivariate distributional regression

After extracting subject-specific DDM parameter estimates, we entered each parameter as an outcome variable in a multivariate Bayesian distributional regression that included subject-specific standard errors. In this model, personality traits predicted each of the key effects from the DDM models (e.g., drift rate for no-go trials on Go/No Go and threshold for flanker as

separate outcomes). We did not allow for residual correlation among the outcome variables.[4] For the flanker task, we observed that Antagonism was associated with lower drift rates toward the correct choice (B = -.40, 95% CI [-.720, -.068]) and lower threshold, B = -.08, 95% CI [-.16, -.003], but did not moderate the effects of trial and previous RT on drift rate. Conversely, Disinhibition heightened the effect of previous RT on drift rate, B = -.15, 95% CI [-.31, .005], but did not predict drift rate across trials or the effect of trial on drift rate. For Go/No Go, Antagonism predicted slower drift rate (toward the No-Go boundary) for No Go trials, *B* = -.45, 95% CI [-.79, -.11]. However, Antagonism did not reliably predict drift rate for Go Trials or threshold, and Disinhibition was not significantly associated with drift rate or threshold on the Go/No Go task. Furthermore, Antagonism and Disinhibition did not significantly moderate drift rate or threshold in recent probes. These results are summarized in Figure 3.

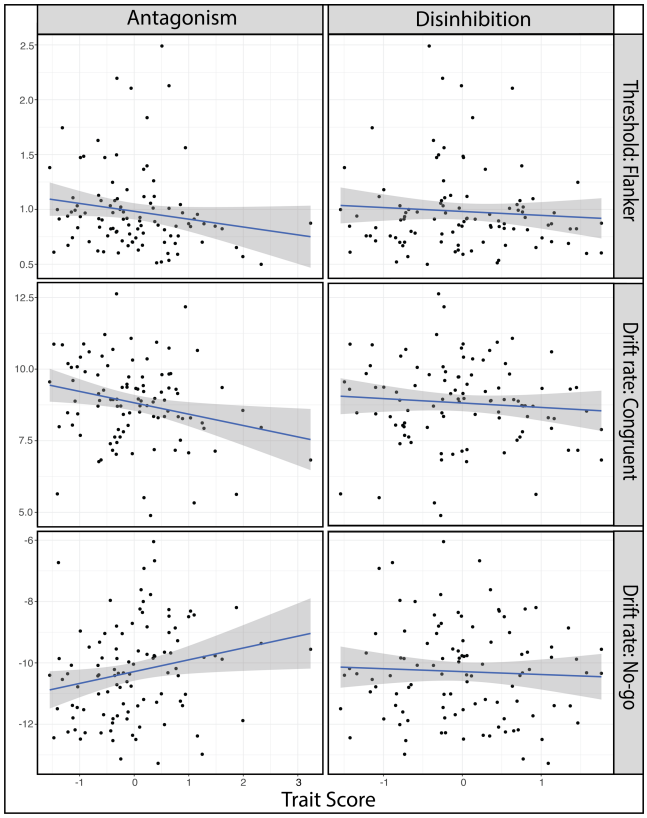

Figure 3 Caption. Scatterplots of significant DDM parameter – Antagonism relationships on the left and accompanying null results for Disinhibition on the right. Importantly, lower drift rate on the flanker task is associated with *slower drift towards the correct option* (regardless of stimulus type), whereas heightened (i.e. "less-negative") drift rate on no-go trials is associated with *slower drift towards the decision to no-go* (see supplemental materials for stimulus vs accuracy coding details).

[4] In a sensitivity analysis, we verified that our results held when allowing for the residual correlation among the outcome variables. Because modeling the residual correlation among outcome variables substantially reduced ESS, we report the estimates from the model in which we did not model the residual correlation among the dependent variables.

Secondary analyses: psychological distress and shared variance across tasks

Finally, we verified that our substantive effects qualitatively held when controlling for psychological distress. In other words, when K10 and STAI were entered as covariates in our distributional models, we did not observe substantive changes in parameter-trait associations (see supplemental materials for detailed results). We also examined whether our effects were attributable to cognitive processes that spanned tasks (e.g. shared variance in drift rate or threshold between tasks) finding that significant trait-parameter associations were focally affected by specific tasks rather than shared across tasks. This suggests that the effects described above hinge upon the specific cognitive demands elicited by different tasks.

Discussion

The objective of our study was to understand the role of cognitive control in Antagonism and Disinhibition, two key facets of externalizing psychopathology. We found that higher Antagonism, but not Disinhibition, was associated with poorer accuracy on two of the three inhibitory control tasks. On the flanker task, we further observed a speed-accuracy tradeoff, such that lower accuracy in more antagonistic people was also associated with faster reaction times. Though each task required response inhibition, they also presented unique demands: the go/no-go task required inhibition of a prepotent motor response (behavioral inhibition), the flanker required suppression of distracting information via attentional filtering (interference control), and recent probes required inhibition of previously encoded information from interfering with current processing (cognitive inhibition).

Turning to the cognitive processes that may underlie inhibitory control, we used hierarchical drift diffusion modeling (HDDM) to examine individual differences in the rate of evidence accumulation (i.e., drift rate) and the level of evidence required to make a response

(i.e., threshold). We found that Antagonism was associated with lower drift rates on (more difficult) no-go trials during the go/no-go task and all trials on the flanker task. Slower evidence accumulation for the no-go choice on no-go trials would lead to more commission errors – choosing to 'go' when one should not have. Lower drift rates on the flanker task indicate that more antagonistic individuals were less efficient in processing the stimulus display, accumulating evidence in favor of the correct option more slowly. Although in isolation, lower drift rates are associated with slower and more variable RTs, the speed-accuracy tradeoff is primary controlled by the threshold parameter (Bogacz et al., 2010). On the flanker task, Antagonism was also associated with a lower threshold, which would give rise to faster, less accurate responding. Combined, the association of Antagonism with lower drift and threshold parameters suggests that antagonistic individuals are less efficient in processing multiple, potentially conflicting stimuli and decide faster on the basis of weaker evidence. On the other hand, Disinhibition was not associated with individual differences in any DDM parameter.

These findings were consistent with our MSEM analyses of accuracy and RTs, which indicated that Antagonism was associated with faster, more inaccurate responding on the flanker task, and with lower accuracy on the flanker and go/no-go tasks. Disinhibition, however, was only marginally associated with lower accuracy on the go/no-go task, but no other performance metrics. Importantly, Antagonism and Disinhibition were positively correlated (r = .34) in our sample, consistent with the idea that their common variance is reflected in a higher-order externalizing factor. Nonetheless, our results suggest that the unique variance associated with Antagonism and Disinhibition may be differentially predictive of distinct control processes that unfold at sub-second resolutions (e.g., under 500ms in our data).

On the whole, our results suggest that Antagonism is associated with a fast-acting, but potentially error-prone, style of processing during tasks that require inhibitory control. There were, however, some differences in the effects across tasks. Although inhibitory control tasks likely have a common foundation (Friedman & Miyake, 2004), they may also tap into different facets of the construct. Following the distinctions outlined by Nigg (2000), we observed evidence of inhibitory control problems on the interference control (flanker) and behavioral inhibition (go/no-go) tasks, but not the cognitive inhibition task (recent probes). This suggests that inhibitory control deficits in more antagonistic individuals may be limited to attentional filtering of conflicting stimuli and suppression of prepotent responses, not problems with manipulation of irrelevant information in working memory. Replicating this pattern of effects in a separate set of tasks, however, will be important to reduce the possibility that our findings are simply due to the 'task impurity problem' (Miyake et al., 2000).

Cognitive control, Disinhibition, and Antagonism: convergence and conflict

Previous studies have reported cognitive control deficits in antagonistic individuals (e.g., Sadeh & Verona, 2008), but the nature of this association has been unclear and inconsistent. General externalizing has been linked to blunted P3 amplitudes in tasks involving novelty detection, interference control, or inhibition of motor responses (Nelson et al., 2011). Externalizing features are also associated with reduced error-related negativity (ERN) responses during tasks that require online adjustment of performance, suggesting that externalizing may be related to poor error processing (Olvet & Hajcak, 2008).

In the psychopathy literature, studies have found that interpersonal-affective features are associated with lower task engagement, but better behavioral adjustment following errors (e.g., Bresin et al., 2014). Such findings can be difficult to interpret, however, given that psychopathy

and its subcomponents reflect a heterogeneous collection of trait constructs that are distributed across multiple domains of the five-factor model (Lynam & Miller, 2015).

In contrast, our study is more firmly anchored in the five-factor model approach to personality. Our results suggest that Antagonism entails inefficient processing of competing stimuli (on the flanker task) or atypical stimuli (no-go stimuli on the go/no-go), as well as a tendency to make a choice quickly before enough information has been obtained (lower threshold on the flanker task, reflecting individual differences in the speed-accuracy tradeoff). These cognitive differences may be, in part, attributable to lower task engagement in antagonistic individuals, consistent with the link between lower P3 amplitudes and callous-unemotional traits (McDonald et al., 2019). Finally, we note that lower drift rates during inhibitory control tasks have been consistently reported in studies of Attention Deficit Hyperactivity Disorder (Huang-Pollock et al., 2017), suggesting an intriguing cognitive link between early disruptive behavior problems and subsequent maladaptive personality (Lewinsohn et al., 1997).

Overall, we did not observe a link between Disinhibition and early inhibitory processes. This null finding is in line with previous ERP studies showing Disinhibition is associated with the late-occurring P3, but not with the earlier-occurring N2 waveform that has been linked to the immediate parsing of conflicting perceptual input (Ribes-Guardiola et al., 2020). One possibility is that Conscientiousness/Disinhibition might be more closely related to components of cognitive control not under investigation here. For instance, previous studies have found that Conscientiousness is positively related to mental set shifting, but not response inhibition or working memory (Fleming et al., 2016). Such a pattern is consistent with recent theoretical accounts linking Conscientiousness to the ability to flexibly plan and prioritize one's goals (Stock & Beste, 2015).

Inhibitory control in Antagonism: trait vulnerability for psychopathology?

Antagonism is a deeply interpersonal trait (with expressions such as social manipulation or obstinacy), so it is noteworthy that our findings link it to cognitive processes unfolding during nonsocial control tasks that seem intuitively distal from interpersonal dynamics. Our results suggest that altered processing of cognitively demanding or conflicting information may be an important component of individual differences in social behavior. This notion is consistent with our broader proposal that cognitive control resources support the rapid integration and flexible arbitration among trait-related goals in order to select actions that lead to goal-consistent long-term outcomes. Though it is conceivable that poor performance on the cognitive tasks was a product of Antagonism (perhaps a desire not to cooperate with the experimenter, or a lack of motivation to perform well), we believe this is unlikely to be the case. We carefully vetted both personality and task behavior to ensure that subjects with systematically poor task performance or invalid personality profiles were excluded from analyses.

Previous research has linked Antagonism to an inability to coordinate one's own goals with those of others in order to navigate complex social contexts (Allen et al., 2017). Social interactions are dynamic, requiring individuals to quickly adapt their behavior to new, potentially conflicting information, and to inhibit their own impulses in order to foster cooperation. Our results suggest that variation in control processes that are deployed quickly (on the order of a few seconds or less) play an important role in determining whether individuals navigate dynamic social contexts with empathy and flexibility, or with callousness and egocentrism.

Personality traits, however, are very unlikely to map one-to-one with single cognitive processes or neural circuits (Allen et al., 2020). Instead, the combination of latent cognitive processes and situational factors interact to produce trait-relevant behaviors. Thus, we warn

against a simplistic view of Antagonism as a trait underpinned purely by impaired inhibitory control. Instead, it is much more likely that in interpersonal situations requiring the utilization of inhibitory control in conjunction with systems implicated in social cognition (e.g., tracking others' intentions or theory of mind), antagonistic individuals may behave more erratically due to impairment in a basic control process. This interpretation is in line with research suggesting that the development of inhibitory control mechanisms in young children is a necessary but not sufficient requirement for the development of theory of mind (Carlson & Moses, 2001). As an example, antagonistic individuals may struggle to resist the emotion-congruent (and myopic) response of yelling at a loafing coworker instead of striving to increase social rapport in the service of building a more accountable work environment.

Importantly, the associations between DDM parameter estimates and self-reported Antagonism held even after controlling for current psychological distress. This suggests that inhibitory control deficits reflect variation in cognitive processing that is specific to Antagonism, as opposed to one's current emotional state. Extreme variation in these cognitive indices may be associated with extreme variability in the traits with which they are associated. As others have noted, however, extremity alone is neither necessary nor sufficient to produce psychopathology (DeYoung & Krueger, 2018). For instance, a lower decision threshold may lead disagreeable individuals to be more decisive, acting quickly even when they have not yet integrated the details of an interpersonal situation. In settings in which errors are not costly, this kind of expediency may be highly valued, and disagreeable individuals may thrive. On the other hand, in settings that demand accurate, deliberate responding, the tendency toward decisional expediency in Antagonism is likely to be counterproductive to one's goals, and thus pathological in nature.

For example, a more disagreeable person who becomes a criminal defense lawyer may be praised for objecting to key points in the courtroom proceedings. However, her status with the judge and jury would surely be reduced if she objected to every point the prosecutor made irrespective of its bearing on the case. Likewise, being as combative in her personal relationships as she is in the courtroom might be disadvantageous for meeting her social goals in close relationships. We note that these conjectures are in line with bivariate correlations in our sample: psychological distress was associated with Antagonism (K10 r = .36 and STAI-S r = .35, respectively) but not with Disinhibition (K10 r = -.02 and STAI-S r = .03, respectively).

Altogether, our results demonstrate that cognitive processes provide a window into the mechanisms by which trait vulnerabilities affect behavior. But this is likely only one part of the story. The extent to which control processes promote or block goal attainment is a key factor in disentangling adaptive personality functioning from psychopathology. Situations that demand cognitive control, particularly inhibitory control, are ubiquitous. During interpersonal interactions, we often must rapidly process complex, conflicting information and short-term impulses to achieve more abstract and integrative goals such as group cohesion or fostering interpersonal harmony. In such situations, a tendency to respond impulsively or egocentrically decreases the probability of succeeding in the pursuit of these complex social goals, though in the moment could increase the probability of achieving short-term goals such as “winning” an argument.

To disentangle personality from psychopathology, it is necessary to understand both the relevant cognitive processes and the circumstances or environments in which these processes promote or inhibit goal attainment. Experimental paradigms that tap into both cognitive control and social cognition may be especially helpful in elaborating and testing this line of thinking.

Paradigms from behavioral economics and decision science such as the ultimatum game have provided useful insights into factors that influence social cooperation (Güth & Kocher, 2014), yet most of these paradigms allow participants to deliberate about each choice, rather than requiring rapid processing of social information. Conversely, simple inhibitory control paradigms such as those presented here require rapid processing of conflicting stimuli, but do not link one's behaviors to outcomes in social contexts. We hope that future research in this area integrates features from both of these experimental traditions to examine how poor inhibitory control, low Agreeableness, and social circumstance interact to produce externalizing psychopathology.

Strengths and Limitations

Importantly, experimental paradigms used to interrogate cognitive processing have been criticized for poor reliability and inconsistent alignment with traditional self-report measures and external criteria (Hedge et al., 2018; Sharma et al., 2014). Such accounts show modest correlations at best that are thought to be caused by low reliability in behavioral indices of impulsivity leading to suppression of between-person variation (Hedge et al., 2018). These problems, however, may largely reflect the use of person-level summary statistics that a) do not retain trial-to-trial variability within persons and b) do not examine individual differences in cognitive processes thought to underlie performance on the task (Haines, Kvam, et al., 2020; Rouder & Haaf, 2019). Recent work on this topic has highlighted the importance of allowing for heterogeneity in experimental effects in data analyses (i.e., relative to simpler ANOVA-style analysis; Bolger et al., 2019) and the value of computational cognitive models that provide testable predictions about the processes that putatively *generate* performance on tasks (Haines, Beauchaine, et al., 2020).

This is the first study that we are aware of to use formal cognitive modeling to assess the processes associated with the two major components of externalizing behavior, Antagonism and Disinhibition, during an inhibitory control task battery. We used the drift diffusion model to estimate individual differences in parameters that reflect the efficiency of cognitive processing, speed-accuracy tradeoff, and stimulus encoding and response preparation time. By parsing these influences into different latent decision processes, the DDM provides more refined measures of the cognitive processes involved in inhibitory control. For example, whereas conventional RT analyses cannot rule out the possibility that trait-related differences may simply reflect slower encoding of the task or motor preparation time, the DDM explicitly parses these out, providing a purer measure of the rate at which an individual accumulates decision-relevant information (i.e., the drift rate). Furthermore, we estimated DDM parameters using a hierarchical Bayesian estimation approach that provides more precise estimates of individual parameters by borrowing strength from the sample. We also used a distributional Bayesian model to relate traits and DDM parameters, thereby propagating uncertainty in the DDM estimates to the trait-DDM associations. Finally, we leveraged multiple conceptually related cognitive tasks (Poldrack & Yarkoni, 2016) to assess for domain-general vs. task-specific modulation of inhibitory control.

The current study also has some noteworthy limitations. First, while our sample is relatively large for a cognitive study, it is modest compared to the broader trait-outcome literature (e.g., Soto, 2019). Second, our sample consisted of unselected young adult participants, most of whom did not have clinically significant levels of psychopathology. Thus, replication of our findings in larger and more diverse samples would demonstrate that our findings scale across the full range of Antagonism and have real-world implications for functioning. Third, our battery was selected to tap into facets of inhibitory control but did not

assess the broader range of processes involved in cognitive control. While this approach sheds new light on cognitive processes associated with Antagonism, our null findings with respect to Disinhibition should not be interpreted as evidence that this trait is unrelated to problems with self-control. Rather, additional paradigms that tap into set-shifting, value-based decision-making, and planning may provide better insights into cognitive processes underlying Disinhibition, as well as corresponding associations with psychopathology (Mukherjee & Kable, 2014).

Conclusion

Deficits in cognitive control have long been associated with externalizing psychopathology. Here, we demonstrated more specifically that Antagonism, but not Disinhibition, is associated with an inefficiency in processing conflicting or rare stimuli, as well as a tendency to prefer speed over accuracy in inhibitory control tasks. Our results point to a potential dysfunction in early-occurring control processes that facilitate inhibiting one's initial tendencies in favor of others' needs and goals. Exerting such control in social interactions may enable agreeable individuals to maintain social harmony and foster cooperation in the service of superordinate goals (such as maintaining healthy relationships). On the other hand, in antagonistic individuals, a limited ability to harness these control processes may undermine social interactions and contribute to erratic and short-sighted behavior.

Acting quickly on one's impulses is unlikely to be sufficient for the development of psychopathology, however. Rather, antagonistic externalizing symptoms likely reflect the cumulative effect of poor cognitive control in with contexts that pit reinforcing, but myopic decisions against long-term social goal attainment. A predisposition toward self-focused and reactive responses in interpersonal exchanges will often culminate in escalating conflicts and difficulties working collaboratively with others. Over time, this pattern may lead to significant

failures in achieving long-term goals, demarcating the inflection point where disagreeable behavior becomes pathological.

References

Alexander, W. H., & Brown, J. W. (2010). Computational Models of Performance Monitoring and Cognitive Control. *Topics in Cognitive Science*, *2*(4), 658–677.

Allen, T. A., Rueter, A. R., Abram, S. V., Brown, J. S., & Deyoung, C. G. (2017). Personality and Neural Correlates of Mentalizing Ability. *European Journal of Personality*, *31*(6), 599–613.

Allen, T. A., Schreiber, A. M., Hall, N. T., & Hallquist, M. N. (2020). From Description to Explanation: Integrating Across Multiple Levels of Analysis to Inform Neuroscientific Accounts of Dimensional Personality Pathology. *Journal of Personality Disorders*, *34*(5), 650–676.

Arnsten, A. F. T. (2009). Stress signalling pathways that impair prefrontal cortex structure and function. *Nature Reviews Neuroscience*, *10*(6), 410–422.

Bastiaansen, L., Hopwood, C. J., Van den Broeck, J., Rossi, G., Schotte, C., & De Fruyt, F. (2016). The twofold diagnosis of personality disorder: How do personality dysfunction and pathological traits increment each other at successive levels of the trait hierarchy? *Personality Disorders: Theory, Research, and Treatment*, *7*(3), 280–292.

Bogacz, R., Hu, P. T., Holmes, P. J., & Cohen, J. D. (2010). Do humans produce the speed-accuracy trade-off that maximizes reward rate? *Quarterly Journal of Experimental Psychology (2006)*, *63*(5), 863–891.

Bolger, N., Zee, K. S., Rossignac-Milon, M., & Hassin, R. R. (2019). Causal processes in psychology are heterogeneous. *Journal of Experimental Psychology: General*, *148*(4), 601–618.

Bresin, K., Finy, M. S., Sprague, J., & Verona, E. (2014). Response monitoring and adjustment: Differential relations with psychopathic traits. *Journal of Abnormal Psychology*, *123*(3), 634–649.

Bürkner, P.-C. (2017). brms: An R Package for Bayesian Multilevel Models Using Stan. *Journal of Statistical Software*, *80*(1), 1–28.

Bürkner, P.-C. (2018). Advanced Bayesian Multilevel Modeling with the R Package brms. *The R Journal*, *10*(1), 395–411.

Carlson, S. M., & Moses, L. J. (2001). Individual Differences in Inhibitory Control and Children's Theory of Mind. *Child Development*, *72*(4), 1032–1053.

Casey, B. J., Thomas, K. M., Welsh, T. F., Badgaiyan, R. D., Eccard, C. H., Jennings, J. R., & Crone, E. A. (2000). Dissociation of response conflict, attentional selection, and expectancy with functional magnetic resonance imaging. *Proceedings of the National Academy of Sciences*, *97*(15), 8728–8733.

Clark, L. A. (1993). *Manual for the Schedule for Nonadaptive and Adaptive Personality (SNAP)*. University of Minnesota Press.

Clark, L. A., Simms, L. J., Wu, K. D., & Casillas, A. (2008). *Manual for the schedule for nonadaptive and adaptive personality (SNAP-2)*.

Depue, R. A., & Collins, P. F. (1999). On the psychobiological complexity and stability of traits. *Behavioral and Brain Sciences*, *22*(3), 541–555.

DeYoung, C. G. (2015). Cybernetic Big Five Theory. *Journal of Research in Personality*, *56*, 33–58.

DeYoung, C. G., & Krueger, R. F. (2018). A Cybernetic Theory of Psychopathology. *Psychological Inquiry*, *29*(3), 117–138.

Dolan, R. J., & Dayan, P. (2013). Goals and Habits in the Brain. *Neuron*, *80*(2), 312–325.

Domjan, M. (2005). Pavlovian Conditioning: A Functional Perspective. *Annual Review of Psychology*, *56*(1), 179–206.

Durston, S., Thomas, K. M., Yang, Y., Uluğ, A. M., Zimmerman, R. D., & Casey, B. J. (2002). A neural basis for the development of inhibitory control. *Developmental Science*, *5*(4), F9–F16.

Eriksen, B. A., & Eriksen, C. W. (1974). Effects of noise letters upon the identification of a target letter in a nonsearch task. *Perception & Psychophysics*, *16*(1), 143–149.

Fleeson, W. (2001). Toward a structure- and process-integrated view of personality: Traits as density distributions of states. *Journal of Personality and Social Psychology*, *80*(6), 1011–1027.

Fleeson, W. (2004). Moving Personality Beyond the Person-Situation Debate: The Challenge and the Opportunity of Within-Person Variability. *Current Directions in Psychological Science*, *13*(2), 83–87.

Fleming, K. A., Heintzelman, S. J., & Bartholow, B. D. (2016). Specifying Associations Between Conscientiousness and Executive Functioning: Mental Set Shifting, Not Prepotent Response Inhibition or Working Memory Updating. *Journal of Personality*, *84*(3), 348–360.

Fossati, A., Somma, A., Borroni, S., Markon, K. E., & Krueger, R. F. (2018). Executive Functioning Correlates of DSM-5 Maladaptive Personality Traits: Initial Evidence from an Italian Sample of Consecutively Admitted Adult Outpatients. *Journal of Psychopathology and Behavioral Assessment*, *40*(3), 484–496.

Friedman, N. P., & Miyake, A. (2004). The Relations Among Inhibition and Interference Control Functions: A Latent-Variable Analysis. *Journal of Experimental Psychology: General*, *133*(1), 101–135.

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2013). *Bayesian Data Analysis, Third Edition*. CRC Press.

Gelman, A., & Rubin, D. B. (1992). Inference from Iterative Simulation Using Multiple Sequences. *Statistical Science*, *7*(4), 457–472.

Güth, W., & Kocher, M. G. (2014). More than thirty years of ultimatum bargaining experiments: Motives, variations, and a survey of the recent literature. *Journal of Economic Behavior & Organization*, *108*, 396–409.

Haines, N., Beauchaine, T. P., Galdo, M., Rogers, A. H., Hahn, H., Pitt, M. A., Myung, J. I., Turner, B. M., & Ahn, W.-Y. (2020). Anxiety Modulates Preference for Immediate Rewards Among Trait-Impulsive Individuals: A Hierarchical Bayesian Analysis. *Clinical Psychological Science*, *8*(6), 1017–1036.

Haines, N., Kvam, P. D., Irving, L. H., Smith, C., Beauchaine, T. P., Pitt, M. A., Ahn, W.-Y., & Turner, B. (2020). *Learning from the Reliability Paradox: How Theoretically Informed Generative Models Can Advance the Social, Behavioral, and Brain Sciences*. PsyArXiv.

Hall, J. R., Bernat, E. M., & Patrick, C. J. (2007). Externalizing Psychopathology and the Error-Related Negativity. *Psychological Science*, *18*(4), 326–333.

Hallquist, M. N., & Dombrovski, A. Y. (2020). Reinforcement Learning Approaches to Computational Clinical Neuroscience. In *Handbook of Research Methods in Clinical Psychology*. Cambridge University Press.

Hedge, C., Powell, G., & Sumner, P. (2018). The reliability paradox: Why robust cognitive tasks do not produce reliable individual differences. *Behavior Research Methods*, *50*(3), 1166–1186.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, *6*(1), 1–55.

Huang-Pollock, C., Ratcliff, R., McKoon, G., Shapiro, Z., Weigard, A., & Galloway-Long, H. (2017). Using the Diffusion Model to Explain Cognitive Deficits in Attention Deficit Hyperactivity Disorder. *Journal of Abnormal Child Psychology*, *45*(1), 57–68.

Jensen-Campbell, L. A., Rosselli, M., Workman, K. A., Santisi, M., Rios, J. D., & Bojan, D. (2002). Agreeableness, conscientiousness, and effortful control processes. *Journal of Research in Personality*, *36*(5), 476–489.

Kessler, R. C., Andrews, G., Colpe, L. J., Hiripi, E., Mroczek, D. K., Normand, S.-L. T., Walters, E. E., & Zaslavsky, A. M. (2002). Short screening scales to monitor population prevalences and trends in non-specific psychological distress. *Psychological Medicine*, *32*(6), 959–976.

Kotov, R., Krueger, R. F., Watson, D., Achenbach, T. M., Althoff, R. R., Bagby, R. M., Brown, T. A., Carpenter, W. T., Caspi, A., Clark, L. A., Eaton, N. R., Forbes, M. K., Forbush, K. T., Goldberg, D., Hasin, D., Hyman, S. E., Ivanova, M. Y., Lynam, D. R., Markon, K., … Zimmerman, M. (2017). The Hierarchical Taxonomy of Psychopathology (HiTOP): A dimensional alternative to traditional nosologies. *Journal of Abnormal Psychology*, *126*(4), 454–477.

Krueger, R. F., Markon, K. E., Patrick, C. J., & Iacono, W. G. (2005). Externalizing Psychopathology in Adulthood: A Dimensional-Spectrum Conceptualization and Its Implications for DSM–V. *Journal of Abnormal Psychology*, *114*(4), 537–550.

Lewinsohn, P. M., Rohde, P., Seeley, J. R., & Klein, D. N. (1997). Axis II psychopathology as a function of Axis I disorders in childhood and adolescence. *Journal of the American Academy of Child & Adolescent Psychiatry*, *36*(12), 1752–1759.

Luna, B., Marek, S., Larsen, B., Tervo-Clemmens, B., & Chahal, R. (2015). An Integrative Model of the Maturation of Cognitive Control. *Annual Review of Neuroscience*, *38*(1), 151–170.

Lynam, D. R., & Miller, J. D. (2015). Psychopathy from a Basic Trait Perspective: The Utility of a Five-Factor Model Approach. *Journal of Personality*, *83*(6), 611–626.

Maia, T. V., Huys, Q. J. M., & Frank, M. J. (2017). Theory-Based Computational Psychiatry. *Biological Psychiatry*, *82*(6), 382–384.

Markon, K. E., Krueger, R. F., & Watson, D. (2005). Delineating the Structure of Normal and Abnormal Personality: An Integrative Hierarchical Approach. *Journal of Personality and Social Psychology*, *88*(1), 139–157.

McDonald, J. B., Bozzay, M. L., Bresin, K., & Verona, E. (2019). Facets of externalizing psychopathology in relation to inhibitory control and error processing. *International Journal of Psychophysiology*.

Meehan, K. B., De Panfilis, C., Cain, N. M., & Clarkin, J. F. (2013). Effortful control and externalizing problems in young adults. *Personality and Individual Differences*, *55*(5), 553–558.

Miller, J. D., Lyman, D. R., Widiger, T. A., & Leukefeld, C. (2001). Personality Disorders as Extreme Variants of Common Personality Dimensions: Can the Five Factor Model Adequately Represent Psychopathy? *Journal of Personality*, *69*(2), 253–276.

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The Unity and Diversity of Executive Functions and Their Contributions to Complex "Frontal Lobe" Tasks: A Latent Variable Analysis. *Cognitive Psychology*, *41*(1), 49–100.

Mukherjee, D., & Kable, J. W. (2014). Value-Based Decision Making in Mental Illness A Meta-Analysis. *Clinical Psychological Science*, 2167702614531580.

Muthén, L. K., & Muthén, B. O. (2017). *Mplus User's Guide (version 8)*. Muthén and Muthén.

Nelson, J. K., Reuter-Lorenz, P. A., Sylvester, C.-Y. C., Jonides, J., & Smith, E. E. (2003). Dissociable neural mechanisms underlying response-based and familiarity-based conflict in working memory. *Proceedings of the National Academy of Sciences*, *100*(19), 11171–11175.

Nelson, L. D., Patrick, C. J., & Bernat, E. M. (2011). Operationalizing proneness to externalizing psychopathology as a multivariate psychophysiological phenotype. *Psychophysiology*, *48*(1), 64–72.

Nigg, J. (2017). Annual Research Review: On the relations among self-regulation, self-control, executive functioning, effortful control, cognitive control, impulsivity, risk-taking, and inhibition for developmental psychopathology. *Journal of Child Psychology and Psychiatry and Allied Disciplines*.

Nigg, J. T. (2000). On inhibition/disinhibition in developmental psychopathology: Views from cognitive and personality psychology and a working inhibition taxonomy. *Psychological Bulletin*, *126*(2), 220–246.

Olvet, D., & Hajcak, G. (2008). The error-related negativity (ERN) and psychopathology: Toward an endophenotype. *Clinical Psychology Review*, *28*(8), 1343–1354.

Patrick, C. J., Curtin, J. J., & Tellegen, A. (2002). Development and validation of a brief form of the Multidimensional Personality Questionnaire. *Psychological Assessment*, *14*(2), 150–163.

Pincus, A. L., & Krueger, R. F. (2015). Theodore Millon's Contributions to Conceptualizing Personality Disorders. *Journal of Personality Assessment*, *97*(6), 537–540.

Poldrack, R. A., & Yarkoni, T. (2016). From Brain Maps to Cognitive Ontologies: Informatics and the Search for Mental Structure. *Annual Review of Psychology*, *67*(1), 587–612.

Posner, M. I., Rothbart, M. K., Vizueta, N., Levy, K. N., Evans, D. E., Thomas, K. M., & Clarkin, J. F. (2002). Attentional mechanisms of borderline personality disorder. *Proceedings of the National Academy of Sciences*, *99*(25), 16366–16370.

Ratcliff, R., & Childers, R. (2015). Individual Differences and Fitting Methods for the Two-Choice Diffusion Model of Decision Making. *Decision*.

Ratcliff, R., & McKoon, G. (2008). The diffusion decision model: Theory and data for two-choice decision tasks. *Neural Computation*, *20*(4), 873–922.

Ratcliff, R., & Smith, P. L. (2004). A Comparison of Sequential Sampling Models for Two-Choice Reaction Time. *Psychological Review*, *111*(2), 333–367.

Ratcliff, R., Smith, P. L., Brown, S. D., & McKoon, G. (2016). Diffusion Decision Model: Current Issues and History. *Trends in Cognitive Sciences*, *20*(4), 260–281.

Ribes-Guardiola, P., Poy, R., Patrick, C. J., & Moltó, J. (2020). Electrocortical measures of performance monitoring from go/no-go and flanker tasks: Differential relations with trait dimensions of the triarchic model of psychopathy. *Psychophysiology*, *57*(6), e13573.

Robinson, M. D. (2004). Personality as Performance: Categorization Tendencies and Their Correlates. *Current Directions in Psychological Science*, *13*(3), 127–129.

Robinson, M. D., & Clore, G. L. (2002). Belief and feeling: Evidence for an accessibility model of emotional self-report. *Psychological Bulletin*, *128*(6), 934–960.

Rothbart, M. K., & Ahadi, S. A. (1994). Temperament and the development of personality. *Journal of Abnormal Psychology*, *103*(1), 55–66.

Rothbart, M. K., Ahadi, S. A., & Evans, D. E. (2000). Temperament and personality: Origins and outcomes. *Journal of Personality and Social Psychology*, *78*(1), 122–135.

Rouder, J. N., & Haaf, J. M. (2019). A psychometrics of individual differences in experimental tasks. *Psychonomic Bulletin & Review*, *26*(2), 452–467.

Sadeh, N., & Verona, E. (2008). Psychopathic Personality Traits Associated with Abnormal Selective Attention and Impaired Cognitive Control. *Neuropsychology*, *22*(5), 669–680.

Sharma, L., Markon, K. E., & Clark, L. A. (2014). Toward a theory of distinct types of "impulsive" behaviors: A meta-analysis of self-report and behavioral measures. *Psychological Bulletin*, *140*(2), 374–408.

Soto, C. J. (2019). How Replicable Are Links Between Personality Traits and Consequential Life Outcomes? The Life Outcomes of Personality Replication Project. *Psychological Science*, *30*(5), 711–727.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., & Van Der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *64*(4), 583–639.

Spielberger, C. D. (1983). *State-Trait Anxiety Inventory for Adults*. Mind Garden.

Stock, A.-K., & Beste, C. (2015). Conscientiousness increases efficiency of multicomponent behavior. *Scientific Reports*, *5*(1), 15731.

Tellegen, A., & Waller, N. G. (2008). Exploring personality through test construction: Development of the Multidimensional Personality Questionnaire. In *The SAGE handbook of personality theory and assessment, Vol 2: Personality measurement and testing* (pp. 261–292). Sage Publications, Inc.

Watts, A. L., Lilienfeld, S. O., Smith, S. F., Miller, J. D., Campbell, W. K., Waldman, I. D., Rubenzer, S. J., & Faschingbauer, T. J. (2013). The double-edged sword of grandiose narcissism: Implications for successful and unsuccessful leadership among U.S. Presidents. *Psychological Science*, *24*(12), 2379–2389.

Wiecki, T. V., Sofer, I., & Frank, M. J. (2013). HDDM: Hierarchical Bayesian estimation of the Drift-Diffusion Model in Python. *Frontiers in Neuroinformatics*, *7*.

Wright, A.G. C., & Kaurin, A. (2020). Integrating Structure and Function in Conceptualizing and Assessing Pathological Traits. *Psychopathology*, 1–9.

Young, S. E., Friedman, N. P., Miyake, A., Willcutt, E. G., Corley, R. P., Haberstick, B. C., & Hewitt, J. K. (2009). Behavioral disinhibition: Liability for externalizing spectrum disorders and its genetic

and environmental relation to response inhibition across adolescence. *Journal of Abnormal Psychology*, *118*(1), 117–130.