Published in final edited form as:

Psychol Med. 2006 November ; 36(11): 1613–1624. doi:10.1017/S0033291706008580.

Reaction time, inhibition, working memory and ‘delay aversion’ performance: genetic influences and their interpretation

JONNA KUNTSI^{1,*}, HANNAH ROGERS¹, GREER SWINARD¹, NORBERT BÖRGER², JAAP van der MEERE², FRUHLING RIJSDIJK¹, and PHILIP ASHERSON¹

¹MRC Social, Genetic and Developmental Psychiatry Centre, Institute of Psychiatry, King’s College London, UK ²Department of Developmental and Experimental Clinical Psychology, University of Groningen, The Netherlands

Abstract

Background—For candidate endophenotypes to be useful for psychiatric genetic research, they first of all need to show significant genetic influences. To address the relative lack of previous data, we set to investigate the extent of genetic and environmental influences on performance in a set of theoretically driven cognitive-experimental tasks in a large twin sample. We further aimed to illustrate how test–retest reliability of the measures affects the estimates.

Method—Four-hundred 7- to 9-year-old twin pairs were assessed individually on tasks measuring reaction time, inhibition, working memory and ‘delay aversion’ performance. Test–retest reliability data on some of the key measures were available from a previous study.

Results—Several key measures of reaction time, inhibition and working-memory performance indicated a moderate degree of genetic influence. Combining data across theoretically related tasks increased the heritability estimates, as illustrated by the heritability estimates of 60% for mean reaction time and 50% for reaction-time variability. Psychometric properties (reliability or ceiling effects) had a substantial influence on the estimates for some measures.

Conclusions—The data support the usefulness of several of the variables for endophenotype studies that aim to link genes to cognitive and motivational processes. Importantly, the data also illustrate specific conditions under which the true extent of genetic influences may be underestimated and hence the usefulness for genetic mapping studies compromised, and suggest ways to address this.

INTRODUCTION

‘Endophenotypes’ reflect intermediate phenotypes between aetiological factors and behavioural disorders or traits. The criteria for endophenotypes useful for psychiatric genetic research were originally proposed by Gottesman and colleagues and include the important criterion of heritability for candidate endophenotypes (Gottesman & Shields, 1973; Gottesman & Gould, 2003).

Yet the current interest in research on cognitive endophenotypes in disorders such as attention deficit hyperactivity disorder (ADHD) has not, overall, been matched by detailed

investigations of the extent of genetic influences on the candidate endophenotypes. Most twin studies on cognitive tasks to date have included relatively small sample sizes, affording limited statistical power for twin model-fitting analyses to estimate the contributions from genetic, shared environmental and child-specific environmental factors to task performance. A recent review of twin studies on measures of attention, executive functions and processing speed (Doyle *et al.* 2005) identified nine studies, but in five of these studies the total number of twin pairs was less than 60.

Twin studies of somewhat larger sample sizes indicate moderate heritability for performance on several of the tasks studied to date. In a study with 236 twin pairs, aged 16–29 years, Ando *et al.* (2001) obtained heritability estimates of 43–49% for measures of verbal and spatial working memory. Anokhin *et al.* (2003) assessed 168 twin pairs aged 17–28 years on the Wisconsin Card Sorting Test, estimating heritabilities of the various indices, including perseverative errors, as 37–46%. On the Stroop colour-word interference task, with a sample of 145 pairs of 12-year-old twins, a moderate degree of heritability (49%) was obtained for interference effect and somewhat higher heritability (70–75%) for time to complete each of the cards (Stins *et al.* 2004). However, performance on another response interference task (the Eriksen flanker task) suggested little genetic influence. The reason for this is not clear, but the authors note that the Stroop is likely to be a more reliable measure than the flanker task (Stins *et al.* 2004). In younger twins – 237 pairs nearly 6 years of age – heritability for reaction time to hits on a go/no-go task was estimated as 54% (Groot *et al.* 2004). On other variables of the go/no-go task (Groot *et al.* 2004), or on a selective attention or working-memory task (Stins *et al.* 2005), the authors could not distinguish between models indicating genetic influence and those indicating familial influence through shared environmental factors only. Studying a sample of 213 twin pairs at ages 16 and 18, Rijdsdijk *et al.* (1998) found moderately strong genetic effects (42–51%) on reaction-time performance on both simple and choice reaction-time tasks at both ages.

With a sample of 356 twins aged 15–18 years, Luciano *et al.* (2001) demonstrated heritabilities of 52% for two-choice, 59% for four-choice and 70% for eight-choice reaction times. An important aspect of the study was the demonstration of how test–retest reliability of the tasks affects the heritability estimates; a subsample of the twins was re-assessed on the tasks for this purpose. After correcting for unreliability, the heritability estimates for the reaction times increased to 79%, 90% and 85% respectively. The heritability estimate for performance on a spatial working-memory task similarly increased from 48% to 93%, after correction for unreliability. Test–retest unreliability deflates heritability estimates, as in twin model fitting such unreliability is automatically included in the child-specific environmental variance that always also includes measurement error. When variance due to measurement error ‘artificially’ explains a portion of the overall variance, the portion of the overall variance due to genetic (or shared environmental) influences will appear less.

Overall, data on the extent of genetic influences on performance on cognitive-experimental tasks are as yet limited and on a limited range of measures. Data are especially lacking for younger children. We aimed to investigate the extent of genetic and environmental influences on a set of theoretically driven tasks in a large sample of 400 pairs of twins aged 7–9 years. The tasks measure inhibition and reaction-time performance, under different event rate and incentive conditions, and ‘delay aversion’, which measures the ability to wait for large, delayed rewards under different overall delay periods. In addition, we used the digit span backwards score from the Wechsler Intelligence Scales as an estimate of working-memory performance. We aimed also to use these data to illustrate and discuss additional factors that can affect heritability estimates. Data from our previous test–retest reliability study enable a consideration of task unreliability on the results.

METHOD

Sample and procedure

Participants are members of the Study of Activity and Impulsivity Levels in children (SAIL), a study of a general population sample of twins at age 7–9 years. The sample was recruited from a birth cohort study, the Twins' Early Development Study (TEDS; Trouton *et al.* 2002), which had invited parents of all twins born in England and Wales during 1994–1996 to enrol. Despite attrition, the TEDS families continue to be fairly representative of the UK population with respect to parental occupation, education and ethnicity (Spinath *et al.* 2003). Zygosity has been determined using a standard zygosity questionnaire, which has been shown to have 95% accuracy (Price *et al.* 2000).

Families on the TEDS register were invited to take part, if they fulfilled the following SAIL project inclusion criteria: twins' birthdates between 1 September 1995 and 31 December 1996; lived within a feasible travelling distance (return daytrip) from the Research Centre; ethnic origin white European (to reduce population heterogeneity for molecular genetic studies); recent participation in TEDS, as indicated by return of questionnaires at either 4- or 7-year data collection point; no extreme pregnancy or perinatal difficulties (15 pairs excluded), specific medical syndromes and chromosomal anomalies (two pairs excluded) or epilepsy (one pair excluded); not participating in other current TEDS substudies (45 pairs excluded); and not on stimulant or other neuropsychiatric medications (two pairs excluded).

The current analyses focus on data obtained following contact with the first 693 suitable families on the register. Of these families, 400 families agreed to participate, reflecting a participation rate of 58%. The final sample consisted of 156 identical (monozygotic, MZ), 110 same-sex non-identical (dizygotic, DZ) and 134 opposite-sex DZ twin pairs. Data from nine individual children were subsequently excluded (five children with IQs below 70 and one child due to each of the following: neurofibromatosis, epilepsy, hypothyroidism and illness during testing). The mean age was 8.39 years ($s.d.=0.26$). All participants gave informed consent and the study was approved by the Institute of Psychiatry Ethical Committee (approval number 286/01).

The families visited the Research Centre for the assessments. Two testers assessed the twins simultaneously in separate testing rooms. The tasks were presented in the following fixed order: delay aversion, Wechsler Intelligence Scale for Children (WISC) picture completion, WISC similarities, go/no-go task, WISC block design, WISC vocabulary, fast task and WISC digit span. The total length of the testing session, including breaks, was approximately 2.5 h.

Measures

Wechsler Intelligence Scales for Children—(Wechsler, 1991). The vocabulary, similarities, picture completion and block design subtests from the WISC were used to obtain an estimate of the child's IQ [prorated following procedures described by Sattler (1992)]. The children's IQs ranged from 73 to 158 (mean=110.63, $s.d.=14.92$). In addition, the digit span subtest was administered to obtain an estimate of working memory (digit span backwards score).

The go/no-go task—(Borger & van der Meere, 2000; Kuntsi *et al.* 2005a; van der Meere *et al.* 1995). On each trial, one of two possible stimuli appeared for 300 ms in the middle of the computer screen. The child was instructed to respond only to the 'go' stimuli and to react as quickly as possible, but to maintain a high level of accuracy. The proportion of 'go' stimuli to 'no-go' stimuli was 4 : 1. The response variables are commission errors (an index

of inhibition), mean reaction time to 'go' stimuli and standard deviation of the reaction times. (Omission errors were rare – mean in each condition 2–12% – and are therefore not included in analyses.)

The children performed the task under three different conditions, matched for length of time on task. The fast condition consisted of 462 trials and had an inter-stimulus-interval (ISI) of 1 s. The ISI dropped to 8 s in the slow presentation condition, which consisted of 72 trials. The order of presentation of the slow and fast conditions was varied randomly across children.

The incentive condition was always administered last. This condition is a modification of the incentive condition used in the study on the stop task (another inhibition task) by Slusarek *et al.* (2001). Each correct response to the letter X and each correct non-response to the letter O earned the child 1 point. The child lost 1 point for each omission error (failure to respond to X) and for each failure to respond within 2 s. Each commission error (incorrect response to O) led to the loss of 5 points. The points were shown in a box, immediately right of the screen centre, and were updated continuously throughout. The child started with 40 points, to avoid the possibility of a negative tally. The child was asked to try to win as many points as possible, and was told that the points will be exchanged for a real prize when the game ends. This condition consisted of 72 trials and had an ISI of 8 s. A practice session preceded each experimental condition.

The fast task—(Kuntsi *et al.* 2005a). The base-line condition consisted of a standard warned four-choice reaction-time task, as outlined in Leth-Steensen *et al.* (2000). A warning signal (four empty circles, arranged side by side) first appeared on the screen. At the end of the fore-period (presentation interval for the warning signal), the circle designated as the target signal for that trial was filled (coloured) in. The child was asked to make a compatible choice response by pressing the response key that directly corresponded in position to the location of the target stimulus. Following a response, the stimuli disappeared from the screen and a fixed inter-trial interval of 2.5 s followed. Speed and accuracy were emphasized equally. If the child did not respond within 10 s, the trial terminated.

First a practice session was administered, during which the child had to respond correctly to five consecutive trials. The baseline condition, with a fore-period of 8 s and consisting of 72 trials, then followed.

To investigate the extent to which a response style characterized by slow and variable speed of responding can be maximally reduced, the task includes a comparison condition that uses a fast event rate (1 s) and incentives. This condition started immediately after the baseline condition and consisted of 80 trials (following the faster event rate conditions in Leth-Steensen *et al.* 2000). The child was told that if s/he will respond really quickly one after another, she will win smiley faces and will get real prizes in the end. The child won a smiley face each time she responded faster than his/her own mean reaction time during the baseline (first) condition consecutively for three trials. The baseline mean reaction time was calculated here based on the middle 94% of responses, therefore excluding extremely fast and extremely slow responses. The smiley faces appeared below the circles in the middle of the screen and were updated continuously. The response variables are mean reaction time and standard deviation of the reaction times, calculated for each condition based on correct responses only.

The Maudsley Index of Childhood Delay Aversion—(Kuntsi *et al.* 2001). This task was originally developed to test the delay aversion hypothesis that children with ADHD

choose immediate, small rewards over large, delayed ones, when this leads to less overall delay (Sonuga-Barke *et al.* 1992).

Two conditions, each with 20 trials, were administered. In each trial the child had a choice between a small immediate reward (1 point involving a 2-s pre-reward delay) and a large delayed reward (2 points involving a 30-s pre-reward delay). In the *no post-reward delay* condition, choosing the small reward led immediately to the next trial; this of course reduced the overall length of the condition. In the *post-reward delay* (control) condition, choosing the small reward led to a delay period of 30 s and choosing the large reward to a delay period of 2 s before the next trial; therefore the overall delay was constant and independent of choice made. The order of the two conditions was randomly chosen for each twin before the session.

The task was presented as a space game, in which the child had to destroy enemy spacecraft (using the computer mouse). There were 20 'missions' in each condition, and the tester removed one counter from the table next to the computer every time the mission ended to indicate the number of missions left. The aim of the game was to earn as many points as possible and the child was told that s/he would receive a real prize in the end. The delay aversion variable used in the analyses is the percentage of choices for the delayed reward. A short practice session preceded each condition.

Analyses

The structural equation-modelling program Mx (Neale, 1997) was used to conduct the genetic analyses. Models were fitted to age- and sex-standardized scores, using raw data analysis. Participants with incomplete data were included in the analyses as Mx provides a method for handling incomplete data by using raw maximum-likelihood estimation, in which a likelihood statistic ($-2LL$) of the data for each observation is calculated. This implies that there is no overall measure of fit (such as a χ^2 value with corresponding p value for the number of degrees of freedom (df), as obtained by fitting directly on observed variance-covariance matrices). Instead, with raw data, there are relative measures of fit: by comparing the $-2LL$ (and df) of our models with the $-2LL$ (and df) of the saturated model (where the maximum number of parameters is estimated to describe the correlational structure between variables) a χ^2 fit index is obtained. The df for this test is equal to the difference in df of the two models (Neale & Cardon, 1992). A χ^2 difference test can be performed to compare the fit of nested models. For non-nested models, the AIC (Akaike's Information Criteria) values are used to compare the fit of alternative models. Low (ideally negative) AIC values indicate less difference between the observed and predicted covariances and therefore better fit (Williams & Holahan, 1994). A difference in AIC between two models of <2 suggests substantial evidence for both models; a difference between 3 and 7 indicates that the higher AIC model has considerably less support; a difference of >10 indicates that the higher AIC model is very unlikely (Burnham & Anderson, 2002). Precision of parameter estimates were obtained by likelihood-based confidence intervals in Mx (Neale & Miller, 1997).

Univariate genetic analyses—The logic behind quantitative genetic analyses of twin data has three parts. First, MZ twins share all their inherited parental chromosomes and are therefore genetically identical, whereas DZ twins, like ordinary full siblings, share on average only half of their parental chromosomes and therefore 50% of inherited genetic variation. For environmental influences MZ and DZ twins are expected to correlate to the same extent. As such, when the similarity of MZ twins is greater than the similarity of DZ twins, this indicates a genetic contribution to the behaviour being measured. In model fitting, this yields a significant variance component called A (additive genetic variance).

Second, if only genes were influencing their behaviour, MZ twins' behaviour should be at least twice as similar as DZ twins'. If, however, DZ twin pairs are less than twice as similar as MZ twin pairs, this indicates that environments the children share in common have enhanced their similarity. In model fitting, this yields a variance component called C (common or shared environmental variance). Third, if MZ twins, despite sharing all their genes, are not perfectly identical in their behaviour, this indicates that experiences unique to each twin have reduced the twins' behavioural similarity. In model fitting, this yields a variance component called E (child-specific environmental variance, which, since it also includes measurement error, cannot be omitted from any model).

The full ACE model is fitted first. Then, to attain the most parsimonious model, parameters which do not significantly contribute to the fit of the model are dropped. The AE and CE models are nested within the full ACE model (i.e. subsets of free parameters in these models are contained in the full model).

RESULTS

Transformations were applied to normalize skewed distributions, as required for the quantitative genetic model fitting. The MZ and DZ within-pair correlations (Table 1) provide rough estimates of the extent to which genetic, shared environmental and child-specific environmental factors contribute to each variable. For most variables MZ correlations were greater than DZ correlations, suggesting genetic influence (see below for further discussion on exceptions). We fitted the three models (ACE, AE and CE) to each variable. Means and standard deviations on task variables are reported in Appendix A.

The AE model provided the best fit for the following variables, indicating no shared environmental influence: four go/no-go task variables [slow condition mean reaction time (MRT) and all variables from fast condition], two fast task variables [baseline MRT and *s.d.* of reaction times (RTs)] and digit span backwards score (Table 1, Appendix B). For the remaining variables, dropping either the A term (CE model) or the C term (AE model) did not significantly worsen the fit, compared to the ACE model. As the difference in the AIC values suggested substantial evidence for both models, we report parameter values from the full ACE model. This cautious, yet necessary, approach may have resulted in slightly underestimated heritabilities for variables where the AIC values indicated better fit for the AE than CE model and, conversely, in slightly overestimated heritabilities for variables where the AIC values indicated better fit for the CE than AE model (delay aversion task variables). The AIC values indicated good model fit for all variables except the go/no-go task variables of *s.d.* of RTs in the slow and incentive conditions and the fast task variable of *s.d.* of RTs in the fast-incentive condition; as such, the parameter estimates for these variables are preliminary.

For the go/no-go task variables heritabilities were estimated as 18–45% for commission errors, 31–54% for MRT and 10–43% for *s.d.* of RTs across the three conditions (Table 1). In the fast task baseline condition heritabilities were estimated as 55% for MRT and 37% for *s.d.* of RTs, and in the fast-incentive condition as 23% for MRT and 17% for *s.d.* of RTs (latter estimate preliminary). For the delay aversion task variables the model fitting produced heritabilities of 18% and 11%, but inspection of the data indicated that the model fitting was compromised due to ceiling effects: 22% (*no post-reward delay* condition) and 34% (*post-reward delay* condition) of the children performed at ceiling on the task. Such ceiling effects artificially increase twin similarity and therefore distort parameter estimates from model fitting. We were therefore not able to obtain reliable estimates of the extent of genetic and environmental influences on delay aversion performance. For digit span backwards score heritability was estimated as 36%.

To examine the extent to which heritability estimates may increase for composite scores, we carried out additional analyses combining the RT data across the go/no-go and fast tasks. Here we focused on baseline data only (baseline condition of fast task and slow condition of the go/no-go task), which reflect RT performance that is unaffected by manipulations with event rate or incentives. The AE model provided the best fit in each case and the AIC values indicated good model fit (Table 1, Appendix B). Heritabilities were estimated as 60% for MRT and 48% for *s.d.* of RTs (Table 1).

As reliability of the measures provides an upper limit for the heritability estimates, any test–retest unreliability will lead to underestimation of both additive genetic (A) and shared environment (C) components. Here we demonstrate how test–retest reliability may have affected the current heritability estimates, using data from our previous test–retest reliability study on the go/no-go and fast tasks on a separate general population sample of 49 children aged 8–13 years, controlling for age effects (Kuntsi *et al.* 2005a; test–retest interval 2 weeks). In model fitting the proportion of variance that is due to test–retest unreliability is automatically included in the ‘E’ parameter that reflects both child-specific environmental variance and measurement error. A test–retest reliability coefficient of 0.63, for example, indicates that 37% of the total variance is due to test–retest unreliability (plus possible transitory environmental influences) and therefore cannot be ‘explained’ within model fitting.

If we remove the variance due to test–retest unreliability and ‘correct’ the parameter estimates accordingly, we obtain heritability estimates as follows (test–retest partial inter-class correlations, correcting for age, in parentheses). For MRT, the heritability estimates increase from 52% to 83% ($r=0.63$) for go/no-go task slow condition, from 54% to 66% ($r=0.85$) for fast condition and from 31% to 49% ($r=0.63$) for incentive condition; for fast task, the heritability estimates increase from 55% to 73% ($r=0.75$) for baseline condition and from 23% to 29% ($r=0.79$) for fast-incentive condition. For *s.d.* of RTs, the heritability estimates increase from 29% to 55% ($r=0.53$) for go/no-go task slow condition, from 43% to 53% ($r=0.82$) for fast condition and from 10% to >100% (due to the low test–retest reliability of $r=0.09$ unless six outliers are excluded; see Kuntsi *et al.* 2005a) for incentive condition; for fast task, the heritability estimates increase from 37% to 70% ($r=0.53$) for baseline condition and from 17% to 26% ($r=0.65$) for fast-incentive condition. For commission errors in the go/no-go task, the heritability estimates increase from 18% to 32% ($r=0.56$) in the slow condition, from 45% to 67% in the fast condition ($r=0.67$) and from 29% to 62% ($r=0.47$) in the incentive condition. For the composite scores based on go/no-go task slow condition and fast task baseline condition data, the heritability estimates increase from 60% to 73% for MRT and from 48% to 68% for *s.d.* of RTs ($r=0.82$ and $r=0.71$ respectively; reported here for the first time).

DISCUSSION

With a large, population-based twin sample we demonstrated a moderate degree of heritability for reaction time, inhibition and working-memory performance. The data support the usefulness of several of the variables for endophenotype studies that aim to link genes to cognitive and motivational processes. Importantly, the data also illustrate specific conditions under which the true extent of genetic influences may be underestimated and hence the usefulness of cognitive-experimental measures for genetic mapping studies compromised.

Mean reaction time on both the go/no-go and fast tasks, and across the different conditions within each task, indicated heritability estimates in the range of 23–55%. For reaction-time variability, the heritability estimates were 37–43% in conditions where a good model fit was obtained, and lower at 10–29% in conditions where model fit was less good, calling for

caution in the interpretation of the estimates. For commission errors, an index of inhibition, heritability estimates across the three go/no-go task conditions were 18–45%. Working-memory performance, as measured using digit span backwards, indicated a similar extent of genetic influence, with heritability estimated at 36%. Shared environmental influences made no or only a small contribution to performance on these tasks. Overall, the heritability estimates confirm and extend previous findings (Rijsdijk *et al.* 1998; Ando *et al.* 2001; Luciano *et al.* 2001; Groot *et al.* 2004).

In contrast, model fitting on the delay aversion data suggested significant shared environmental effects and less genetic influence. Further inspection of the data indicated that the model fitting had been affected by ceiling effects. Such ceiling effects artificially inflate twin similarity for both MZ and DZ twins, resulting in an overestimation of shared environmental influences and underestimation of genetic influences. As such, the standard model fitting does not reveal the true extent of genetic influences on this task.

We also illustrated how reliability of the measures influences heritability estimates. Using data from our previous test–retest reliability study on the go/no-go and fast tasks (Kuntsi *et al.* 2005a), we showed how, despite moderate-to-good test–retest reliability for most variables, true heritability will be underestimated, as all deviations from perfect reliability will increase estimates for the ‘E’ parameter that reflects not only child-specific environmental influences but also measurement error. The larger the estimate for ‘E’, the less variance remains to be accounted for by genetic and shared environmental influences. Explained in another way, the test–retest reliability sets an upper limit for the heritability estimate. This issue has been largely ignored (but see Luciano *et al.* 2001), despite the important implications. If test–retest reliability data are not available, it is impossible to know whether a low heritability estimate could simply reflect poor reliability of the measure. As also illustrated here, comparisons of heritability estimates across conditions, tasks, or indeed studies, are problematic if reliability of the measures is unknown: the variations in the size of the heritability estimate may reflect variations in reliability rather than in the true extent of genetic influences. Correcting the estimates for test–retest unreliability suggested strong genetic influences on task performance, with heritability estimated as up to 50–80% for the majority of the variables.

Yet, if aiming to choose measures suitable for molecular genetic studies, the key focus will be on maximizing reliability rather than statistically correcting for test–retest unreliability within twin model fitting. The extent of detectable heritability is crucially important for the success of mapping genes to cognitive variables. We illustrated how, by creating composite scores, reliability can be improved and heritability estimates increased. Summing data on baseline performance across the go/no-go and fast tasks resulted in heritability estimates of around 60% for mean reaction time and 50% for reaction-time variability (with 95% confidence intervals of around 50–70% and 40–60% respectively). These increases in the heritability estimates reflect the improved reliability: test–retest reliability coefficients improved from 0.63 and 0.75 for the individual mean reaction times to 0.82 for the composite score, and from 0.53 for individual reaction-time variability (for both tasks) to 0.71 for the composite score. This is also seen in the heritability estimates corrected for test–retest unreliability, which were of similar magnitude for the composite scores as for the individual scores. Summed components are in general expected to have better reliability than single variables (Rousson *et al.* 2002) and are commonly used in many areas of measurement, including IQ estimates and behavioural rating scores. In contrast, this has received scant attention in cognitive endophenotype studies to date.

With the final SAIL sample – an estimated 700 twin pairs – we will be able to take the current analyses further by carrying out multivariate genetic analyses. Multivariate twin

model-fitting analyses will indicate the extent to which there are shared *versus* unique genetic and environmental influences on multiple variables (see, for example, Kuntsi *et al.* 2005*b*) and, as such, will inform the creation of composite scores.

Acknowledgments

The Study of Activity and Impulsivity Levels in children (SAIL) is funded by a project grant from the Wellcome Trust (GR070345MF). Thanks go to all who make this research possible: the TEDS-SAIL families, who give their time and support so unstintingly; Eda Salih, Rebecca Gibbs, Kayley O'Flynn, Suzi Marquis and Rebecca Whittemore; and everyone on the TEDS team.

APPENDIX A: Means and standard deviations on task variables

	Twin 1	Twin 2
	Mean (s.d.)	Mean (s.d.)
Go/no-go task slow condition		
MRT		
MZ	616.70 (138.93)	619.20 (136.42)
DZ	606.66 (138.33)	600.00 (133.26)
S.D. of RTs		
MZ	234.45 (151.05)	226.27 (147.16)
DZ	244.29 (162.96)	241.08 (168.39)
Commission errors		
MZ	54.88 (23.30)	52.82 (23.11)
DZ	56.22 (23.31)	57.29 (22.09)
Go/no-go task fast condition		
MRT		
MZ	446.62 (61.97)	449.53 (59.00)
DZ	428.75 (62.89)	435.45 (61.18)
S.D. of RTs		
MZ	174.83 (54.79)	175.25 (56.37)
DZ	165.20 (57.07)	166.61 (56.19)
Commission errors		
MZ	51.51 (16.07)	51.73 (17.32)
DZ	53.28 (15.44)	51.29 (16.76)
Go/no-go task incentive condition		
MRT		
MZ	591.45 (134.65)	582.39 (115.12)
DZ	559.79 (106.45)	558.67 (102.95)
S.D. of RTs		
MZ	161.01 (94.04)	161.11 (80.03)
DZ	147.10 (72.25)	153.41 (75.20)
Commission errors		
MZ	31.32 (20.23)	33.29 (22.25)
DZ	34.10 (21.99)	35.68 (20.40)
Fast task baseline condition		

	Twin 1	Twin 2
	Mean (s.d.)	Mean (s.d.)
MRT		
MZ	1014.89 (232.48)	1037.15 (220.82)
DZ	991.56 (227.50)	1020.26 (262.74)
S.D. of RTs		
MZ	437.81 (277.71)	479.68 (313.11)
DZ	441.68 (273.89)	467.82 (367.12)
Fast task fast-incentive condition		
MRT		
MZ	699.23 (162.07)	693.42 (171.32)
DZ	693.03 (177.59)	668.66 (163.92)
S.D. of RTs		
MZ	229.81 (143.69)	242.33 (225.97)
DZ	222.04 (145.54)	218.26 (127.72)
Go/no-go (slow) and fast task (baseline) combined		
MRT		
MZ	1624.19 (312.43)	1656.47 (309.57)
DZ	1588.90 (301.69)	1614.27 (346.26)
S.D. of RTs		
MZ	665.25 (365.09)	703.19 (391.63)
DZ	676.36 (360.42)	698.16 (428.49)
Delay aversion		
No post-reward delay		
MZ	64.27 (30.04)	63.07 (26.61)
DZ	68.06 (27.01)	64.20 (28.73)
Post-reward delay		
MZ	79.83 (22.67)	80.96 (19.12)
DZ	81.75 (19.66)	81.80 (19.59)
Digit span backwards		
MZ	4.35 (1.40)	4.24 (1.37)
DZ	4.25 (1.26)	4.39 (1.53)

Mean (s.d.) on raw (untransformed) data.

MRT, Mean reaction time; RT, reaction time.

APPENDIX B: Fit of quantitative genetic models

Model	-2LL	df	χ^2	df	<i>p</i>	χ^2	df	<i>p</i>	AIC
Go/no-go task slow condition									
MRT									
Saturated model	310.96	757							
1. ACE	327.49	777	16.53	20	0.68				-23.47

Model	-2LL	df	χ^2	df	p	χ^2	df	p	AIC
2. AE	328.34	778	17.38	21	0.69	0.85	1	0.36	-24.62
3. CE	333.03	778	22.07	21	0.40	5.54	1	0.02	-19.93
S.D. of RTs									
Saturated model	18.23	757							
1. ACE	70.09	777	51.86	20	<0.001				11.86
2. AE	70.18	778	51.95	21	<0.001	0.09	1	0.76	9.95
3. CE	72.29	778	54.06	21	<0.001	2.20	1	0.14	12.06
Commission errors									
Saturated model	2193.79	757							
1. ACE	2205.78	777	11.99	20	0.92				-28.01
2. AE	2205.78	778	11.99	21	0.94	0	1	1.00	-30.01
3. CE	2208.24	778	14.45	21	0.85	2.46	1	0.12	-27.55
Go/no-go task fast condition									
MRT									
Saturated model	2152.62	761							
1. ACE	2157.72	781	5.10	20	1.00				-34.90
2. AE	2157.72	782	5.10	21	1.00	0	1	1.00	-36.90
3. CE	2172.71	782	20.09	21	0.52	14.99	1	<0.001	-21.91
S.D. of RTs									
Saturated model	164.13	761							
1. ACE	198.80	781	34.67	20	0.02				-5.33
2. AE	199.01	782	34.88	21	0.03	0.21	1	0.65	-7.12
3. CE	202.66	782	38.53	21	0.01	3.86	1	<0.05	-3.47
Commission errors									
Saturated model	2157.84	761							
1. ACE	2173.13	781	15.29	20	0.76				-24.71
2. AE	2173.37	782	15.53	21	0.80	0.24	1	0.62	-26.47
3. CE	2177.90	782	20.06	21	0.52	4.77	1	0.03	-21.94
Go/no-go task incentive condition									
MRT									
Saturated model	385.43	751							
1. ACE	406.05	771	20.62	20	0.42				-19.38
2. AE	406.58	772	21.15	21	0.45	0.53	1	0.47	-20.85
3. CE	409.06	772	23.63	21	0.31	3.01	1	0.08	-18.37
S.D. of RTs									
Saturated model	3528.30	773							
1. ACE	3631.48	793	103.18	20	<0.001				63.18
2. AE	3632.89	794	104.59	21	<0.001	1.41	1	0.24	62.59
3. CE	3631.51	794	103.21	21	<0.001	0.03	1	0.86	61.21
Commission errors									
Saturated model	2164.51	751							
1. ACE	2177.39	771	12.88	20	0.89				-27.12

Model	-2LL	df	χ^2	df	p	χ^2	df	p	AIC
2. AE	2177.39	772	12.88	21	0.91	0	1	1.00	-29.12
3. CE	2180.22	772	15.71	21	0.79	2.83	1	0.09	-26.29
Fast task baseline condition									
MRT									
Saturated model	125.68	725							
1. ACE	146.68	745	21.00	20	0.40				-19.00
2. AE	146.83	746	21.15	21	0.45	0.15	1	0.70	-20.85
3. CE	154.87	746	29.19	21	0.11	8.19	1	0.004	-12.81
S.D. of RTs									
Saturated model	3412.31	725							
1. ACE	3443.20	745	30.89	20	0.06				-9.11
2. AE	3443.20	746	30.89	21	0.08	0.00	1	1.00	-11.11
3. CE	3449.17	746	36.86	21	0.02	5.97	1	0.01	-5.14
Fast task fast-incentive condition									
MRT									
Saturated model	308.51	724							
1. ACE	343.00	744	34.49	20	0.02				-5.51
2. AE	343.33	745	34.82	21	0.03	0.33	1	0.57	-7.18
3. CE	344.32	745	35.81	21	0.02	1.32	1	0.25	-6.19
S.D. of RTs									
Saturated model	2985.39	725							
1. ACE	3087.69	745	102.30	20	<0.001				62.30
2. AE	3087.73	746	102.34	21	<0.001	0.04	1	0.84	60.34
3. CE	3088.32	746	102.93	21	<0.001	0.63	1	0.43	60.93
Go/no-go (slow) and fast task (baseline) combined									
MRT									
Saturated model	115.60	718							
1. ACE	137.44	738	21.84	20	0.35				-18.16
2. AE	137.85	739	22.25	21	0.39	0.41	1	0.52	-19.75
3. CE	147.29	739	31.69	21	0.06	9.85	1	0.002	-10.31
S.D. of RTs									
Saturated model	221.91	718							
1. ACE	237.24	738	15.33	20	0.76				-24.67
2. AE	237.24	739	15.33	21	0.81	0	1	1.00	-26.67
3. CE	245.23	739	23.32	21	0.33	7.99	1	0.005	-18.68
Delay aversion									
No post-reward delay									
Saturated model	2014.47	708							
1. ACE	2031.45	728	16.98	20	0.65				-23.02
2. AE	2033.53	729	19.06	21	0.58	2.08	1	0.15	-22.94
3. CE	2032.43	729	17.96	21	0.65	0.98	1	0.32	-24.04

Model	−2LL	df	²	df	p	²	df	p	AIC
Post-reward delay									
Saturated model	1992.45	702							
1. ACE	2007.08	722	14.63	20	0.80				−25.37
2. AE	2010.79	723	18.34	21	0.63	3.71	1	0.05	−23.66
3. CE	2007.48	723	15.03	21	0.82	0.40	1	0.53	−26.97
Digit span backwards									
Saturated model	2181.62	763							
1. ACE	2204.14	783	22.52	20	0.31				−17.48
2. AE	2204.14	784	22.52	21	0.37	0	1	1.00	−19.48
3. CE	2211.51	784	29.89	21	0.09	7.37	1	0.007	−12.11

The AE and CE models were compared to the ACE model. Best-fitting model indicated in bold.

MRT, Mean reaction time; RT, reaction time; −2LL, likelihood statistic; AIC, Akaike's Information Criteria.

REFERENCES

- Ando J, Ono Y, Wright MJ. Genetic structure of spatial and verbal working memory. *Behavior Genetics*. 2001; 31:615–624. [PubMed: 11838538]
- Anokhin AP, Heath AC, Ralano A. Genetic influences on frontal brain function: WCST performance in twins. *Neuroreport*. 2003; 14:1975–1978. [PubMed: 14561932]
- Borger N, van der Meere J. Motor control and state regulation in children with ADHD: a cardiac response study. *Biological Psychology*. 2000; 51:247–267. [PubMed: 10686368]
- Burnham, KP.; Anderson, DR. *Model Selection and Multimodel Inference: A Practical Information-theoretic Approach*. 2nd edn. Springer-Verlag; New York: 2002.
- Doyle AE, Willcutt EG, Seidman LJ, Biederman J, Chouinard VA, Silva J, Faraone SV. Attention-deficit/hyperactivity disorder endophenotypes. *Biological Psychiatry*. 2005; 57:1324–1335. [PubMed: 15950005]
- Gottesman II, Gould TD. The endophenotype concept in psychiatry: etymology and strategic intentions. *American Journal of Psychiatry*. 2003; 160:636–645. [PubMed: 12668349]
- Gottesman II, Shields J. Genetic theorizing and schizophrenia. *British Journal of Psychiatry*. 1973; 122:15–30. [PubMed: 4683020]
- Groot AS, de Sonnevile LM, Stins JF, Boomsma DI. Familial influences on sustained attention and inhibition in preschoolers. *Journal of Child Psychology and Psychiatry*. 2004; 45:306–314. [PubMed: 14982244]
- Kuntsi J, Andreou P, Ma J, Borger NA, van der Meere J. Testing assumptions for endophenotype studies in ADHD: reliability and validity of tasks in a general population sample. *BMC Psychiatry*. 2005a; 5:40. [PubMed: 16262903]
- Kuntsi J, Rijdsdijk F, Ronald A, Asherson P, Plomin R. Genetic influences on the stability of attention-deficit/hyperactivity disorder symptoms from early to middle childhood. *Biological Psychiatry*. 2005b; 57:647–654. [PubMed: 15780852]
- Kuntsi J, Stevenson J, Oosterlaan J, Sonuga-Barke EJS. Test-retest reliability of a new delay aversion task and executive function measures. *British Journal of Developmental Psychology*. 2001; 19:339–348.
- Leth-Steensen C, Elbaz ZK, Douglas VI. Mean response times, variability, and skew in the responding of ADHD children: a response time distributional approach. *Acta Psychologica (Amsterdam)*. 2000; 104:167–190.
- Luciano M, Wright M, Smith GA, Geffen GM, Geffen LB, Martin NG. Genetic covariance among measures of information processing speed, working memory, and IQ. *Behavior Genetics*. 2001; 31:581–592. [PubMed: 11838535]

- Neale, MC. Mx: Statistical Modelling. 4th edn. Department of Psychiatry; Richmond, VA: 1997.
- Neale, MC.; Cardon, LR. Methodology for Genetic Studies of Twins and Families. Kluwer Academic Publishers; Dordrecht: 1992.
- Neale MC, Miller MB. The use of likelihood-based confidence intervals in genetic models. *Behavior Genetics*. 1997; 27:113–120. [PubMed: 9145549]
- Price TS, Freeman B, Craig I, Petrill SA, Ebersole L, Plomin R. Infant zygosity can be assigned by parental report questionnaire data. *Twin Research*. 2000; 3:129–133. [PubMed: 11035484]
- Rijsdijk FV, Vernon PA, Boomsma DI. The genetic basis of the relation between speed-of-information-processing and IQ. *Behavioural Brain Research*. 1998; 95:77–84. [PubMed: 9754879]
- Rousson V, Gasser T, Seifert B. Assessing intrarater, interrater and test-retest reliability of continuous measurements. *Statistics in Medicine*. 2002; 21:3431–3446. [PubMed: 12407682]
- Sattler, JM. Assessment of Children: WISC-III and WPPSI-R Supplement. Jerome M. Sattler; San Diego: 1992.
- Slusarek M, Velling S, Bunk D, Eggers C. Motivational effects on inhibitory control in children with ADHD. *Journal of the American Academy of Child and Adolescent Psychiatry*. 2001; 40:355–363. [PubMed: 11288778]
- Sonuga-Barke EJS, Taylor E, Sembi S, Smith J. Hyperactivity and delay aversion – I. The effect of delay on choice. *Journal of Child Psychology and Psychiatry*. 1992; 33:387–398. [PubMed: 1564081]
- Spinath FM, Ronald A, Harlaar N, Price TS, Plomin R. Phenotypic ‘g’ early in life: On the etiology of general cognitive ability in a large population sample of twin children aged 2 to 4 years. *Intelligence*. 2003; 31:195–210.
- Stins JF, de Sonneville LM, Groot AS, Polderman TC, van Baal CG, Boomsma DI. Heritability of selective attention and working memory in preschoolers. *Behavior Genetics*. 2005; 35:407–416. [PubMed: 15971022]
- Stins JF, van Baal GC, Polderman TJ, Verhulst FC, Boomsma DI. Heritability of Stroop and flanker performance in 12-year old children. *BMC Neuroscience*. 2004; 5:49. [PubMed: 15579206]
- Trouton A, Spinath FM, Plomin R. Twins early development study (TEDS): a multivariate, longitudinal genetic investigation of language, cognition and behavior problems in childhood. *Twin Research*. 2002; 5:444–448. [PubMed: 12537874]
- van der Meere J, Stermerdink N, Gunning B. Effects of presentation rate of stimuli on response inhibition in ADHD children with and without tics. *Perceptual Motor Skills*. 1995; 81:259–262. [PubMed: 8532467]
- Wechsler, D. Wechsler Intelligence Scale for Children. 3rd edn. The Psychological Corporation; London: 1991.
- Williams LJ, Holahan PJ. Parsimony-based fit indices for multiple-indicator models: do they work? *Structural Equation Modeling*. 1994; 1:161–189.

Table 1

Within-pair Pearson correlations and estimates of additive genetic (a^2), shared environmental (c^2) and child-specific environmental (e^2) contributions to task performance (full ACE model, and best-fitting model if different, with 95% confidence intervals)

Model	<i>r</i>		Full ACE model						Best-fitting model					
	MZ	DZ	a^2	c^2	e^2	χ^2 (df)	<i>p</i>	AIC	a^2	c^2	e^2	χ^2 (df)	<i>p</i>	AIC
Go/no-go task slow condition														
MRT	0.51	0.30	0.38 (0.06–0.60)	0.12 (0.00–0.37)	0.50 (0.40–0.62)	16.53 (20)	0.68	-23.47	0.52 (0.41–0.61)	—	0.48 (0.39–0.59)	17.38 (21)	0.69	-24.62
S.D. of RTs	0.31	0.20	0.29 (0.00–0.45)	0.04 (0.00–0.30)	0.67 (0.55–0.83)	51.86 (20)	<0.001	11.86	—	—	—	—	—	—
Commission errors	0.23	0.03	0.18 (0.00–0.31)	0.00 (0.00–0.17)	0.82 (0.69–0.95)	11.99 (20)	0.92	-28.01	—	—	—	—	—	—
Go/no-go task fast condition														
MRT	0.53	0.23	0.54 (0.32–0.63)	0.00 (0.00–0.17)	0.46 (0.37–0.57)	5.10 (20)	1.00	-34.90	0.54 (0.43–0.63)	—	0.46 (0.37–0.57)	5.10 (21)	1.00	-36.90
S.D. of RTs	0.38	0.24	0.35 (0.00–0.53)	0.06 (0.00–0.32)	0.59 (0.47–0.73)	34.67 (20)	0.20	-5.33	0.43 (0.31–0.53)	—	0.57 (0.46–0.69)	34.88 (21)	0.03	-7.12
Commission errors	0.46	0.25	0.38 (0.04–0.55)	0.07 (0.00–0.32)	0.56 (0.45–0.69)	15.29 (20)	0.76	-24.71	0.45 (0.34–0.55)	—	0.55 (0.45–0.66)	15.53 (21)	0.80	-26.47
Go/no-go task incentive condition														
MRT	0.43	0.24	0.31 (0.00–0.52)	0.10 (0.00–0.37)	0.59 (0.48–0.72)	20.62 (20)	0.42	-19.38	—	—	—	—	—	—
S.D. of RTs	0.23	0.15	0.10 (0.00–0.33)	0.09 (0.00–0.26)	0.80 (0.67–0.93)	103.18 (20)	<0.001	63.18	—	—	—	—	—	—
Commission errors	0.30	0.13	0.29 (0.00–0.41)	0.00 (0.00–0.25)	0.71 (0.59–0.85)	12.88 (20)	0.89	-27.12	—	—	—	—	—	—
Fast task baseline condition														
MRT	0.52	0.29	0.49 (0.16–0.64)	0.05 (0.00–0.30)	0.46 (0.36–0.59)	21.00 (20)	0.40	-19.00	0.55 (0.44–0.64)	—	0.45 (0.36–0.56)	21.15 (21)	0.45	-20.85
S.D. of RTs	0.38	0.14	0.37 (0.09–0.49)	0.00 (0.00–0.19)	0.63 (0.51–0.77)	30.89 (20)	0.06	-9.11	0.37 (0.23–0.49)	—	0.63 (0.51–0.77)	30.89 (21)	0.08	-11.11
Fast task fast-incentive condition														
MRT	0.32	0.19	0.23 (0.00–0.45)	0.09 (0.00–0.33)	0.68 (0.55–0.83)	34.49 (20)	0.02	-5.51	—	—	—	—	—	—
S.D. of RTs	0.21	0.11	0.17 (0.00–0.33)	0.03 (0.00–0.24)	0.80 (0.67–0.94)	102.30 (20)	<0.001	62.30	—	—	—	—	—	—
Go/no-go (slow) and fast task (baseline) combined														
MRT	0.57	0.33	0.50 (0.20–0.68)	0.08 (0.00–0.32)	0.41 (0.32–0.53)	21.84 (20)	0.35	-18.16	0.60 (0.49–0.68)	—	0.40 (0.32–0.51)	22.25 (21)	0.39	-19.75
S.D. of RTs	0.46	0.21	0.48 (0.17–0.58)	0.00 (0.00–0.23)	0.52 (0.42–0.66)	15.33 (20)	0.76	-24.67	0.48 (0.35–0.58)	—	0.52 (0.42–0.65)	15.33 (21)	0.81	-26.67
Delay aversion														
No post-reward delay	0.39	0.29	0.18 (0.00–0.49)	0.20 (0.00–0.40)	0.62 (0.50–0.75)	16.98 (20)	0.65	-23.02	—	—	—	—	—	—
Post-reward delay	0.41	0.32	0.11 (0.00–0.45)	0.27 (0.00–0.43)	0.62 (0.50–0.74)	14.63 (20)	0.80	-25.37	—	—	—	—	—	—

Model	<i>r</i>			Full ACE model					Best-fitting model				
	MZ	DZ		a ²	e ²	χ^2 (df)	<i>p</i>	AIC	a ²	e ²	χ^2 (df)	<i>p</i>	AIC
Digit span backwards													
	0.40	0.12		0.36 (0.13–0.48)	5.7 × 10 ⁻¹³ (0.00–0.16)	22.52 (20)	0.31	-17.48	0.36 (0.24–0.48)	—	22.52 (21)	0.37	-19.48

MRT, Mean reaction time; RT, reaction time; AIC, Akaike's Information Criteria.