Prescription Stimulants' Effects on Healthy Inhibitory Control, Working Memory, and Episodic Memory: A Meta-analysis

Irena P. Ilieva, Cayce J. Hook, and Martha J. Farah

Abstract

■ The use of prescription stimulants to enhance healthy cognition has significant social, ethical, and public health implications. The large number of enhancement users across various ages and occupations emphasizes the importance of examining these drugs' efficacy in a nonclinical sample. The present meta-analysis was conducted to estimate the magnitude of the effects of methylphenidate and amphetamine on cognitive functions central to academic and occupational functioning, including inhibitory control, working memory, short-term episodic memory, and delayed episodic memory. In addition, we examined the evidence for publication bias. Forty-eight studies (total of 1,409 participants) were included in the analyses. We found evidence for small but significant stimulant enhancement effects on inhibitory control and short-term episodic memory. Small effects on working memory reached significance, based on one of our two analytical approaches. Effects on delayed episodic memory were medium in size. However, because the effects on long-term and working memory were qualified by evidence for publication bias, we conclude that the effect of amphetamine and methylphenidate on the examined facets of healthy cognition is probably modest overall. In some situations, a small advantage may be valuable, although it is also possible that healthy users resort to stimulants to enhance their energy and motivation more than their cognition. ■

INTRODUCTION

The scientific and popular literatures both document the use of prescription medications by healthy young people to enhance cognitive performance in school and on the job (e.g., Smith & Farah, 2011; Talbot, 2009). This practice, called "cognitive enhancement," has provoked wide discussion of its potential social, ethical, and public health consequences (Greely, 2013; Sahakian & Morein-Zamir, 2011; Farah et al., 2004). Recently, another question concerning cognitive enhancement has arisen: To what degree do the medications used for cognitive enhancement in fact improve the abilities of cognitively normal individuals?

In view of the prevalence of cognitive enhancement and the intensity of academic and policy interest in this practice, it is surprising that the answer to this question has not been clearly established. The empirical literature on the effects of these stimulants on cognition in normal participants has yielded variable results, with some reviewers doubting their efficacy altogether. For example, in reviewing the literature on the cognitive effects of methylphenidate, Repantis, Schlattmann, Laisney, and Heuser (2010) concluded that they were "not able to provide sufficient evidence of positive effects in healthy

University of Pennsylvania

individuals from objective tests." Hall and Lucke (2010) stated: "There is very weak evidence that putatively neuro-enhancing pharmaceuticals in fact enhance cognitive function." Advokat (2010) concluded her review of the literature by stating that "studies in non-ADHD adults suggest that stimulants may actually impair performance on tasks that require adaptation, flexibility and planning."

Smith and Farah (2011) attempted to test the hypothesis that stimulants enhance cognitive performance in normal healthy participants with a systematic literature review. We included studies of amphetamine and methylphenidate's effects on episodic and procedural memory and three categories of executive function: working memory, inhibitory control, and third category of other executive function tasks that did not fit into either of the first two. Results were mixed, with large effects, small effects, and null effects all reported. For example, over a third of the executive function studies reported null results. One interpretation of this pattern is that the drugs confer a small benefit, which may fail to be detected in some studies because of inadequate power. The other possibility is that chance positive findings, combined with publication bias, may be responsible for the positive evidence that exists in the literature. Thus, despite the large literature included in our review, we were forced to conclude that "there remains great uncertainty regarding the size and robustness of these effects." Meta-analysis is a method that

can distinguish between the competing interpretations of the findings in the cognitive enhancement literature.

The primary goal of the present meta-analysis is to obtain a quantitative estimate of the cognitive effects of the stimulants amphetamine and methylphenidate. They are commonly prescribed for the treatment of attention deficit hyperactivity disorder (ADHD) but are frequently diverted for enhancement use by students and others (e.g., Wilens et al., 2008; Poulin, 2007; McCabe, Teter, & Boyd, 2006). Guided by the findings of Smith and Farah's (2011) review, we focus on the cognitive processes that seemed most likely to be enhanced by stimulants, specifically inhibitory control, working memory, and episodic memory. In addition, because this earlier review found the strongest evidence of episodic memory enhancement after long delays between learning and test, we distinguish between episodic memory tested soon after learning (within 30 min after learning trials) and episodic memory tested after longer intervals (1 hr to 1 week).

The meta-analysis has two additional goals: one is to test hypotheses about moderators of the effects, that is, differences between studies that might account for the variability in effectiveness noted across different studies. For example, perhaps one of the stimulants is effective and the other less so, or perhaps low doses are more effective than higher doses. The final goal is to assess the role of publication bias in shaping the literature and potentially inflating effect size estimates. This would happen if, as hypothesized (e.g., Smith & Farah, 2011), underpowered studies obtained large statistically significant effects by chance and thereby entered the literature while the balancing effects of smaller or null results from similar studies remained unpublished.

METHODS

Search Strategies

Online databases PubMed and PsychInfo were searched with key words *amphetamine* and *methylphenidate*, each combined with each of the following: *executive function, executive control, cognitive control, inhibitory control, inhibition, working memory, flanker, stop signal task, stop task, no-go, card-sort, ID/ED, set shifting, Sternberg memory, Stroop, Digit Span, memory, learning, recall, recognition,* and *retention.* These searches were narrowed to exclude research on nonhuman participants, qualitative studies, and nonempirical publications (e.g., review articles, meta-analyses, lectures, news articles, etc.). In addition, the reference sections of the following review articles were searched for relevant articles: Chamberlain et al. (2011), Smith and Farah (2011), Advokat (2010), and Repantis et al. (2010). Finally, we searched the list of articles being reviewed by an American Academy of Neurology committee studying cognitive enhancement on which the last author serves. All

research published through the end of December 2012 were eligible.

We also sought relevant unpublished data to include in the meta-analyses. Twenty researchers active in the area were contacted for unpublished data on amphetamine or methylphenidate effects on episodic memory, working memory, or inhibitory control in healthy nonelderly adults. In addition, 14 requests were made for additional data from studies published in the past 10 years but reporting insufficient data to calculate effect sizes. This led to obtaining two data sets of studies in progress or in submission as well as additional effect size data from four published reports. An additional unpublished pilot data set from our laboratory was also included.

Criteria for Study Eligibility

Publication Type and Language

Empirical investigations in any report format were eligible for inclusion in the meta-analysis. These included journal articles as well as dissertations, conference presentations, and unpublished data sets. The latter three were considered in an attempt to minimize the influence of publication bias on the obtained effect size estimates. Only reports in English were included.

Participants

Eligible participants were young and middle-aged adults. Research on children, elderly, criminal, or mentally ill participants was excluded. Studies were also excluded if the experimental procedure entailed sleep deprivation.

Research Design: Methodological Quality

A double-blind, placebo-controlled design was required for inclusion. This criterion aimed to maximize the methodological quality of the meta-analyzed material.

Research Design: Intervention

Eligible interventions were orally administered amphetamine and methylphenidate, with drugs administered before the start of the cognitive protocol (e.g., not after learning in a memory experiment). We only included research on single-dose administration (the only study on the effect of repeated administration was excluded because of lack of consistency of intervention strength with the rest of the available research). In the included studies, the interval between drug administration and the cognitive task ranged between 30 min and 4.5 hr for amphetamine studies and 40 min and 4.5 hr for research on methylphenidate. These intervals are within the medications' window of effectiveness (Volkow et al., 1998; Angrist, Corwin, Bartlik, & Cooper, 1987; Vree & Van Rossum, 1970). In addition, it is not unreasonable to

suspect that these waiting times have ecological validity, with users working or studying similar intervals after drug intake.

Studies including multiple intervention arms, such as different drugs or TMS, were included only if the effects of amphetamine and methylphenidate could be assessed in isolation (e.g., without concurrent TMS) and compared with placebo.

Cognitive Systems under Investigation

Four abilities central to academic and professional work were included based on the findings of Smith and Farah's (2011) literature review. They were *inhibitory control*, the ability to override dominant, habitual, or automatic responses for the sake of implementing more adaptive, goal-directed behaviors; *working memory*, the capacity to temporarily store and manipulate information in the service of other ongoing cognitive functions; *episodic memory*, the ability to encode, store, and retrieve task-relevant information, assessed *shortly* after learning (i.e., within 30 min) and at *longer delays* (1 hr to 1 week). Whenever task descriptions were not sufficient to identify the cognitive function tested, the data were excluded.

Outcome Measures

Performance can be measured by RT, overall accuracy, or specific types of error such as misses or false alarms. Overall in the literature, research reports varied in the types and number of outcome measures reported for each task. To maintain the validity and consistency of outcome measures in our analyses, we designed an a priori outcome selection procedure, as shown in Table 1. Our outcome selection strategy favored the most widely used and construct valid measures but also included second-best options, whenever our first choices were not reported. In general, we favored error measures over RT measures unless accuracy was near ceiling; in which case, RT data, if available, were coded. On tests of inhibitory control, instead of overall accuracy, more specific accuracy measures (or the relationships thereof) were used, such as a measure of false alarms on go/no-go tasks or the contrast in performance on incongruent and congruent trials of Flanker and Stroop. Whenever relevant, our main outcome measure was tailored to the specific design of the task. Particularly, two variants of the stop signal task of inhibitory control have been used in the examined literature: a version where the probability of stopping is allowed to vary and is the main measure of inhibition (e.g., Fillmore, Kelly, & Martin, 2005) and a version where the probability of stopping is held constant (e.g., De Wit, Enggasser, & Richards, 2002; Logan, Schachar, & Tannock, 1997), in which case stop signal RT is the main outcome. Eligible outcome measures for each task are shown in Table 1.

Process of Determining Study Eligibility

The search process, summarized in Figure 1, led to the identification of 1,799 titles, which were narrowed down to 1,505 after 294 duplicate articles were removed. After screening the titles of these articles, additional 1,304 reports were excluded for not meeting the inclusion criteria. The remaining 201 studies were assessed for eligibility by applying the exclusion criteria to the abstract and, in case of insufficient data, to the full text.

Of the remaining 201 studies, 73 were excluded because the measured cognitive constructs (e.g., simple RT, sustained attention, creativity, intelligence, fear conditioning, motor performance, reward processing, probabilistic learning, etc.) were outside the scope of the present review. Twelve studies failed to meet the criteria for eligible participants (mice: $n = 1$; elderly participants: $n = 6$; children: $n = 2$; mentally ill participants: $n = 2$, including one study on ADHD and one study on cocaine abuse; criminal participants: $n = 1$). Eighteen reports lacked a double-blind placebo-controlled design (when these design features were not explicitly mentioned, the study was excluded). Sixteen reports were excluded because of ineligible intervention. These included four studies that tested drugs other than amphetamine or methylphenidate, four studies in which drugs were administered intravenously, four studies conducted in the context of sleep deprivation, two studies in which outcomes were measured under TMS, one study in which drug administration followed (as opposed to preceding) learning, and one study that tested the effect of multiple drug doses. Seven studies in language other than English were excluded. Four studies could not be retrieved from available online and article sources. Four studies were excluded because of duplicating the data of already included research. In 19 of the remaining otherwise eligible studies, reported data were insufficient to calculate effect size. The final analyses were based on 48 articles reporting at least one relevant effect size (44 published reports, 3 unpublished data sets, and 1 dissertation with 1,409 participants). The first and second authors independently conducted the eligibility determination procedures; disagreements were resolved by consensus after reviewing the experimental reports.

Coding Procedures

All studies were coded by the first author, according to a standardized coding manual. Coded variables included means and standard deviations for performance under drug and placebo, sample size, outcome measure, effect direction, significance level, and several moderators. The moderators and rationale for examining their effects were that following:

(1) Drug (methylphenidate vs. amphetamine): This moderator analysis was conducted to examine if amphetamine and methylphenidate differ in their cognitive

Table 1. Eligible Measures for Examined Tasks

Cognitive Construct	Task	Eligible Measure(s)	Reference Supporting Choice of Measure
Inhibitory control	Stop Signal task	Depending on task design:	
		• Stop signal RT (mean go RT minus mean stop delay)	• Logan et al. (1997)
		• Probability of inhibiting a response	• Lappin and Eriksen (1966)
	Go/no-go	• False alarms or no-go accuracy	• Aron, Robbins, and Poldrack (2004); Helmers, Young, and Pihl (1995)
	Wisconsin Card Sort	• Perseverative errors	• Heaton, Chelune, Talley, Kay, and Curtis (1993)
		• If unavailable: accuracy	
	ID/ED	• Perseverative extradimensional shift errors	• Rogers et al. (1999)
	Flanker	• Difference or ratio between accuracy in the congruent and incongruent conditions	• Eriksen and Eriksen (1974)
		• If unavailable: incongruent condition accuracy	
		• If accuracy was at ceiling, corresponding RTs were coded	
	Stroop	• Difference or ratio between accuracy in the congruent and incongruent conditions	• Stroop (1935)
		• If unavailable: incongruent condition accuracy	
		• If accuracy was at ceiling, corresponding RTs were coded	
	Antisaccade task	• Error saccades toward the target	• Everling and Fischer (1998)
Working memory	n-back[a]	• d', difference between hits and false alarms, or overall accuracy	• Jaeggi, Buschkuehl, Perrig, and Meier (2010); Kane, Conway, Miura, and Colflesh (2007)
		• If unavailable: omissions or hit rate	
		• When the accuracy measures from the list above were at ceiling, RTs were coded instead[b]	
	Rapid Information Processing	• Processing rate (digits presented per minute)	• Fillmore et al. (2005)
	Sternberg	• Load effect	• Sternberg (1966)
		• If unavailable: accuracy	
		• If accuracy was at ceiling, corresponding RTs were coded[b]	
	Digit Span	• Accuracy	• The Psychological Corporation (2002)
		• If unavailable: longest length of correctly reported item	
	CANTAB Spatial Working Memory	• Within and between search errors	• Owen, Downes, Sahakian, Polkey, and Robbins (1990)
		• If unavailable: within- or between-search errors	

Corrected Proof

Table 1. (continued)

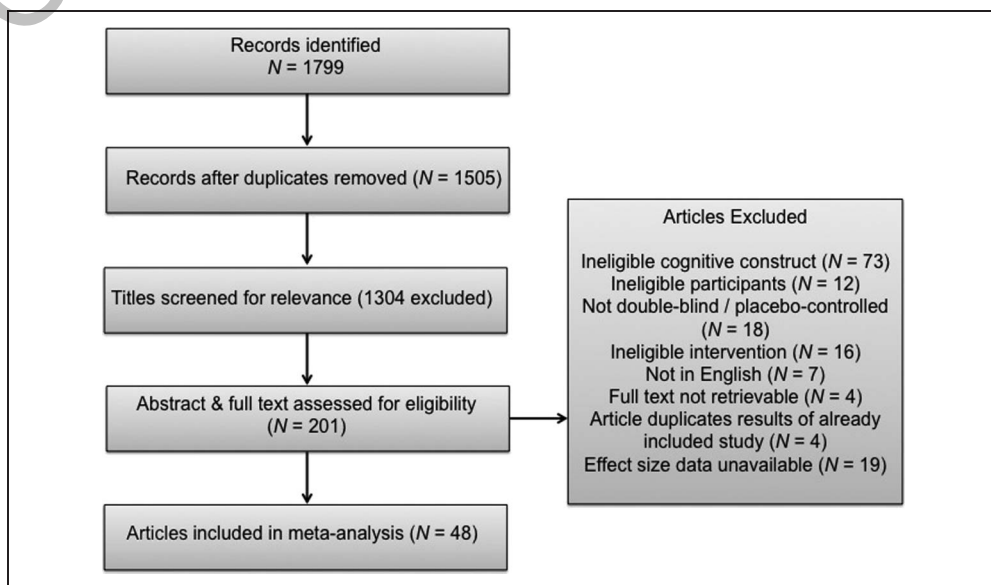| Cognitive Construct	Task	Eligible Measure(s)	Reference Supporting Choice of Measure
	Spatial Delayed Response	• Accuracy	• Postle, Jonides, Smith, Corkin, and Growdon (1997)
	Other WM measures	• d' or accuracy	• Jaeggi et al. (2010), Kane et al. (2007)
		• For spatial tasks: error to position and positional fit	
		• If unavailable: omission errors	
Immediate and delayed episodic memory	Recall (free and cued) and recognition tests	• Sensitivity (d' or a'), proportion of hits minus proportion of false alarms, accuracy, or number of trials to criterion	• Henson, Rugg, Shallice, and Dolan (2000)
		• If unavailable: hit rate	

[a]Only data from 2- and 3-back tasks were coded, excluding data from 0-back conditions (which capitalize on sustained attention more than working memory) and 1-back conditions (which, while taxing some working memory components, such as online maintenance, minimally tax other facets of working memory, such as monitoring and manipulation). Thus, we only included the *n*-back conditions that maximized the possibility of detecting drug effect and minimized the possibility of ceiling effects.

[b]When the data did not allow an inference about the presence or absence of ceiling or floor effects (i.e., the floor or the ceiling of the scale was not clearly defined or apparent), both accuracy and RT measures were coded.

enhancement potential. To our knowledge of the enhancement literature, no previous study has compared the enhancement effects of these two medications.

(2) Dose (low vs. high): The cognitive effects of stimulants are dose dependent (e.g., Cooper, et al. 2005). In examining the role of dose in enhancement effects, we defined a "high" dose as amphetamine ≥ 20 mg and methylphenidate ≥ 40 mg. Doses below these benchmarks were coded as "low."

(3) Caffeine restriction (present vs. absent): We explored the possibility that stimulants may be especially helpful in countering caffeine withdrawal, while possibly having limited effects on non-caffeine-withdrawn individuals.

The presence or absence of instructions to abstain from caffeinated beverages on the day of the experiment was coded as a possible moderator.

(4) Gender distribution in the sample (percent male participants): In the past, higher rates of enhancement use have been reported among male students (e.g., Tèter, Mccabe, Cranford, Boyd, & Guthrie, 2005), and differences in stimulants' subjective effects have been shown to vary as a function of gender and menstrual phase (White, Justice, & DeWit, 2002). The percentage of men in the study sample was therefore tested as a moderator.

(5) Risk of ceiling or floor effects (suspected vs. not): Ceiling and floor effects could attenuate the estimated

Figure 1. Process of determining study eligibility.


```
Records identified
N = 1799
        ↓
Records after duplicates removed (N = 1505)
        ↓
Titles screened for relevance (1304 excluded)
        ↓
Abstract & full text assessed for eligibility
(N = 201)                                    → Articles Excluded

                                               Ineligible cognitive construct (N = 73)
                                               Ineligible participants (N = 12)
                                               Not double-blind / placebo-controlled
                                               (N = 18)
                                               Ineligible intervention (N = 16)
                                               Not in English (N = 7)
                                               Full text not retrievable (N = 4)
                                               Article duplicates results of already
                                               included study (N = 4)
                                               Effect size data unavailable (N = 19)
        ↓
Articles included in meta-analysis (N = 48)
```

effect size. In these analyses, we examined whether the effect size in studies with no restriction of range differed from the effect size estimate in studies with suspected floor or ceiling effects. A study was coded as being at risk of range restriction if the larger among the means in the drug and placebo conditions was less than 1 *SD* away from the scale's floor or if the smaller mean was less than 1 *SD* away from the scale's ceiling. In case of moderation, our goal was to focus on the effect size estimate in the group of studies without suspected floor or ceiling effects.

(6) Reason to publish if drug effects are null (present vs. absent): For the purpose of assessing publication bias for reports of behavioral effects of stimulants, we distinguished between effect sizes from studies that focused only on the effects of amphetamine or methylphenidate on healthy individuals and studies that also included clinical groups, other drugs, or nonbehavioral measures such as PET, fMRI, EEG, or ERP. We expected that smaller stimulant enhancement effects would be published in the context of studies addressing multiple questions (because of the higher likelihood of a positive finding given multiple measures and the greater resources invested in testing clinical populations, measuring neural activity, and administering multiple interventions).

(7) For working memory tasks: Stimulus type (verbal vs. visual vs. spatial) and type of working memory subprocess measured (maintenance vs. maintenance plus manipulation). Different stimuli types and working memory subprocesses are supported by different brain structures, raising a possibility for differential susceptibility to stimulant effects (e.g., Martinussen, Hayden, Hogg-Johnson, & Tannock, 2005; Wager & Smith, 2003).

Effect sizes were calculated using means and standard deviations. Where these descriptives were not presented, we estimated them from published graphs. We favored descriptive over inferential statistics based on previous research showing that, in repeated-measures designs (most of the included studies), effect size estimates from descriptive statistics are less biased than those from repeated-measures inferential statistics (Dunlap, Cortina, Vaslow, & Burke, 1996). In the absence of descriptive data, we estimated effect sizes from *F* (provided *df* = 1), *t*, and/ or *p* values. If effect sizes were directly reported, we estimated their confidence intervals for $r_{equivalent}$ (Rosenthal & Rubin, 2003) and converted the values to *d*. When data were unavailable from either reports or from graphs, they were requested from authors.

The second author independently coded a random sample of 44% of the means and standard deviations (including data, estimated from graphs) in the placebo and drug conditions. Analyses of reliability showed excellent agreement (two-way mixed-model intraclass correlation coefficient for absolute agreement > .99 in all cases).

Handling of Missing Data

Effect size data could not be retrieved or calculated from 19 reports. We performed all meta-analyses excluding all missing data. We did not impute data in missing cells because we had no reason to infer either zero or average sizes of these unreported effects (Cooper, 2010). In other words, we had no sufficient data to ensure that these analyses would improve our effect size estimates, instead of introducing error.

Statistical Methods

Effect Size Metrics

Hedge's *g* was used as the primary effect size measure, whereby a value of .2 is conventionally considered small, .5 is considered medium, and .8 is considered large. Hedge's *g* is obtained by multiplying the effect size Cohen's *d* by a coefficient *J*, which corrects for the tendency for studies with small sample sizes to bias the mean effect size positively because of publication bias: $J = 1 - \left(\frac{3}{4 \times df - 1}\right)$. In combining effect sizes, each was weighted by an estimate of its precision, that is, the inverse of the squared standard error of the effect size.

For within-subject designs, employed in most of the meta-analyzed articles, we have the option of calculating the effect size in two ways. Typically, for such designs, a measure of performance change is scaled by units of variability of change. This addresses the question, "How much drug-related benefit can one expect, relative to the variability of change scores in the sample?" Alternatively, the effect size can be expressed as the size of the drug treatment effect on performance, measured in units of performance variability, as in between-subject designs. Specifically, using this approach, the difference in performance attributable to the drug is measured against the standard deviation of the sample's placebo performance. In effect, this addresses the question, "how far along the distribution of normal performance does the drug push subjects?" This question is very appropriate to the study of cognitive enhancement when used to gain a competitive edge relative to an unmedicated population. In addition, some authors have argued that "subject differences are always of theoretical interest" because "they are present in the population to which we want to generalize," justifying the calculation of effect sizes from either within- or between-subject designs in units of variability (Cortina & Nouri, 2000, p. 49). We report both types of effect size analysis here, placing primary emphasis on effect sizes measured relative to normal variability.

To conduct our primary analyses, we included research with both within- and between-subject designs and relied on effect sizes calculated as shown below. These formulas, typically used for between-subject designs, were modified so that the observed standard deviations in the placebo condition are entered in the

analyses as values for both medication and placebo conditions. In particular:

$$g = J \times \frac{M_{\text{DRUG}} - M_{\text{PBO}}}{SD_{\text{POOLED}}}$$

where $SD_{\text{POOLED}} = \sqrt{\frac{\left((N_{\text{DRUG}} - 1)SD_{\text{PBO}}^2 + (N_{\text{PBO}} - 1)SD_{\text{PBO}}^2\right)}{N_{\text{DRUG}} + N_{\text{PBO}} - 2}}$ and $SE = SD_{POOLED} \times \sqrt{\frac{1}{N_{DRUG}} + \frac{1}{N_{PBO}}}$.

In these analyses, t, F, and p values were used to derive effect sizes from between-subject designs, using the following formulas: Hedge's $g = J \times \frac{t\sqrt{2}}{\sqrt{\frac{2N_1 N_2}{N_1 + N_2}}}$; $SE = \sqrt{\frac{1}{N_1} + \frac{1}{N_2} + \frac{d^2}{2(N_1 + N_2)}}$ (after converting p and F to t).

Inferential statistics from within-subject designs were not included in these analyses because they inherently reflect drug effects relative to variability of change, rather than relative to performance variability.

Our secondary analyses focused on the change score (drug minus placebo), specifically the average benefit because of drug, relative to variability of change within the sample. Only within-subject designs contained information relevant to this question. To calculate Hedge's g for change in within-subject designs, the following formulas were used.

$$g = J \times \left(\frac{(M_{\text{DRUG}} - M_{\text{PBO}}) \times \sqrt{2(1 - Corr)}}{SD_{\text{DIFF}}} \right)$$

where $SD_{\text{DIFF}} = \sqrt{SD_{\text{DRUG}}^2 + SD_{\text{PBO}}^2 - 2Corr SD_{DRUG}SD_{PBO}}$ and $SE = \frac{SD_{\text{DIFF}}}{\sqrt{N}}$.

Alternatively, Hedge's $g = J \frac{t}{\sqrt{N}}$ and $SE = \frac{1}{\sqrt{N}} \times \sqrt{1 + \frac{d^2}{2}}$. These formulas require the value of the correlation between repeated measures, which were not reported in the published studies. These values, necessary to adjust for the dependency between repeated measures in effect size calculations, were estimated based on similar data sets.[1,2]

Handling of Studies with More than One Effect Size

One of the assumptions of meta-analysis is that each effect size comes from an independent sample. If this assumption is violated by the inclusion of more than one effect size per study, between-study variance will be underestimated, and the significance of the summary effect size will be overestimated. The following steps were taken to reduce the available data to a single effect size per study.

(1) When effect sizes for more than one construct per study were available, data on each construct (i.e., inhibition, working memory, and short- and long-term episodic memory) were separated in an individual meta-analysis.
(2) When multiple doses of a drug were compared with placebo within the same study, effect size data from all doses were coded and averaged.

(3) When, in a given study, effect sizes were reported for more than one eligible task and/or measure per construct, a single average effect size estimate per construct was obtained.
(4) When outcome data were available from various time intervals after the administration of the drug (e.g., when inhibitory control was tested 1, 2, and 3 hr after drug administration or when long-term episodic memory was measured at various retention intervals), the average effect size was entered in the main analyses.

Fixed vs. Random Effects Model

A fixed effects model assumes that the only source of effect size variability is sampling error. It therefore produces an effect size estimate that describes the analyzed studies but cannot be generalized to other trials. By contrast, in a random effects model, variability is assumed to arise from both sampling error and between-study variability. Effect sizes derived from this model can be generalized to research outside of the analyzed studies. For the present meta-analysis, we selected a random effects model because of the variability between individual studies in each meta-analysis (different drugs, doses, waiting times between drug administration and testing, measures of each specific cognitive function, individual differences between samples) and also because we wanted to generalize the findings beyond the examined research.

Estimation of Heterogeneity

Tests for heterogeneity determine whether the dispersion of the individual effect sizes around their mean value is greater than predicted solely on the basis of subject-level sampling error. One of the tests employed uses the Q statistic, which, if significant, rejects a null hypothesis of homogeneity. The second test, based on the I^2 statistic, generates an estimate of the between-study variance as a percentage of the total variance (between subjects plus subject level). Conventions for low, moderate, and high heterogeneity correspond to I^2 values of 25, 50, and 75, respectively (Lipsey & Wilson, 2001).

Moderator Analyses

Most commonly in the literature, moderator analyses are conducted only after a finding of significant heterogeneity. In contrast to this approach, we decided to conduct moderator analyses regardless of the results of the heterogeneity tests because a homogeneous set of findings may emerge either in the absence of moderators or in the presence of moderators whose effects cancel each other out.

We examined the effect of the dichotomous moderators described earlier using mixed effects analyses. This analytical model assumes that the effect size variation is due to a combination of systematic associations between

moderators and effect sizes, random differences between studies, and subject-level sampling error. Finally, the moderating role of gender composition (measured as percent male) was examined through meta-regression, given the continuous nature of this moderator.

A feature of the data on some moderator variables demanded the following modification in some of the analyses. When analyzing the moderating role of dose, ceiling/floor effects, working memory stimulus type, and working memory subprocess, there were several cases of more than one level of the moderating variable for per study (e.g., this occurred when more than one drug dose was administered per sample or when floor/ceiling effects were suspected for one outcome within a study but not for another). In these cases, we relied on two approaches to analysis. First, to satisfy the assumption of independence between effect sizes, we excluded studies that included data on more than one level of each moderator variable. In a second version of the analyses, we used the shifting-unit method of analysis (Cooper, 2010). The shifting-unit method allows violation of the assumption of meta-analysis in which a study can contribute an effect size to each level of the moderator (e.g., high and low doses). The advantage of the first approach is that the analysis assumptions remain unviolated; the advantage of the second approach is that it makes use of maximum possible data points. The findings based on the two approaches were in agreement, so we only report data based on the second one.

Publication Bias

Publication bias refers to the greater tendency of studies with significant results to be published than nonsignificant findings. Publication bias can therefore bias the results of meta-analyses because the more significant findings typically have larger effect sizes than those remaining in file drawers (Lipsey & Wilson, 2001). To minimize bias in the current meta-analysis, we made efforts to locate and retrieve unpublished data (see Search Strategies above). In addition, we used three methods to assess the evidence for publication bias and the stability of the effect size estimates and to determine unbiased effect sizes: funnel plots, fail-safe N, and trim and fill (Lipsey & Wilson, 2001). These analyses were conducted without correcting effect sizes by the factor J, as described earlier. Only data from published reports were included in these analyses.

A funnel plot permits a qualitative test of publication bias by showing the effect sizes of the analyzed studies plotted against an estimate of those studies' precision (the inverse of standard error of the effect size in our graphs). Effect size estimates from more accurate studies (toward the top of the graph) should cluster closely around the true effect size, whereas effect sizes from less accurate studies should appear more broadly dispersed below. In the absence of publication bias, the more broadly dispersed effect size estimates should extend in a roughly symmetrical arrangement to either side of the more accurate

estimates. A negative skew, where points in the lower left quadrant appear to be missing, is consistent with the operation of publication bias.

In cases of publication bias, the trim-and-fill procedure calculates an unbiased estimate of the effect size. In this procedure, the most extreme positive effects are removed ("trimmed") from analysis, and a mirror image of the trimmed effect sizes with the opposite direction is then imputed. Unbiased estimates of the overall effect size and its variance are calculated, respectively, from the trimmed and filled data.

The fail-safe N indicates the number of studies with a zero effect size that, if added to the analysis, would render the obtained mean effect size nonsignificant. The value of fail-safe N is considered large (and publication bias, an unlikely influence on the effect size estimate) if it exceeds $5k + 10$, where k is the number of meta-analyzed studies (Rothstein, Sutton, & Borenstein, 2006).

Tests for Outliers

The presence of outlier effect sizes was assessed through the sample-adjusted meta-analytic deviancy (SAMD) statistic. For each study, the value of this statistic represents the difference between this study's effect size and the point estimate of the effect size uninfluenced by this study, a difference weighed by the relevant variance terms (Huffcutt & Arthur, 1995). An effect size was considered an outlier if it met both of the following two criteria (Sockol, Epperson, & Barber, 2011): First, in a scree plot of the distribution of absolute SAMD values, it deviates markedly from the slope (Huffcutt & Arthur, 1995). Second, it falls in the top or bottom 2.5% of the SAMD distribution (which approximates a t distribution). This conservative, two-pronged method for outlier detection was chosen because outliers could result from either error or true between-study variation (Sockol et al., 2011).

Software

The data were analyzed primarily using Comprehensive Meta-Analysis 2.0, with the exception of meta-regression analyses, completed in R 3.0.0.

RESULTS
Overview of Results

We report meta-analyses for the effects of stimulants on the four constructs of interest: inhibitory control, working memory, short-term episodic memory, and delayed episodic memory. Two sets of results are presented, corresponding to the two different ways of measuring effect sizes from within-subject designs described earlier. For each cognitive construct, we first present meta-analyses of within- and between-subject studies combined, measuring the size of the drug effect relative to variability in the

normal population. We then present the effect sizes estimated in separate meta-analyses for within-subject and matched-group studies using the formula for within-subject effect sizes described earlier. For the main analyses, we also report the results of moderator analyses and three measures related to publication bias. In reporting our secondary analyses, we do not detail the results of moderator and publication bias analyses, which, in all cases, were qualitatively similar to the results in our main analyses. Most effect sizes were small. Evidence of publication bias emerged in two cognitive domains. Characteristics of all effect sizes (outcomes, magnitude of effect, sample sizes, values of moderator variables) are presented in Tables 2–5.

Stimulants' Effects on Healthy Inhibitory Control

Twenty-five studies (including two unpublished) reported sufficient data to calculate the size of stimulants' effect on inhibitory control. After examining the values of the SAMD statistic, no value fell within the top or bottom 2.5% of the distribution or notably deviated from the relatively flat line of the scree plot. Not all of these studies were suitable for each analysis, that is, a study whose effect size was derived from a repeated-measures t value was excluded from analyses relative to normal variability; and a between-subject study was excluded from analyses relative to variability of change. Data for calculating effect sizes relative to normal variability were available from 24 studies (see Table 2); effect size relative to variability of gain scores was also estimated from 24 studies.

Stimulants' mean effect on inhibitory control, when measured relative to normative variability of performance, was small but significantly different from zero: Hedge's $g = 0.20$, 95% CI [0.11, 0.30]. Effect size measured relative to the variability of gain scores was similarly small and significantly different from zero: Hedge's $g = 0.19$, 95% CI [0.11, 0.26]. No evidence for between-study heterogeneity emerged: $Q(23) = 7.82, p > .99, I^2 = 0.00$. Moderator analyses indicated that none of the candidate moderators impacted significantly the stimulant effects on cognition (all $ps > .20$).

A funnel plot based only on the published studies ($N = 22$) showed no evidence for publication bias: The distribution of studies was roughly symmetrical (Figure 2). The trim-and-fill procedure led to the exclusion of no study, and the adjusted effect size estimates remained the same as reported above. However, the fail-safe N method indicated that 39 studies (less than two studies per each published report) with an effect size of zero would nullify the obtained results. Taken together, the lack of negative skew in the funnel plot and the robustness of the effect size estimate to trim-and-fill adjustment converge to suggest that the effect estimate obtained for inhibitory control is most likely not affected by publication bias. In other words, there is no evidence to suspect that the relatively modest number of studies needed to nullify the result has remained in file drawers.

Stimulants' Effects on Healthy Working Memory

Effect size data on stimulants' effects on working memory were available from 23 studies, three of which were unpublished. None of the effect sizes were outliers by our criteria. Relevant statistics for calculating effect size relative to normal variability were available from 20 studies (Table 3). Effect size relative to variability of gain scores was calculated based on 23 studies with within-subject or matched-group designs.

Our main analyses indicated a near-significant small stimulant effect on working memory: Hedge's $g = 0.13$, 95% CI [−0.02, 0.27]. When measured relative to variability of the gain scores, the effect size was again estimated to be small but, this time, reached significance: $g = 0.13$, 95% CI [0.06, 0.20]. There was no significant evidence for heterogeneity: $Q(19) = 7.74, p = .99, I^2 = 0.00$. Moderator analyses were performed, but no evidence emerged for moderation by any of the examined variables.

The funnel plots, based on published studies only (Figure 3), showed slightly negative skew. The trim-and-fill procedure trimmed 4 data points, reducing the above-reported effect size to a nonsignificant trend of $d = 0.06$, 95% CI [−0.08, 0.20]. Because the gain score effect size was significant, whereas the primary effect size was not, here, we also report the trim-and-fill results from our secondary analyses, where the effect size was again reduced to nonsignificant: $d = 0.06$, 95% CI [−0.03, 0.15], given a negatively skewed funnel plot. Taken together, the trim-and-fill correction and the skew of the funnel plot suggest the presence of publication bias. Fail-safe N analyses were obviated by the lack of significance in the obtained effect size estimate.

Stimulants' Effects on Healthy People's Short-term Episodic Memory

Fourteen effect sizes (one unpublished) were considered for inclusion in the meta-analysis. Two SAMD values, equaling −8.53 (Burns, House, Fensch, & Miller, 1967) and 2.18 (Zeeuws, Deroost, & Soetens, 2010a), exceeded the cutoff for exclusion and deviated markedly from the relatively flat line on the scree plot of absolute SAMD values. Therefore, these studies were excluded from further analyses after confirming correct data entry.

On the basis of 12 studies (see Table 4), the mean effect of stimulants on short-term episodic memory, relative to normal variation of performance, was small but significant: Hedge's $g = 0.20$, 95% CI [0.01, 0.38]. This was similar to the result observed when the effect size was measured relative to variability of gain scores (12 studies): Hedge's $g = 0.22$, 95% CI [0.09, 0.35]. No evidence for heterogeneity emerged in our main analyses: $Q(11) = 4.44, p = .96, I^2 = 0.00$. Moderator analyses indicated no significant influence of any of the examined moderators (all $ps > .64$).

Table 2. Stimulant Enhancement of Inhibitory Control: Effect Sizes and Study Characteristics

Study	N	Target of Recruitment	Age (M)	Age (Range)	% Male	Education (Years)	Mental Health Assessment	Caffeine Restriction	Drug	Dose (mg)	Dose Coding	Test	Design	Floor or Ceiling?	Other Reason to Publish?	Hedge's g
Acheson and de Wit (2008)	28	Not specified	23.45	18–45	54.54	≥12	SCID-I, SCL–90, MAST	No	Amp	20	High	Stop signal task	Within-subject	Not suspected	No	0.23
Agay, Yechiam, Carmel, and Levkovitz (2010)	25	Local community	32.65	21–50	46.15	M = 14.4	SCID-I, ASRS, CAARS, WURS	No	Mph	15	Low	TOVA (commissions)	Between-subject	Possible	No	0
Allman et al. (2010)	24	Not specified	Not reported	18–34	70.83	Not reported	SCID-I	No	Amp	21	High	Antisaccade task	Within-subject	Not suspected	No	0.35
Barch and Carter (2005)	22	Local community	36.6	Not reported	55	M = 16	SCID, family history of psychosis	No	Amp	17.5	Low	Stroop	Within-subject	Not suspected	No	0.23
Costa et al. (2013)	46	University and local community	23.65	18–30	100	Not reported	Clinician interview (not specified)	Yes	Mph	40	High	Stop signal task, go/no-go	Within-subject	Not suspected	No	0.07
de Bruijn, Hulstijn, Verkes, Ruigt, and Sabbe (2004)	12	not specified	22.58	19–39	58.33	Not reported	Not described	No	Amp	15	Low	Flanker	Within-subject	Not suspected	No	0.22
De Wit (2012)	207	Not specified	Not reported	18–35	52.43	≥12	SCID-I, SCL-90	No	Amp	5, 10, 20	Both	Stop signal task	Within-subject	Not suspected	No	0.21
De Wit et al. (2000)	20	University and local community	25.9	21–35	70	≥12	Semi-structured psychiatric screening, SCL-90	No	Amp	10, 20	Both	Stop signal task	Within-subject	Not suspected	No	0.28
De Wit et al. (2002)	36	University and local community	24	18–44	50	>12	Semistructured psychiatric interview, SCL90, MAST	No	Amp	10, 20	Both	Stop signal task, go/no-go	Within-subject	Not suspected	No	0.35
Engert, Joober, Meaney, Hellhammer, and Pruessner (2009)	43	University	22.2	Not reported	100	Not reported	Mini-SCID	Yes	Mph	20	Low	WCST	Between-subject	Not suspected	No	0.11
Farah (2012)	15	University and local community	Not reported	Not reported	25	Not reported	Self-reported history of diagnosis	No	Amp	10	Low	Flanker	Within-subject	Not suspected	No	0.22

Study	N	Sample source	Age	Range	%		Screening		Drug	Dose		Task	Design			ES
Fillmore et al. (2005)	22	Local community	21.5	18–30	45.45	$M = 14.1$ (range: 12–17)	Not described	Yes	Amp	7.5, 15	Low	Stop signal task	Within-subject	Not suspected	No	0.1
Hamidovic et al. (2009)	93	Not specified	22.3	18–35	53.76	≥12	Structured clinical interview, SCL-90, MAST	No	Amp	5, 10, 20	Both	Stop signal task	Within-subject	Not suspected	Yes	0.2
Hester et al. (2012)	27	University	22	18–35	100	Not reported	MINI, Kessler K10	Yes	Mph	30	Low	Go/no-go (modified)	Within-subject	Not suspected	Yes	0.18
Ilieva et al. (2013)	43	University and local community	24	21–30	50	Not reported	Self-reported history of diagnosis	No	Amp	20	High	Go/no-go, flanker	Within-subject	Not suspected	No	0.05
Kelly et al. (2006)	20	University and local community	21.7	Not reported	50	$M = 14.25$	Not described	No	Amp	8, 15	Low	Stop signal task	Within-subject	Not suspected	No	0.09
Linssen, Vuurman, Sambeth, and Riedel (2012)	19	Local community	23.4	19–37	100	Not reported	Not described	No	Mph	10, 20, 40	Both	Stop signal task	Within-subject	Not suspected	No	0.35
Mattay et al. (1996)	8	Not specified	25	22–32	50	Not reported	Not described	Yes	Amp	17.5	Low	WCST	Within-subject	Not suspected	Yes	0.08
Moeller et al. (2012)	15	Local community	38.9	Not reported	93.33	$M = 13.9$	SCID-I, Addiction Severity Index	No	Mph	20	Low	Stroop	Within-subject	Not suspected	Yes	0.37
Nandam et al. (2011)	24	University	23	18–35	100	Not reported	M.I.N.I., Kessler K10	No	Mph	30	Low	Stop signal task	Within-subject	Not suspected	Yes	0.58
Pauls et al. (2012)	16	University	23.6	19–30	100	Not reported	ASRS; assessment of other psychopathology not described	Yes	Mph	40	High	Stop signal task	Within-subject	Not suspected	Yes	0.32
Servan-Schreiber, Carter, Bruno, and Cohen (1998)	8	University	Not reported	24–39	50	Not reported	SCID-I	No	Amp	17.5	Low	Flanker	Within-subject	Not suspected	No	0.72
Sofuoglu, Waters, Mooney, and Kosten (2008)	10	Local community	27.7	Not reported	58.33	Not reported	Psychiatric examination (not specified)	No	Amp	20	High	Go/no-go	Within-subject	Possible	Yes	−0.36
Theunissen, Elvira, van den Bergh, and Ramaekers (2009)	16	Local community	21.8	19–29	31.25	Not reported	Not described	Yes	Mph	20	Low	Stop signal task	Within-subject	Not suspected	Yes	−0.01
Overall effect size																0.20

Table 3. Stimulant Enhancement of Working Memory: Effect Sizes and Study Characteristics

Study	N	Target of Recruitment	Age (M)	Age (Range)	% Male	Education (Years)	Mental Health Assessment	Caffeine Restriction	Drug	Dose (mg)	Dose Coding	Test	Stimulus Modality	Type of WM Processing	Design	Floor or Ceiling?	Other Reason to Publish?	Hedge's g
Agay et al. (2010)	26	Local community	32.65	21–50	56.25	M = 14.4	SCID-I, ASRS, CAARS, WURS	No	Mph	15	Low	Digit span	Verbal	Maintenance only, maintenance + manipulation	Between-subject	Not suspected	Yes	0.22
Agay (2012)	19	Local community	Not reported	20–40	Not reported	Not reported	SCID-I	No	Mph	20	Low	Digit Span, CANTAB Spatial WM	Spatial, verbal	Maintenance only, maintenance + manipulation	Within-subject	Not suspected	Yes	0.23
Barch and Carter (2005)	22	Local community	36.6	Not reported	55	M = 16	SCID, family history of psychosis	No	Amp	17.5	Low	Spatial working memory (8-sec delay, single and dual tasks)	Spatial	Maintenance only, maintenance + manipulation	Within-subject	Not suspected	Yes	0.10
Dorflinger (2005)	20	Not specified	27.9	24–34	Not reported	M = 16.5 (range: 12–22)	Not described	No	Mph	14, 28	Low	2-back, 3-Back	Verbal	Maintenance + manipulation	Within-subject	Not suspected	Yes	0.15
Farah (2012)	16	University and local community	Not reported	Not reported	25	Not reported	Self-reported history of diagnosis	No	Amp	10	Low	Digit Span, 2-back	Visual, verbal	Maintenance only, maintenance + manipulation	Within-subject	Not suspected	No	−0.10
Fillmore et al. (2005)	22	Local community	21.5	18–30	45.45	M = 14.1 (range: 12–17)	Not described	Yes	Amp	7.5, 15	Low	Rapid Inf. processing	Verbal	Maintenance + manipulation	Within-subject	Not suspected	No	0.25
Ilieva et al. (2013)	43	University and local community	24	21–30	50	Not reported	Self-reported history of diagnosis	No	Amp	20	High	Digit Span, 2-back	Visual, verbal	Maintenance only, maintenance + manipulation	Within-subject	Not suspected	No	0.01
Kelly et al. (2006)	20	University and local community	21.7	Not reported	50	M = 14.25	Not described	No	Amp	7.5, 15	Low	Rapid Inf. Processing	Verbal	Maintenance + manipulation	Within-subject	Not suspected	No	0.38
Linssen et al. (2012)	19	Local community	23.4	19–37	100	Not reported	Not described	Yes	Mph	10, 20, 40	Low, high	Spatial working memory	Spatial	Maintenance + manipulation	Within-subject	Not suspected	No	0.41
Marquand et al. (2011)	15	University and local community	Not reported	20–39	100	Not reported	Interview (not specified)	Yes	Mph	30	Low	Spatial working memory (unrewarded condition)	Spatial	Maintenance only	Within-subject	Not suspected	Yes	−0.11

Study	N																	
Mattay et al. (2000)	10	Not specified	Not reported	Not reported	80	Not reported	Not described	Yes	Amp	17.5	Low	2-back, 3-back	Verbal	Maintenance + manipulation	Within-subject	Not suspected	Yes	−0.04
Mattay et al. (2003)	26	Not specified	Not reported	<45	40.74	Not reported	Not described	No	Amp	17.5	Low	2-back, 3-back	Verbal	Maintenance + manipulation	Within-subject	Not suspected	Yes	−0.23
Mehta et al. (2000)	10	Not specified	34.8	Not reported	100	Not reported	Not described	No	Mph	40	High	CANTAB Spatial Working Memory	Spatial	Maintenance + manipulation	Within-subject	Not suspected	Yes	0.27
Mintzer and Griffiths (2003)	20	Not specified	30	19–52	70	12–16, M = 14	Not described	No	Amp	20	High	Digit Recall, 2-back	Verbal	Maintenance only, maintenance + manipulation	Within-subject	Possible for one measure	Yes	0.25
Mintzer and Griffiths (2007)	18	Not specified	23	18–39	61.11	12–21, M = 15	Not described	No	Amp	20, 30	High	2-back, 3-back, modified Sternberg	Verbal	Maintenance + manipulation	Within-subject	Not suspected	Yes	0.15
Oken, Kishiyama, and Salinsky (1995)	23	Not specified	25	21–39	47.83	Not reported	Not described	Yes	Mph	14	Low	Digit Span	Verbal	Maintenance only, maintenance + manipulation	Within-subject	Not suspected	Yes	−0.14
Ramasubbu, Singh, Zhu, and Dunn (2012)	13	University	28	Not reported	38.46	Not reported	Not described	Yes	Mph	20	Low	2-back	Verbal	Maintenance + manipulation	Within-subject	Not suspected	Yes	0.62
Schmedtje, Oman, Letz, and Baker (1988)	8	Not specified	Not reported	Not reported	Not reported	Not reported	Not described	No	Amp	5	Low	Digit Span, pattern memory	Visual, verbal	Maintenance only, maintenance + manipulation	Within-subject	Not suspected	Yes	0.30
Silber, Croft, Papafotiou, and Stough (2006)	20	Local community	25.4	21–32	50	≥11	Clinician screening (not specified)	Yes	Amp	5	Low	Digit Span	Verbal	Maintenance only, maintenance + manipulation	Within-subject	Not suspected	Yes	0.18
Studer et al. (2010)	11	Not specified	29.7	Not reported	45.45	Not reported	Not described	No	Mph	20	Low	Visual working memory	Visual	Maintenance + manipulation	Within-subject	Possible for one measure	Yes	0.10
Overall effect size																		0.13

Table 4. Stimulant Enhancement of Short-term Episodic Memory: Effect Sizes and Study Characteristics

Study	N	Target of Recruitment	Age (M)	Age (Range)	% Male	Education (Years)	Mental Health Assessment	Caffeine Restriction	Drug	Dose (mg)	Dose Coding	Test	Retention Interval	Design	Floor or Ceiling?	Other Reason to Publish?	Hedge's g
Farah (2012)	16	University and local community	Not reported	Not reported	25	Not reported	Self-reported history of diagnosis	No	Amp	10	Low	Word recall, word recognition, face recognition	30 min	Within-subject	Not suspected	No	0.07
Fleming, Bigelow, Weinberger, and Goldberg (1995)	17	Local community	27.5	Not reported	52.94	$M = 15.8$	SCID-I and II	No	Amp	20	High	Paired associates, Rey Verbal Learning Test	Immediate	Within-subject	Possible for one measure	No	0.16
Linssen et al. (2012)	19	Local community	23.4	19–37	100	Not reported	Not described	No	Mph	10, 20, 40	Low, high	Word recall, word recognition	Immediate; 30 min	Within-subject	Possible for one measure	No	0.20
Soetens et al. (1995), Study 1	18	Not specified	Not reported	19–25	100	Not reported	Not described	No	Amp	10	Low	Word recall	Immediate	Within-subject	Not suspected	No	0.23
Soetens et al. (1995), Study 2	14	Not specified	Not reported	19–25	100	Not reported	Not described	No	Amp	10	Low	Word recall	Immediate	Within-subject	Not suspected	No	0.39
Soetens et al. (1995), Study 4	12	Not specified	Not reported	19–25	100	Not reported	Not described	No	Amp	10	Low	Word recognition	Immediate	Within-subject	Not suspected	No	0.29
Soetens et al. (1995), Study 5	12	Not specified	Not reported	19–25	100	Not reported	Not described	No	Amp	10	Low	Word recognition	Immediate	Within-subject	Not suspected	No	0.31
Unrug, Coenen, and van Luijtelaar (1997)	12	University	24	19–27	50	Not reported	Not described	Yes	Mph	20	Low	Word recall	20 min	Within-subject	Possible for one measure	Yes	0.32
Willett (1962)	37	Not specified	Not reported	Not reported	0	Not reported	Not described	No	Amp	10	Low	Learning of nonword lists	Immediate	Between-subject	Not suspected	No	0.22
Zeeuws, Deroost, and Soetens (2010b), Study 1	24	University	21.1	18–25	100	Not reported	Not described	No	Amp	10	Low	Word recognition	Immediate	Within-subject	Not suspected	No	−0.17
Zeeuws et al. (2010b), Study 2	16	University	21.4	18–25	100	Not reported	Not described	No	Amp	10	Low	Word recognition	Immediate	Within-subject	Not suspected	No	−0.13
Zeeuws and Soetens (2007)	36	University	Not reported	18–25	100	Not reported	Not described	No	Amp	10	Low	Word recognition	Immediate; 30 min	Within-subject	Not suspected	No	0.45
Overall effect size																	0.20

Corrected Proof

Table 5. Stimulant Enhancement of Long-term Episodic Memory: Effect Sizes and Study Characteristics

Study	N	Target of Recruitment	Age (M)	Age (Range)	% Male	Education (Years)	Mental Health Assessment	Caffeine Restriction	Drug	Dose (mg)	Dose Coding	Test	Retention Interval	Design	Floor or Ceiling?	Other Reason to Publish?	Hedge's g
Brignell, Rosenthal, and Curran (2007)	36	Not specified	22.8	18–35	Not reported	Not reported	Not described	No	Mph	40	High	Recognition memory for narratives	1 hr, 1 day	Between-subject	Not suspected	Yes	0.52
Ilieva et al. (2013)	18	University and local community	24	21–30	50	Not reported	Self-reported history of diagnosis	No	Amp	20	High	Word recall and recognition, face recognition	2 hr	Within-subject	Not suspected	No	0.01
Mintzer and Griffiths (2003)	16	Not specified	30	19–52	70	12–16, M = 14	Not described	No	Amp	20	High	Word recall and recognition	2 hr	Within-subject	Not suspected	Yes	0.24
Mintzer and Griffiths (2007)	20	Not specified	23	18–39	61.11	12–21, M = 15	Not described	No	Amp	20, 30	High	Word recall and recognition	2 hr	Within-subject	Not suspected	Yes	0.33
Soetens et al. (1995), Study 1	44	Not specified	Not reported	19–25	100	Not reported	Not described	No	Amp	10	Low	Word recall	1 day	Within-subject	Not suspected	No	0.71
Soetens et al. (1995), Study 2	18	Not specified	Not reported	19–25	100	Not reported	Not described	No	Amp	10	Low	Word recall	1 hr, 1 day	Within-subject	Not suspected	No	0.58
Soetens et al. (1995), Study 4	14	Not specified	Not reported	19–25	100	Not reported	Not described	No	Amp	10	Low	Word recall	1 day, 2 days, 3 days	Within-subject	Not suspected	No	0.58
Soetens et al. (1995), Study 5	12	Not specified	Not reported	19–25	100	Not reported	Not described	No	Amp	10	Low	Word recognition	1 day, 1 week	Within-subject	Not suspected	No	0.74
Zeeuws et al. (2010b), Study 1	12	University	21.1	18–25	100	Not reported	Not described	No	Amp	10	Low	Word recognition	1 hr, 1 day, 1 week	Within-subject	Not suspected	No	0.69
Zeeuws et al. (2010b), Study 2	24	University	21.4	18–25	100	Not reported	Not described	No	Amp	10	Low	Word recognition	1 hr, 1 day, 1 week	Within-subject	Not suspected	No	0.18
Zeeuws and Soetens (2007)	16	University	Not reported	18–25	100	Not reported	Not described	No	Amp	10	Low	Word recognition	1 hr, 1 day	Within-subject	Not suspected	No	0.80
Overall effect size																	0.45
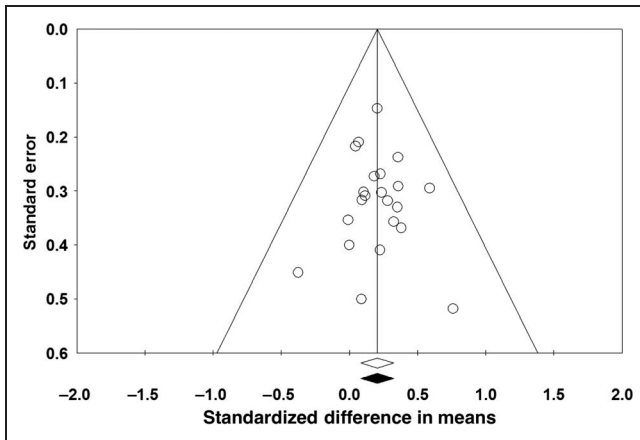
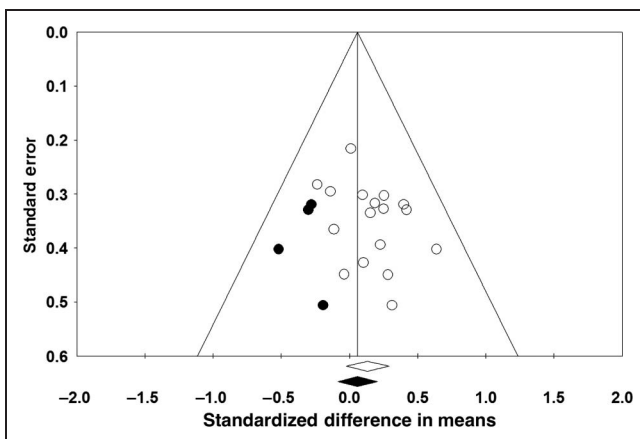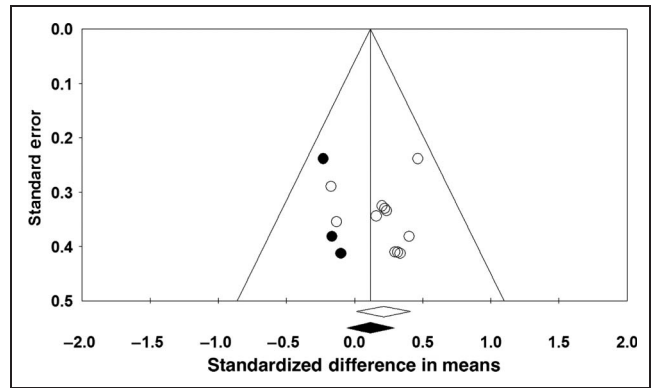Figure 2. Funnel plot of publication bias: inhibitory control.

Figure 4. Funnel plot of publication bias: short-term episodic memory. Darkened data points represent studies "filled" by trim and fill.

A funnel plot, based on the 11 published studies, showed slightly negative asymmetry (Figure 4); yet, inconsistent with publication bias, the largest study had the largest effect. The trim-and-fill procedure trimmed three studies, reducing the effect size estimate to a non-significant $d = 0.12$, 95% CI [−0.06, 0.29]. The fail-safe N procedure showed that a mere two studies with an effect size of zero would be needed to nullify the obtained effect, casting doubt on this effect's robustness.

Stimulant Effects on Healthy People's Delayed Episodic Memory

Twelve effect sizes describing stimulants' effects on delayed episodic memory were reported. One outlier was excluded, given an SAMD value of 3.35 (Zeeuws et al., 2010a), which fell in the top 2.5% of the distribution of SAMD scores.

On the basis of the remaining 11 effect sizes, estimated relative to normal variability (see Table 5), stimulants' mean effect on delayed episodic memory was significantly different from zero and medium in size: $g = .45$, 95% CI [0.27, 0.63]. Similarly, analyses focusing on the

mean gain, relative to the sample's variability of change, showed a medium-sized effect: Hedge's $g = 0.44$, 95% CI [0.26, 0.62]. There was no evidence for significant between-study heterogeneity: $I^2 = 0.00$, $Q(10) = 9.67$, $p = .47$. We found a small but significant moderating effect of gender, $Q(1) = 7.44$, $p < .01$, β = 0.01, with larger drug effects for larger proportions of men in samples. In addition, there was a significant moderating effect of dose, $Q(1) = 5.49$, $p = .02$, indicating a larger effect for the smaller dose, Hedge's $g = 0.64$, 95% CI [0.40, 0.88], than the larger dose, Hedge's $g = 0.20$, 95% CI [−0.08, 0.48]. Note that these moderation effects are confounded with each other and with research group: All studies that used low doses of stimulants came from the same research group, tested only male participants, and tended to include memory tests at longer retention intervals (up to 1 week), whereas among tests of the high drug dose, the percentage of men in the sample ranged between 48% and 70% and retention intervals, with one exception, were 2 hr. No other factors were found to significantly moderate stimulants' effects (all $ps > .52$).

The funnel plot of these studies was negatively skewed, suggesting publication bias (Figure 5). The

Figure 3. Funnel plot of publication bias: working memory. Darkened data points represent studies "filled" by trim and fill.
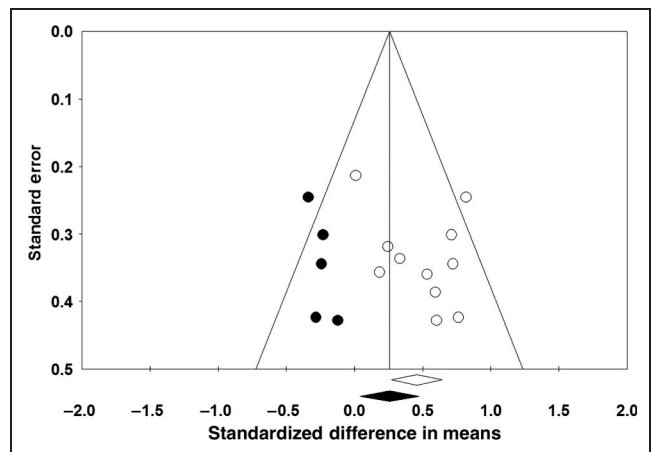

Figure 5. Funnel plot of publication bias: long-term episodic memory. Darkened data points represent studies "filled" by trim and fill.

trim-and-fill method trimmed five studies, reducing the estimated effect size to $d = 0.26$, 95% CI [0.04, 0.47]. According to the fail-safe N procedure, 59 studies were needed to nullify the significance level of the result. The negative skew of the funnel plot, combined with the trim-and-fill correction, suggests the presence of publication bias and indicates that the true effect size may be small. It is important to note, though, that inferences from the funnel plot are qualified by the presence of significant moderation (see Lau, Ioannidis, Terrin, Schmid, & Olkin, 2006). In particular, the studies with the six largest effect sizes came from the same laboratory and tested the effect of a low stimulant dose on male-only samples, in part, over relatively longer retention intervals. Four of the five remaining studies with smaller effect sizes came from other research groups and examined the effects of a high stimulant dose on a mixed-gender sample over relatively shorter delays. Thus, the funnel plot might reflect true publication bias or might be driven by between-study differences. If the latter is the case, the trim-and-fill-adjusted effect size may be underestimating the true effect size (e.g., Peters, Sutton, Jones, Abrams, & Rushton, 2007). Unfortunately, the proposed methods of unconfounding publication bias and moderating factors (e.g., conducting funnel plot analyses within a subgroup of studies) are applicable only to large meta-analyses (see Peters et al., 2010).

DISCUSSION
Summary and Interpretation of Results

Earlier research has failed to distinguish whether stimulants' effects are small or whether they are nonexistent (Ilieva et al., 2013; Smith & Farah, 2011). The present findings supported generally small effects of amphetamine and methylphenidate on executive function and memory. Specifically, in a set of experiments limited to high-quality designs, we found significant enhancement of several cognitive abilities. We found a small but significant degree of enhancement of inhibitory control and short-term episodic memory. Effects on working memory were small and significant in one of our two analyses. Delayed episodic memory was unique in showing a medium-sized effect. However, both working memory and delayed episodic memory findings were qualified by possible publication bias.

Theoretically, the relatively more pronounced effects of delayed episodic memory, in comparison with short-term episodic memory, suggest that stimulants may be affecting most potently memory consolidation in comparison with encoding or retrieval. This conclusion is consistent with previous proposals (e.g., Soetens, Hueting, Casaer, & D'Hooge, 1995; see also McGaugh & Roozendaal, 2009) but, again, qualified by the possibility of publication bias.

Several potentially important moderators were tested because of their scientific relevance for understanding the effects of stimulants on cognition and their practical relevance in determining whether stimulants might be more effective cognitive enhancers under some circumstances than others. Moderator analyses yielded only a few significant findings. Stimulant effects on delayed episodic memory were moderated by gender, with larger effects for samples with more men, and by dosage, with larger effects for smaller doses. Unfortunately, these two moderators were confounded in the studies analyzed and also confounded with research laboratory and retention interval, so we cannot draw firm conclusions about the effects of gender or dose.

Where no effects of moderators were found, this may be because of uncertainty or imprecision in moderator coding, for instance, the dichotomization of drug dose or the possibility of nonlinear relationships between drug effect and dose. Finally, partly for the sake of limiting the number of comparisons and partly because of limited availability of relevant information, we examined only a subset of all relevant moderators. For instance, we did not explore the moderating role of participant age, level of education, waiting time between drug administration and testing, length of testing session, or time of day. Moderators of great interest, which might be expected to affect results based on previous studies but which could not be assessed because of insufficient available data, include individuals' baseline cognitive ability and individuals' variants of dopamine-related genes such as COMT and DRD2 (see Hamidovic et al., 2009; Mattay et al., 2003; but see also Wardle, Hart, Palmer, & de Wit, 2013, for null results). Consistent with the nonmonotonic relation between dopamine activity and performance, there is evidence that stimulants can impair performance in normal individuals who are especially high performing (Farah, Haimm, Sankoorikal, & Chatterjee, 2009; De Wit et al., 2002; De Wit, Crean, & Richards, 2000; Mattay et al., 2000). It remains possible that some individuals who would not qualify for a diagnosis of ADHD could nevertheless benefit from stimulants to a greater degree than indicated by the present results and that some individuals could be impaired.

Could the effects documented here be driven by undiagnosed psychopathology in some participants? One might expect participants with ADHD or depression to perform better on stimulants and participants with anxiety disorders or bipolar disorder to be impaired by these drugs. Unfortunately, few publications included comprehensive, detailed description of procedure through which psychopathology was screened out, making it difficult to assess the quality of assessment. Nevertheless, all reports explicitly described their samples as "healthy" or "nonclinical." Thus, it is possible but unlikely that unrecognized mental illness is responsible for the pattern of obtained findings.

Neuroethical Implications

What do the results reported here imply for neuroethical issues surrounding the use of stimulants for enhancement? Should we be concerned about the fairness of students and workers competing with the help of stimulant drugs?

Is there a genuine benefit to be weighed against the risks of using these prescription drugs for enhancement? The overall small effects of stimulants on healthy people's inhibitory control and working and episodic memory might be taken to mean that these drugs would not deliver a practically significant performance advantage. If so, one might argue that neuroethical discussions are therefore moot at best (and encouraging of a false belief in the drugs' efficacy at worst, e.g., Hall & Lucke, 2010).

The present findings should temper these and other more skeptical assessments of stimulant medications for cognitive enhancement of healthy, cognitively normal individuals. Although the reported effects are smaller than these of some other cognitive enhancement techniques (e.g., mindfulness meditation, for which near-medium effects on inhibitory control have been documented; see Sedlmeier et al., 2012, for a meta-analysis), the present findings show stimulant benefits comparable with the effects of other commonly used enhancement tools (e.g., physical exercise, the cognitive effects of which have been found to be similarly small; see Chang, Labban, Gapin, & Etnier, 2012, for a meta-analysis). Furthermore, small effects can make a difference in academic and professional outcomes. Even on a single occasion, a small effect might make the difference between good and very good performance or between passing a school entrance or licensing examination or failing. It is also possible that these drugs may give a larger boost to cognitive functions not examined here (e.g., sustained attention, processing speed); to people not specifically studied in this meta-analysis (e.g., healthy participants with low cognitive performance or specific genotypes); or to performance under conditions not tested here, for example, fatigue, sleep deprivation, distraction, or repeated stimulant use (e.g., Breitenstein et al., 2006). It is also possible that stimulants enhance cognitive performance in real-world contexts at least in part through effects on users' affective states. They have been found to alter users' emotions about, and interest in, tasks otherwise seen as boring and unrewarding (Ilieva & Farah, 2013; Vrecko, 2013).

The results of this meta-analysis cannot address the important issues of individual differences in stimulant effects or the role of motivational enhancement in helping perform academic or occupational tasks. However, they do confirm the reality of cognitive enhancing effects for normal healthy adults in general, while also indicating that these effects are modest in size.

Notes

1. Correlations were obtained from Ilieva, Boland, and Farah (2013; Flanker, Go/No-Go, *n*-back, Digit Span Backward and Forward, delayed memory for words and faces; 46 participants); Mintzer and Griffiths (2007; *n*-back, Sternberg memory task, delayed memory for words; 18 participants); and Hamidovic, Dlugos, Skol, Palmer, and de Wit (2009), combined with a set of unpublished data from Dr. Harriet de Wit's laboratory (Stop Signal task, 299 participants). When correlations for a given task (e.g., *n*-back) were available from more than one data set, we estimated a composite through meta-analyzing the available correlations based on a random effects model. In the case of tasks for which data on observed correlations were lacking, we imputed an estimate of the correlation for the corresponding cognitive construct obtained through meta-analyzing the available observed correlations for tasks within that construct based a random effects model.
2. To estimate the potential for error in case of inaccurate imputed correlations, we repeated our main analyses three times after imputing uniform correlation values of .2, .5, and .8 across all effect sizes. This led to minimal changes in the reported patterns of findings (largest change in effect size was $g = 0.02$).

REFERENCES

Acheson, A., & de Wit, H. (2008). Bupropion improves attention but does not affect impulsive behavior in healthy young adults. *Experimental and Clinical Psychopharmacology, 16,* 113–123.

Advokat, C. (2010). What are the cognitive effects of stimulant medications? Emphasis on adults with attention-deficit/hyperactivity disorder (ADHD). *Neuroscience & Biobehavioral Reviews, 34,* 1256–1266.

Agay, N. (2012). [Cognitive effects of stimulants]. Unpublished raw data.

Agay, N., Yechiam, E., Carmel, Z., & Levkovitz, Y. (2010). Non-specific effects of methylphenidate (Ritalin) on cognitive ability and decision-making of ADHD and healthy adults. *Psychopharmacology, 210,* 511–519.

Allman, A.-A., Benkelfat, C., Durand, F., Sibon, I., Dagher, A., Leyton, M., et al. (2010). Effect of d-amphetamine on inhibition and motor planning as a function of baseline performance. *Psychopharmacology, 211,* 423–433.

Angrist, B., Corwin, J., Bartlik, B., & Cooper, T. (1987). Early pharmacokinetics and clinical effects of oral d-amphetamine in normal subjects. *Biological Psychiatry, 22,* 1357–1368.

Aron, A. R., Robbins, T. W., & Poldrack, R. A. (2004). Inhibition and the right inferior frontal cortex. *Trends in Cognitive Sciences, 8,* 170–177.

Barch, D. M., & Carter, C. S. (2005). Amphetamine improves cognitive function in medicated individuals with schizophrenia and in healthy volunteers. *Schizophrenia Research, 77,* 43–58.

Breitenstein, C., Flöel, A., Korsukewitz, C., Wailke, S., Bushuven, S., & Knecht, S. (2006). A shift of paradigm: From noradrenergic to dopaminergic modulation of learning? *Journal of the Neurological Sciences, 248,* 42–47.

Brignell, C. M., Rosenthal, J., & Curran, H. V. (2007). Pharmacological manipulations of arousal and memory for emotional material: Effects of a single dose of methylphenidate or lorazepam. *Journal of Psychopharmacology, 21,* 673–683.

Burns, J. T., House, R. F., Fensch, F. C., & Miller, J. G. (1967). Effects of magnesium pemoline and dextroamphetamine on human learning. *Science, 155,* 849–851.

Chamberlain, S. R., Robbins, T. W., Winder-Rhodes, S., Müller, U., Sahakian, B. J., Blackwell, A. D., et al. (2011).

Translational approaches to frontostriatal dysfunction in attention-deficit/hyperactivity disorder using a computerized neuropsychological battery. *Biological Psychiatry, 69,* 1192–1203.

Chang, Y. K., Labban, J. D., Gapin, J. I., & Etnier, J. L. (2012). The effects of acute exercise on cognitive performance: A meta-analysis. *Brain Research, 1453,* 87–101.

Cooper, H. (2010). *Research synthesis and meta-analysis.* A step-by-step approach. Thousand Oaks, CA: Sage.

Cooper, N. J., Keage, H., Hermens, D., Williams, L. M., Bebrota, D., & Clark, C. R. (2005). The dose-dependent effect of methylphenidate on performance, cognition and psychophysiology. *Journal of Integrative Neuroscience, 4,* 123–144.

Cortina, J. M., & Nouri, H. (2000). Effect size for ANOVA designs. Thousand Oaks, CA: Sage.

Costa, A., Riedel, M., Pogarell, O., Menzel-Zelnitschek, F., Schwarz, M., Reiser, M., et al. (2013). Methylphenidate effects on neural activity during response inhibition in healthy humans. *Cerebral Cortex, 23,* 1179–1189.

de Bruijn, E. R. A., Hulstijn, W., Verkes, R. J., Ruigt, G. S. F., & Sabbe, B. G. C. (2004). Drug-induced stimulation and suppression of action monitoring in healthy volunteers. *Psychopharmacology, 177,* 151–160.

De Wit, H. (2012). [Stimulants' effects on stop signal task performance]. Unpublished raw data.

De Wit, H., Crean, J., & Richards, J. B. (2000). Effects of d-amphetamine and ethanol on a measure of behavioral inhibition in humans. *Behavioral Neuroscience, 114,* 830–837.

De Wit, H., Enggasser, J. L., & Richards, J. B. (2002). Acute administration of d-amphetamine decreases impulsivity in healthy volunteers. *Neuropsychopharmacology, 27,* 813–825.

Dorflinger, J. (2005). *Methylphenidate and working memory: An investigation using fMRI* (Unpublished dissertation). Rosalind Franklin University of Medicine and Science, Chicago, IL.

Dunlap, W. P., Cortina, J. M., Vaslow, J. B., & Burke, M. J. (1996). Meta-analysis of experiments with matched groups or repeated measures designs. *Psychological Methods, 1,* 170–177.

Engert, V., Joober, R., Meaney, M. J., Hellhammer, D. H., & Pruessner, J. C. (2009). Behavioral response to methylphenidate challenge: Influence of early life parental care. *Developmental Psychobiology, 51,* 408–416.

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,* 143–149.

Everling, S., & Fischer, B. (1998). The antisaccade: A review of basic research and clinical studies. *Neuropsychologia, 36,* 885–899.

Farah, M. J. (2012). [Stimulants' cognitive enhancement effects in healthy people]. Unpublished raw data.

Farah, M. J., Haimm, C., Sankoorikal, G., & Chatterjee, A. (2009). When we enhance cognition with Adderall, do we sacrifice creativity? A preliminary study. *Psychopharmacology, 202,* 541–547.

Farah, M. J., Illes, J., Cook-Deegan, R., Gardner, H., Kandel, E., King, P., et al. (2004). Neurocognitive enhancement: What can we do and what should we do? *Nature Reviews Neuroscience, 5,* 421–425.

Fillmore, M. T., Kelly, T. H., & Martin, C. A. (2005). Effects of d-amphetamine in human models of information processing and inhibitory control. *Drug and Alcohol Dependence, 77,* 151–159.

Fleming, K., Bigelow, L. B., Weinberger, D. R., & Goldberg, T. E. (1995). Neuropsychological effects of amphetamine

may correlate with personality characteristics. *Psychopharmacology Bulletin, 31,* 357–362.

Greely, H. T. (2013). Some first steps toward responsible use of cognitive-enhancing drugs by the healthy. *The American Journal of Bioethics, 13,* 39–41.

Hall, W. D., & Lucke, J. C. (2010). The enhancement use of neuropharmaceuticals: More scepticism and caution needed. *Addiction, 105,* 2041–2043.

Hamidovic, A., Dlugos, A., Skol, A., Palmer, A. A., & de Wit, H. (2009). Evaluation of genetic variability in the dopamine receptor D2 in relation to behavioral inhibition and impulsivity/sensation seeking: An exploratory study with d-amphetamine in healthy participants. *Experimental and Clinical Psychopharmacology, 17,* 374–383.

Heaton, R. K., Chelune, G. J., Talley, J. L., Kay, G. G., & Curtis, G. (1993). *Wisconsin Card Sorting Test (WCST) manual (revised and expanded).* Odessa, Ukraine: Psychological Assessment Resources, Inc.

Helmers, K. F., Young, S. N., & Pihl, R. O. (1995). Assessment of measures of impulsivity in healthy male volunteers. *Personality and Individual Differences, 19,* 927–935.

Henson, R. N. A., Rugg, M. D., Shallice, T., & Dolan, R. J. (2000). Confidence in recognition memory for words: Dissociating right prefrontal roles in episodic retrieval. *Journal of Cognitive Neuroscience, 12,* 913–923.

Hester, R., Nandam, L. S., O'Connell, R. G., Wagner, J., Strudwick, M., Nathan, P. J., et al. (2012). Neurochemical enhancement of conscious error awareness. *The Journal of Neuroscience, 32,* 2619–2627.

Huffcutt, A. I., & Arthur, W. (1995). Development of a new outlier statistic for meta-analytic data. *Journal of Applied Psychology, 80,* 327–334.

Ilieva, I., Boland, J., & Farah, M. J. (2013). Objective and subjective cognitive enhancing effects of mixed amphetamine salts in healthy people. *Neuropharmacology, 64,* 496–505.

Ilieva, I., & Farah, M. J. (2013). Enhancement stimulants: Perceived motivational and cognitive advantages. *Frontiers in Neuroscience-Neuropharmacology, 7,* 198.

Jaeggi, S. M., Buschkuehl, M., Perrig, W. J., & Meier, B. (2010). The concurrent validity of the *n*-back task as a working memory measure. *Memory, 18,* 394–412.

Kane, M. J., Conway, A. R., Miura, T. K., & Colflesh, G. J. (2007). Working memory, attention control, and the *n*-back task: A question of construct validity. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33,* 615–622.

Kelly, T. H., Robbins, G., Martin, C. A., Fillmore, M. T., Lane, S. D., Harrington, N. G., et al. (2006). Individual differences in drug abuse vulnerability: d-Amphetamine and sensation-seeking status. *Psychopharmacology, 189,* 17–25.

Lappin, J. S., & Eriksen, C. W. (1966). Use of a delayed signal to stop a visual reaction-time response. *Journal of Experimental Psychology, 72,* 805–811.

Lau, J., Ioannidis, J. P. A., Terrin, N., Schmid, C. H., & Olkin, I. (2006). Evidence based medicine: The case of the misleading funnel plot. *British Medical Journal, 333,* 597–600.

Linssen, A. M. W., Vuurman, E. F. P. M., Sambeth, A., & Riedel, W. J. (2012). Methylphenidate produces selective enhancement of episodic memory consolidation in healthy volunteers. *Psychopharmacology, 221,* 611–619.

Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis.* Thousand Oaks, CA: Sage.

Logan, G. D., Schachar, R. J., & Tannock, R. (1997). Impulsivity and inhibitory control. *Psychological Science, 8,* 60–64.

Marquand, A. F., De Simoni, S., O'Daly, O. G., Williams, S. C., Mourão-Miranda, J., & Mehta, M. A. (2011). Pattern

classification of working memory networks reveals differential effects of methylphenidate, atomoxetine, and placebo in healthy volunteers. *Neuropsychopharmacology, 36,* 1237–1247.

Martinussen, R., Hayden, J., Hogg-Johnson, S., & Tannock, R. (2005). A meta-analysis of working memory impairments in children with attention-deficit/hyperactivity disorder. *Journal of the American Academy of Child and Adolescent Psychiatry, 44,* 377–384.

Mattay, V. S., Berman, K. F., Ostrem, J. L., Esposito, G., Horn, J. D. V., Bigelow, L. B., et al. (1996). Dextroamphetamine enhances "neural network-specific" physiological signals: A positron-emission tomography rCBF study. *The Journal of Neuroscience, 16,* 4816–4822.

Mattay, V. S., Callicott, J. H., Bertolino, A., Heaton, I., Frank, J. A., Coppola, R., et al. (2000). Effects of dextroamphetamine on cognitive performance and cortical activation. *Neuroimage, 12,* 268–275.

Mattay, V. S., Goldberg, T. E., Fera, F., Hariri, A. R., Tessitore, A., Egan, M. F., et al. (2003). Catechol O-methyltransferase val158-met genotype and individual variation in the brain response to amphetamine. *Proceedings of the National Academy of Sciences, U.S.A., 100,* 6186–6191.

McCabe, S. E., Teter, C. J., & Boyd, C. J. (2006). Medical use, illicit use, and diversion of prescription stimulant medication. *Journal of Psycho-active Drugs, 38,* 43–56.

McGaugh, J. L., & Roozendaal, B. (2009). Drug enhancement of memory consolidation: Historical perspectives and neurobiological implications. *Psychopharmacology, 202,* 3–14.

Mehta, M. A., Owen, A. M., Sahakian, B. J., Mavaddat, N., Pickard, J. D., & Robbins, T. W. (2000). Methylphenidate enhances working memory by modulating discrete frontal and parietal lobe regions in the human brain. *The Journal of Neuroscience, 20,* RC65.

Mintzer, M. Z., & Griffiths, R. R. (2003). Triazolam-amphetamine interaction: Dissociation of effects on memory versus arousal. *Journal of Psychopharmacology, 17,* 17–29.

Mintzer, M. Z., & Griffiths, R. R. (2007). A triazolam/ amphetamine dose–effect interaction study: Dissociation of effects on memory versus arousal. *Psychopharmacology, 192,* 425–440.

Moeller, S. J., Honorio, J., Tomasi, D., Parvaz, M. A., Woicik, P. A., Volkow, N. D., et al. (2012). Methylphenidate enhances executive function and optimizes prefrontal function in both health and cocaine addiction. *Cerebral Cortex, 24,* 643–653.

Nandam, L. S., Hester, R., Wagner, J., Cummins, T. D. R., Garner, K., Dean, A. J., et al. (2011). Methylphenidate but not atomoxetine or citalopram modulates inhibitory control and response time variability. *Biological Psychiatry, 69,* 902–904.

Oken, B. S., Kishiyama, S. S., & Salinsky, M. C. (1995). Pharmacologically induced changes in arousal: Effects on behavioral and electrophysiologic measures of alertness and attention. *Electroencephalography and Clinical Neurophysiology, 95,* 359–371.

Owen, A. M., Downes, J. J., Sahakian, B. J., Polkey, C. E., & Robbins, T. W. (1990). Planning and spatial working memory following frontal lobe lesions in man. *Neuropsychologia, 28,* 1021–1034.

Pauls, A. M., O'Daly, O. G., Rubia, K., Riedel, W. J., Williams, S. C. R., & Mehta, M. A. (2012). Methylphenidate effects on prefrontal functioning during attentional-capture and response inhibition. *Biological Psychiatry, 72,* 142–149.

Peters, J. L., Sutton, A. J., Jones, D. R., Abrams, K. R., & Rushton, L. (2007). Performance of the trim and fill method in the presence of publication bias and between-study heterogeneity. *Statistics in Medicine, 26,* 4544–4562.

Peters, J. L., Sutton, A. J., Jones, D. R., Abrams, K. R., Rushton, L., & Moreno, S. G. (2010). Assessing publication bias in meta-analyses in the presence of between-study heterogeneity. *Journal of the Royal Statistical Society: Series A, 173,* 575–591.

Postle, B. R., Jonides, J., Smith, E. E., Corkin, S., & Growdon, J. H. (1997). Spatial, but not object, delayed response is impaired in early Parkinson's disease. *Neuropsychology, 11,* 171–179.

Poulin, C. (2007). From attention-deficit/hyperactivity disorder to medical stimulant use to the diversion of prescribed stimulants to non-medical stimulant use: Connecting the dots. *Addiction, 102,* 740–751.

Ramasubbu, R., Singh, H., Zhu, H., & Dunn, J. F. (2012). Methylphenidate-mediated reduction in prefrontal hemodynamic responses to working memory task: A functional near-infrared spectroscopy study. *Human Psychopharmacology: Clinical and Experimental, 27,* 615–621.

Repantis, D., Schlattmann, P., Laisney, O., & Heuser, I. (2010). Modafinil and methylphenidate for neuroenhancement in healthy individuals: A systematic review. *Pharmacological Research, 62,* 187–206.

Rogers, R. D., Blackshaw, A. J., Matthews, K., Middleton, H. C., Hawtin, K., Crowley, C., Hopwood, A., et al. (1999). Tryptophan depletion impairs stimulus-reward learning while methylphenidate disrupts attentional control in healthy young adults: Implications for the monoaminergic basis of impulsive behaviour. *Psychopharmacology, 146,* 482–491.

Rosenthal, R., & Rubin, D. B. (2003). Requivalent: A simple effect size indicator. *Psychological Methods, 8,* 492–496.

Rothstein, H. R., Sutton, A. J., & Borenstein, M. (2006). *Publication bias in meta-analysis: Prevention, assessment and adjustments.* Chichester, UK: Wiley.

Sahakian, B. J., & Morein-Zamir, S. (2011). Neuroethical issues in cognitive enhancement. *Journal of Psychopharmacology, 25,* 197–204.

Schmedtje, J. F., Jr., Oman, C. M., Letz, R., & Baker, E. L. (1988). Effects of scopolamine and dextroamphetamine on human performance. *Aviation, Space, and Environmental Medicine, 59,* 407–410.

Sedlmeier, P., Eberth, J., Schwarz, M., Zimmermann, D., Haarig, F., Jaeger, S., et al. (2012). The psychological effects of meditation: A meta-analysis. *Psychological Bulletin, 138,* 1139–1171.

Servan-Schreiber, D., Carter, C. S., Bruno, R. M., & Cohen, J. D. (1998). Dopamine and the mechanisms of cognition: Part II. d-amphetamine effects in human subjects performing a selective attention task. *Biological Psychiatry, 43,* 723–729.

Silber, B. Y., Croft, R. J., Papafotiou, K., & Stough, C. (2006). The acute effects of d-amphetamine and methamphetamine on attention and psychomotor performance. *Psychopharmacology, 187,* 154–169.

Smith, E. M., & Farah, M. J. (2011). Are prescription stimulants "smart pills"? The epidemiology and cognitive neuroscience of prescription stimulant use by normal healthy individuals. *Psychological Bulletin, 137,* 717–741.

Sockol, L. E., Epperson, C. N., & Barber, J. P. (2011). A meta-analysis of treatments for perinatal depression. *Clinical Psychology Review, 31,* 839–849.

Soetens, E., Hueting, J. E., Casaer, S., & D'Hooge, R. (1995). Effect of amphetamine on long-term retention of verbal material. *Psychopharmacology, 119,* 155–162.

Sofuoglu, M., Waters, A. J., Mooney, M., & Kosten, T. (2008). Riluzole and d-amphetamine interactions in humans. *Progress in Neuro-Psychopharmacology and Biological Psychiatry, 32,* 16–22.

Sternberg, S. (1966). High-speed scanning in human memory. *Science, 153,* 652–654.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology, 18,* 643–662.

Studer, P., Wangler, S., Diruf, M. S., Kratz, O., Moll, G. H., & Heinrich, H. (2010). ERP effects of methylphenidate and working memory load in healthy adults during a serial visual working memory task. *Neuroscience Letters, 482,* 172–176.

Talbot, M. (2009, April). Brain gain: The underground world of neuroenhancing drugs. *The New Yorker*. Retrieved from www.newyorker.com/reporting/2009/04/27/090427fa_fact_talbot.

Tèter, C. J., Mccabe, S. E., Cranford, A., Boyd, C. J., & Guthrie, S. K. (2005). Prevalence and motives for illicit use of prescription stimulants in an undergraduate student sample. *Journal of Clinical Psychopharmacology, 53,* 253–262.

The Psychological Corporation. 2002. *Updated WAISIII-WMS-III technical manual*. San Antonio, TX: Pearson.

Theunissen, E. L., Elvira, Jde L., van den Bergh, D., & Ramaekers, J. G. (2009). Comparing the stimulant effects of the H1-antagonist fexofenadine with 2 psychostimulants, modafinil and methylphenidate. *Journal of Clinical Psychopharmacology, 29,* 439–443.

Unrug, A., Coenen, A., & van Luijtelaar, G. (1997). Effects of the tranquillizer diazepam and the stimulant methylphenidate on alertness and memory. *Neuropsychobiology, 36,* 42–48.

Volkow, N. D., Wang, G.-J., Fowler, J. S., Gatley, S. J., Logan, J., Ding, Y.-S., et al. (1998). Dopamine transporter occupancies in the human brain induced by therapeutic doses of oral methylphenidate. *American Journal of Psychiatry, 155,* 1325–1331.

Vrecko, S. (2013). Just how cognitive is "cognitive enhancement"? On the significance of emotions in university students' experiences with study drugs. *AJOB Neuroscience, 4,* 4–12.

Vree, T. B., & Van Rossum, J. M. (1970). Kinetics of metabolism and excretion of amphetamines in man. *Amphetamines and Related Compounds,* 165–190.

Wager, T. O. R. D., & Smith, E. E. (2003). Neuroimaging studies of working memory: A meta-analysis. *Cognitive, Affective & Behavioral Neuroscience, 3,* 255–274.

Wardle, M. C., Hart, A. B., Palmer, A. A., & de Wit, H. (2013). Does COMT genotype influence the effects of d-amphetamine on executive functioning? *Genes, Brain and Behavior, 12,* 13–20.

White, T. L., Justice, A. J. H., & de Wit, H. (2002). Differential subjective effects of d-amphetamine by gender, hormone levels and menstrual cycle phase? *Pharmacology, Biochemistry And Behavior, 73,* 729–741.

Wilens, T. E., Adler, L. A., Adams, J., Sgambati, S., Rotrosen, J., Sawtelle, R., et al. (2008). Misuse and diversion of stimulants prescribed for ADHD: A systematic review of the literature. *Journal of the American Academy of Child & Adolescent Psychiatry, 47,* 21–31.

Willett, R. A. (1962). The effect of a stimulant and a depressant drug on the serial rote learning of nonsense syllables. *Psychopharmacologia, 3,* 23–34.

Zeeuws, I., Deroost, N., & Soetens, E. (2010a). Effect of an acute d-amphetamine administration on context information memory in healthy volunteers: Evidence from a source memory task. *Human Psychopharmacology: Clinical and Experimental, 25,* 326–334.

Zeeuws, I., Deroost, N., & Soetens, E. (2010b). Verbal memory improved by d-amphetamine: Influence of the testing effect. *Human Psychopharmacology: Clinical and Experimental, 25,* 377–387.

Zeeuws, I., & Soetens, E. (2007). Verbal memory performance improved via an acute administration of d-amphetamine. *Human Psychopharmacology: Clinical and Experimental, 22,* 279–287.