Psychological Bulletin

The Comparative Analysis of Intelligence

Mary Flaim and Aaron P. Blaisdell

The Comparative Analysis of Intelligence

Mary Flaim and Aaron P. Blaisdell
University of California, Los Angeles

The study of intelligence in humans has been ongoing for over 100 years, including the underlying structure, predictive validity, related cognitive measures, and source of differences. One of the key findings in intelligence research is the uniform positive correlations among cognitive tasks. This has been replicated with every cognitive test battery in humans. Nevertheless, many other aspects of intelligence research have revealed contradictory lines of evidence. Recently, cognitive test batteries have been developed for animals to examine similarities to humans in cognitive structure. The results are inconsistent, but there is evidence for some similarities. This article reviews the way intelligence and related cognitive abilities are assessed in humans and animals and suggests a different way of devising test batteries for maximizing between-species comparisons.

Public Significance Statement
This review highlights the difficulty in understanding why different cognitive abilities are related to each other in human and nonhuman animals. Cognitive abilities are related in similar ways across species, but test batteries should be refined to strengthen these findings and increase translational significance.

Keywords: general intelligence, *g* factor, comparative cognition, cognitive test battery

Why do some people live longer, healthier lives, or obtain higher levels of education, or gravitate toward more cognitively demanding careers? What underlies individual differences in reaction time (RT), working memory, or learning? Many researchers have argued that individual differences in intelligence underlie each of these, as it is the best, though not a perfect, predictor of many of these (Brodnick & Ree, 1995; Conway, Cowan, Bunting, Therriault, & Minkoff, 2002; Deary, Whiteman, Starr, Whalley, & Fox, 2004; Gottfredson, 2002; Gottfredson & Deary, 2004; Jensen, 1998; Ree & Earles, 1992; Schmidt, 2011, 2014; Sheppard & Vernon, 2008; but see Gutman & Schoon, 2013; Heckman, Pinto, & Savelyev, 2013 for the importance of "noncognitive" factors and Ceci, 1991 on how schooling is a causal factor for performance on intelligence measures). Intelligence is typically measured with a full-scale IQ (FSIQ) test. The FSIQ contains a battery of diverse tasks designed to assess different aspects of cognition, including basic math skills, matrix reasoning, spatial reasoning, verbal comprehension, and memory, though the specific content can vary across tests (Johnson, Bouchard, Krueger, McGue, & Gottesman, 2004; Reynolds, Floyd, & Niileksela, 2013; Schrank & McGrew, 2001), and concerns have been raised about how often these tests

reflect western education or culture (Ceci, 1991; Nisbett, 2009; Serpell, 2000; Wicherts, Dolan, & van der Maas, 2010). Capturing these individual differences across all these tasks with a single metric may appear to overlook important factors. Perhaps a person is terrible at math, for example, but has exceptional verbal comprehension. Nevertheless, for a large majority of people, performance typically correlates across all tasks—despite their diversity (Carroll, 1993; Deary, 2000).

Charles Spearman (1904) was the first to report a uniform positive correlation among diverse cognitive tasks in people and he called this the "positive manifold." This finding has continued to be replicated ever since (Carroll, 1993; Deary, 2000; Jensen, 1998). When these positive correlational matrices are subjected to factor analysis, one factor that explains approximately half of the variance in performance is extracted. It is called the *g* factor (Carroll, 1993; Jensen, 1998; Spearman, 1904). When tasks load onto a factor, it indicates how much variance in that task can be explained by the factor. All cognitive tasks load in the appropriate direction onto *g*, but not all tasks load equally (Nisbett, 2009; Reynolds et al., 2013; Weiss, Keith, Zhu, & Chen, 2013). The tasks that show the highest loading onto this *g* factor involve reasoning, abstraction, task complexity, and task novelty, irrespective of how the information is presented within the tasks themselves (Ackerman & Cianciolo, 2000; Jensen, 1998; but see Ceci, 1996 for difficulties determining how "abstract" a problem is). Many researchers have described the *g* factor as "indifferent to the indicator" because loading depends more on the complexity and abstraction of the task rather than on a specific type of measure, or in this usage the "indicator" (Jensen, 1998; Spearman, 1904). This is why *g* is thought to reflect a general cognitive ability and has

[ID] Mary Flaim and [ID] Aaron P. Blaisdell, Department of Psychology, University of California, Los Angeles.

Correspondence concerning this article should be addressed to Mary Flaim, Department of Psychology, University of California, Los Angeles, Psychology Building 1285, Box 951563, Los Angeles, CA 90095-1563. E-mail: meflaim@ucla.edu

predictive validity in a variety of contexts (Deary et al., 2004; Gottfredson, 2002).

While the *g* factor typically accounts for half of the variance in performance, it can depend on the number, reliability, and familiarity of the tasks used, and the variation in the sample tested (Ackerman & Cianciolo, 2000; Colom, Abad, García, & Juan-Espinosa, 2002). All cognitive tasks measure *g*, but they also measure narrower abilities and contain task-specific variance (Carroll, 1993; Gustafsson, 2003; Jensen, 1998). The most accurate measures of *g* will be obtained with a diverse test battery. Even with a diverse test battery, the intercorrelations or extracted factor can be smaller than expected because of measurement (un)reliability and range restriction (Jensen, 1998; Viswesvaran, Ones, Schmidt, Le, & Oh, 2014). Range restriction limits the amount of variability in the sample, but the variability across subjects is exactly what factor analysis is attempting to explain (Jensen, 1998)! This lack of variability will lower the value of the extracted factor, but this is primarily because of sample characteristics. Range restriction is a crucial factor when assessing *g* in small samples with a small number of tasks. While a large and representative sample can help with range restriction, the same cannot be said for the effects of measurement reliability. Moderate to low task reliability (.5–.8) attenuates the subsequent correlations, which can impact later factor analysis since more of the variance will be because of random error or transient factors unrelated to *g* (Fan, 2003; Jensen, 1998). Some statistical methods (like structural equation modeling, *SEM*, and confirmatory factor analysis [CFA]) can account for this, but multiple regression does not, increasing the likelihood of false positives (Westfall & Yarkoni, 2016). Nevertheless, many studies have handled these challenges beautifully. The results from various large FSIQ tests conducted with representative samples indicate that, even though the exact test content can vary, the same *g* factor is extracted (Johnson et al., 2004; Reynolds et al., 2013; Schrank & McGrew, 2001). The *g* factor is robust against different methods of analysis, populations, cultures, and test batteries (Carroll, 1993, 2003; Chabris, 2007; Deary, 2000; Warne & Burningham, 2019; but see Wicherts et al., 2010) and is relatively stable throughout the life span starting at 2 years old (Deary, Pattie, & Starr, 2013; Gignac, 2014; Spinath, Ronald, Harlaar, Price, & Plomin, 2003).

While *g* can account for a large amount of cross-task variance in individual differences, additional variance can be explained by group factors. Some tasks show stronger correlations with each other, forming a subgroup. For example, in a test battery with three verbal measures and three math measures, there is a stronger correlation within verbal measures and within math measures than between both domains (Carroll, 1993; Jensen, 1998). Group factors are more strongly affected by test battery composition (Carroll, 2003; Cattell, 1987; Johnson et al., 2004). The most commonly found group factors include fluid intelligence (*Gf*), crystallized intelligence (*Gc*), quantitative reasoning (*Gq*), visual processing (*Gv*), processing speed (*Gs*), and memory, though the exact terminology can vary (Carroll, 1993; Cattell, 1987; Hakstian & Cattell, 1978). Most research has focused on *Gf* and *Gc* (Kvist & Gustafsson, 2008; Nisbett, 2009; but see Johnson & Bouchard, 2005).

Gf is the ability to solve novel and complex problems, in particular those that require relational reasoning, and frequently use shapes and figures in the tasks as opposed to words. *Gf* loads very highly, and sometimes perfectly, onto *g* (Benson, Hulac, & Kranzler, 2010; Bickley, Keith, & Wolfle, 1995; Carroll, 1993; Gustafsson, 1984; Kvist & Gustafsson, 2008), though the strength of loading depends on many factors, such as sample homogeneity, number of tests in the assessment, and methods of analysis (Blair, 2006; Carroll, 2003; Kan, Kievit, Dolan, & van der Maas, 2011; Thorsen, Gustafsson, & Cliffordson, 2014). It is still debated to what degree measures of *Gf* are dependent on school exposure and culture. Some researchers argue that, because *Gf* tasks typically do not use language or memorized facts, this means it relies less on prior knowledge or schooling, and thus should be viewed as culture free (Cattell & Horn, 1978; Jensen, 1998; Kent, 2017). Other researchers have argued that the increased emphasis on formal schooling and increase in visual stimuli in a given culture has resulted in improvements in these tasks in subsequent generations, indicating that these tasks are dependent on school and culture (Baker et al., 2015; Cahan & Noyman, 2001; Ceci, 1991; Nisbett, 2009; Pietschnig & Voracek, 2015). This debate aside, since many measures of *Gf* do not rely on language, they are an important target assessment for assessing *g* in nonhumans (see below). While the ideal way to measure any construct is with a variety of measures, the Raven's Progressive Matrices (RPM) is a quintessential example of a *Gf* task (Carpenter, Just, & Shell, 1990; Nisbett, 2009). The RPM is a series of partially completed matrices, where the participant is tasked with selecting the choice option that will correctly complete the matrix from a set of distractors (Raven, 1941). Each item in the matrix is transformed, sometimes in multiple ways, across the rows and columns of the matrix. The participant must infer the underlying rules and correctly apply them to find the correct answer (Raven, 1941, 2008). The test items progressively increase in difficulty, with few people correctly answering the final questions (Carpenter et al., 1990). Despite the strong visuospatial component, the RPM is used as a measure of reasoning (Schweizer, 2007; but see Gignac, 2015; Stephenson & Halpern, 2013).

Gc reflects the ability to correctly use and apply learned knowledge (Kvist & Gustafsson, 2008). Some researchers have emphasized the role of language and verbal storage, and rely strongly on vocabulary measures to assess this ability (Ackerman & Cianciolo, 2000; Reynolds & Turek, 2012; Rolfhus & Ackerman, 1999). When a more diverse battery is used, however, the extent to which verbal comprehension overlaps with *Gc* has varied (Carroll, 2003; Kan et al., 2011; Schipolowski, Wilhelm, & Schroeders, 2014). In humans, knowledge is typically gained and tested through language, which could have led to the debate about what is the nature of *Gc* (Keith & Reynolds, 2010; Schipolowski et al., 2014). Nevertheless, despite *Gc* being predominantly measured through language, language is not the only way to assess knowledge. As we discuss in more detail later, nonlanguage methods are needed to assess *Gc* in nonhuman animals. Researchers that have utilized a more comprehensive test of knowledge to measure *Gc* have found that it loads highly onto *g* and is a better predictor of academic and job performance compared with *Gf*, particularly for older adults (Postlethwaite, 2011; Schmidt, 2014).

Despite these general issues with group factors, it has been consistently found that *Gf* and *Gc* load highly onto the *g* factor and covary with each other (Carroll, 1993; Schipolowski et al., 2014). One potential explanation for this covariation is provided by investment theory (Kvist & Gustafsson, 2008). Because *Gf* influ-

ences a person's ability to understand or learn from novel problems, it is theorized that *Gf* is used when learning new information. As this information is acquired, it then becomes *Gc*. The initial ability level, determined by *Gf*, will determine the efficacy of this investment or learning when it is specifically directed according to interest, leading to more specific knowledge gains, or during more passive exposure, leading to more general gains (Cattell, 1987; Schmidt, 2011, 2014).

It is possible to discuss *g* at two different levels, as a statistical finding referred to as psychometric *g*, and as a psychological construct. Psychometric *g* is not controversial (Blair, 2006; Carroll, 1993; Jensen, 1998). Positive correlations across diverse cognitive tasks are no longer seen as surprising. Further, many researchers do not argue that a factor analysis will produce one factor that can account for half of the variance (Conway & Kovacs, 2015; Van der Maas et al., 2006). The status of *g* as a psychological construct, however, is still heavily debated. Despite the fact that *g* has been consistently reported in over a century's worth of research, and we know which tasks consistently load highly onto *g*, there remains no consensus as to what *g* actually is (Carroll, 1993; Cattell, 1987; Chabris, 2007; Deary, 2000; Gottfredson, 2002; Gustafsson, 1984, 2003; Jensen, 1998; Kovacs & Conway, 2016; Van der Maas et al., 2006); *g*'s ontological status remains a mystery. The following is a brief overview of many popular theories of *g* today. Our aim is not to review an exhaustive list of all current theories of *g*, nor a nuanced treatment of the theories that are discussed. Additionally, this article is not an endorsement of any particular theory of *g*. Rather, the goal of this article is to provide a general background about theories of intelligence for readers outside the expert community. Furthermore, we also do not cover the vast literature on cognitive abilities in infants or the developmental aspects of general intelligence. Again, this is because we ultimately are interested in discussing tests of *g* in adult nonhuman animals. Undoubtedly, once such tests can be reliably developed, they should allow for the investigation of how developmental processes contribute to *g* in nonhuman animals, but such a discussion would be premature now. For excellent empirical research and theories on the development of *g*, see Blaga et al. (2009); Bornstein et al. (2006); Coyle, Pillow, Snyder, and Kochunov (2011); Demetriou et al. (2018); Fagan, Holland, and Wheeler (2007); Rose, Feldman, Jankowski, and Van Rossem (2008); and Spinath et al. (2003).

One proposal is that *g* is a single entity that is related to a wide variety of cognitive abilities because it *causes* differences between individuals in those abilities (Brown, Le, & Schmidt, 2006; Carroll, 1993; Gustafsson, 1984, 2003; Schmidt, 2011, 2014, 2017

right panel of Figure 1). Even though this perspective purports *g* as a single entity, it does not necessarily reflect one physical structure or psychological process (Jensen, 1998). In attempting to identify the physical substrates of *g*, a variety of results have been found including relevant genes (Plomin & von Stumm, 2018), neural networks (Duncan et al., 2000), neural substrates (Schmitt, Raznahan, Liu, & Neale, 2020), and developmental processes (Garlick, 2002). It is unlikely that any one of these alone is responsible for *g* and more likely that there is a dynamic interaction between all of these physical substrates and with the environment (Ceci, 1991; Chabris, 2007; Garlick, 2002; Jensen, 1998; Kan, Wicherts, Dolan, & van der Maas, 2013; Schmitt et al., 2020; Van der Maas et al., 2006). Indeed, some researchers argue that schooling has robust and potentially causal effects on the physical substrates that could underlie *g* (Baker et al., 2015). At the psychological construct level, other researchers theorize that elementary cognitive process underlie *g*, meaning that differences in one or more of these basic abilities is predominately the reason behind differences in *g* (Gignac, 2014; Jensen, 1998, p. 260). Working memory (WM), short-term memory (STM), processing speed, associative learning, and response inhibition have all been proposed as components of *g* (Conway et al., 2002; Deary, 2000; Dempster, 1991; Jensen, 1998; Kaufman, DeYoung, Gray, Brown, & Mackintosh, 2009; Sheppard & Vernon, 2008; left panel of Figure 1).

Other researchers argue that *g* is actually a statistical artifact. These theories state that more complex tasks (that are more *g* loaded) require a broader array of *independent* resources. Even though these processes are independent, the nature of the tasks creates a correlation (Bartholomew, Deary, & Lawn, 2009; Kovacs & Conway, 2016) or that the developmental trajectory creates mutually beneficial interactions between independent abilities (Rose et al., 2008; Van der Maas et al., 2006). In the next section, we review the relationship between *g* and the cognitive mechanisms listed earlier to investigate this issue. These cognitive mechanisms were investigated because of the rich literature that is available to review and because the potential role they play when investigating *g* across species.

Related Cognitive Factors

Working Memory

WM describes the ability to hold a limited amount of information over the short term (seconds to minutes). What differentiates WM from STM is that WM involves manipulating the stored information or engaging in a secondary task while the to-be-
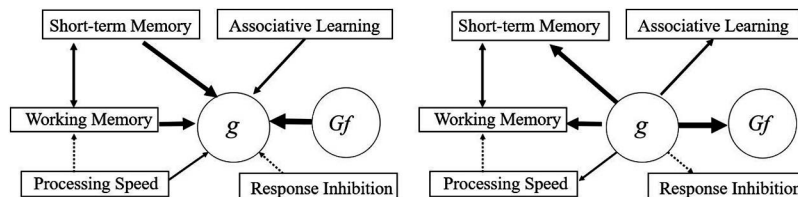

Figure 1. A diagram representing the reviewed cognitive abilities and their relationship to *g* and to each other. The thickness of the lines represents the strength of the relationship, while the type of line (solid or dashed) represents the consistency of the relationship. The direction of the arrows indicate the thereotical causal relationship. *Gf* = fluid intelligence.

recalled information is held in memory (Baddeley, 2003; Conway et al., 2002). For example, STM might involve holding in memory a list of items until their recall is requested, while WM would involve performing mathematical operations, counting, or some other transformation while encoding a list of to be recalled items. Some common WM tasks are the complex span task, n-back task (Au et al., 2015; Shelton, Elliott, Matthews, Hill, & Gouvier, 2010), and reverse span task (Oberauer, Süß, Schulze, Wilhelm, & Wittmann, 2000).

In the complex span task, there is a competing demand that is interspersed between to-be-remembered items (Conway et al., 2002; Engle, Tuholski, Laughlin, & Conway, 1999). For example, in the operation span task, participants must verify if the solution to a given equation is correct or incorrect, before being presented with the to-be-recalled word, letter, or number (Conway et al., 2002). Different variations of the complex span task include verifying if a sentence is logical or counting the number of squares before being presented with the to-be-remembered items. The degree of interference between the interleaved task and the items for recall can vary by changing the similarity between them. Some experiments found worse recall performance when the interleaved task and the to-be remembered item are highly similar, for example both involve words or visuospatial judgments (Jarrold, Tam, Baddeley, & Harvey, 2011; Shah & Miyake, 1996), but this is not consistently found across all item types (Bayliss, Jarrold, Gunn, & Baddeley, 2003). Performance on these different span tasks are correlated, but the correlation is different from unity (Bayliss et al., 2003; Conway et al., 2002). It is possible that this is partially because of measurement error, but an exploratory factor analysis extracted three factors, which they interpreted to be verbal storage, visuospatial storage, and a general processing factor (Bayliss et al., 2003). This indicates that each span task captures more specific and general properties of WM, which is consistent with theoretical conceptions (Baddeley, 2003).

In the n-back task, participants are presented with a continuous stream of items. As each item is presented, the participant must decide whether it matches an item presented n trials ago, with the range typically extending from 0–3 (Jaeggi, Buschkuehl, Jonides, & Perrig, 2008). For example, in the stream: fish, peanut, cup, fish, pipe, dog, dog, car, phone, car; a response would be required to the second presentation of "dog" in a 1-back task, "car" in a 2-back task, and "fish" in a 3-back task. Generally, the larger the n, the more difficult the task. Thus, participants must continuously update the items in their WM, while simultaneously comparing each current item to the appropriate item n-back.

In the reverse-span task, participants are presented with a series of letters or numbers, and then they must repeat them in reverse order. This means that participants are simultaneously holding and transposing the information so that it can be presented in reverse order (Jensen, 1998; Oberauer et al., 2000).

Performance on these WM tasks is correlated with measures of g, but the reason for this correlation is not well understood. Some researchers have shown that WM training improves performance on WM and highly g-loaded tasks, indicating that WM is a subcomponent of g (Jaeggi et al., 2008; Schmiedek, Lövdén, & Lindenberger, 2010). Others argue, however, that these improvements are hollow—that is, they stem from non-g factors, like test familiarity or strategy adjustments during test battery completion (Colom et al., 2002; Colom et al., 2013; Estrada, Ferrer, Abad, Román, & Colom, 2015). Additionally, not all researchers have shown WM-training effects on g (Chooi & Thompson, 2012; Harrison et al., 2013; Redick et al., 2013). These mixed effects could mean that WM is used during these g loaded tasks in holding necessary information, but the ability to correctly identify which information is necessary is unique to g. Being able to hold more information or handle competing demands more effectively does not necessarily indicate an improved ability for abstract reasoning. This dissociation between WM and abstract reasoning could indicate that there is a causal relationship between WM and g, but that differences in g cause differences in WM, not the other way around. Theoretically, from this perspective an increase in g should result in an increase in WM performance as well. Alternatively, there could be another, more general factor that underlies the efficacy of WM and g-loaded tasks. Training on WM that fails to improve this underlying factor should result in little impact for performance on g loaded measures. Finally, WM and g could be independent cognitive abilities and the reason for the correlation is because of task impurity.

The diversity of tasks used to measure WM obstructs determining the relationship between g and WM. Each type of WM task places different demands on WM, and thus they may not all be measuring the same construct (Aben, Stapert, & Blokland, 2012). Supporting this is the fact that these tasks do not always strongly correlate with each other (Au et al., 2015; Jaeggi et al., 2008; Kane, Conway, Miura, & Colflesh, 2007; but see Schmiedek, Lövdén, & Lindenberger, 2014; Wilhelm, Hildebrandt, & Oberauer, 2013). Furthermore, WM tasks sometimes strongly correlate with STM measures (Aben et al., 2012; Colom, Abad, Quiroga, Shih, & Flores-Mendoza, 2008; Conway et al., 2005; St Clair-Thompson, 2010; Figure 1), further obscuring relationships between tasks and the underlying constructs they purportedly measure. Likewise, WM is not necessarily a unitary construct, but may itself consist of separate processes, such as attention (Baddeley, 2003), processing speed (Unsworth, Redick, Heitz, Broadway, & Engle, 2009), and STM capacity (Conway et al., 2002), among others (Kovacs & Conway, 2016; Schmiedek et al., 2014; Wilhelm et al., 2013). Support for the relationship between g and these other processes have all been reported (Chuderski, Taraday, Nęcka, & Smoleń, 2012; Conway et al., 2002; Unsworth et al., 2009; Figure 1). Thus, it is possible that different WM tasks differentially tap into these alternative processes (or subcomponents).

Short-Term Memory

STM is the ability to hold information over a delay period, without an explicit competing task or manipulation requirement (Unsworth & Engle, 2007). The information being held in STM is subject to capacity limits and decay over time (Cowan, 2008). Performance on WM and STM tasks tends to be correlated, likely because both involve the short-term retention of information (Aben et al., 2012; Colom et al., 2008; Conway et al., 2002; Figure 1). STM and WM are not dichotomous constructs; rather, tasks fall on a continuum depending on how demanding is the secondary task (Aben et al., 2012; Engle et al., 1999). For example, requiring participants to repeat a letter, preventing them from verbally rehearsing the to-be-remembered items, is still considered a STM task because the secondary task is of low difficulty (Conway et al., 2002; Engle et al., 1999). Nevertheless, as discussed above, WM

does have some unique properties (Conway et al., 2002; Cowan, 2008). Some researchers have found that STM alone is related to *g* using *SEM* and CFA (Colom et al., 2008; Martínez et al., 2011; Figure 1). Yet, others have failed to find a unique relationship between STM and *g* using *SEM*, but these utilized different construct measures (Conway et al., 2002). Given these challenges, researchers tend to focus on WM when investigating the relationship between STM processes and *g*.

Processing Speed

The term "processing speed" is used to describe a variety of tasks that can vary in complexity and memory demands (Deary, 2000). These tasks typically assess how quickly a participant can detect a change in the environment, perceive the difference between two stimuli, or transform stimuli. Some tasks that require detecting a change in the environment are based off of Hick's law, that RT will increase linearly with increases in information a task requires (Hick, 1952). The Jensen box is an example of an apparatus that utilizes the principal of Hick's Law (Deary, 2000; Jensen & Munro, 1979). The Jensen box consists of a home key and 1, 2, 4, 6, or 8 stimuli placed equidistant from the home key in a semicircular arrangement. Participants must keep their finger resting on the home key until one of the stimuli in the array changes (e.g., color or brightness). Participants are instructed to touch the changed stimulus as quickly as possible (Deary, Der, & Ford, 2001; Jensen, 1982; Vickrey & Neuringer, 2000). Even at its most simple, when the array only has one stimulus, there is still a relationship between RT and intelligence, with the correlation ranging between −.18 to −.22 (Deary, 2000; Doebler & Scheffler, 2016; Sheppard & Vernon, 2008). Another processing speed task is the digit-symbol substitution task (Conway et al., 2002; Hoyer, Stawski, Wasylyshyn, & Verhaeghen, 2004). In this task, participants are given a conversion table of digits and a corresponding symbol. The symbols are usually simple shapes or a series of connected lines that do not resemble letters. Participants must complete a table of numbers with the appropriate corresponding symbol, or they are shown various digit-symbol pairs and must determine if the pairs are valid or invalid according to the conversion table as quickly and accurately as possible (Conway et al., 2002; Hoyer et al., 2004). The conversion table is always present, so this task does not rely on memory processes. This task is commonly included in FSIQ tests (Benson et al., 2010). Even though processing-speed tasks appear simple, they show a consistent, modest relationship to *g*, with correlations typically ranging from −.22 to −.4, such that faster or shorter RTs correlate with higher scores on intelligence tests (Deary, 2000; Doebler & Scheffler, 2016; Sheppard & Vernon, 2008; Vernon, 1983; Figure 1). These two tasks are a very select subset of all the different processing speed tasks that are used (Deary, 2000; Sheppard & Vernon, 2008).

Why processing speed shows a consistent relationship with *g* is not well understood (Deary, 2000). There is some evidence that processing speed influences how quickly a competing task can be performed in WM tasks (Conway et al., 2002; Unsworth et al., 2009; Figure 1). Therefore, processing speed may only be related to *g* because it influences WM. When WM tasks are also included, processing speed is no longer directly related to *g* (Conway et al., 2002). It is also possible that processing speed, WM, and *g* rely on the same underlying mechanism or process. There are a large variety of processing speed tasks, however, so it is unclear if these different tasks measure the same underlying construct (Stankov & Roberts, 1997). Tasks used to show a relationship between processing speed and *g* differ greatly from those used to study how processing speed influences WM (Colom et al., 2008; Conway et al., 2002; Deary, 2000; Stankov & Roberts, 1997). Thus, it is difficult to determine if processing speed and WM show the same relationship across all of these tasks.

Response Inhibition

Some researchers have suggested that response inhibition is a crucial factor underlying differences in intelligence (Dempster, 1991). Response inhibition is the ability to suppress unwanted motor responses or thoughts and can be measured with antisaccade, Stroop, Go/No-go (GNG), and stop signal tasks (Friedman et al., 2006; Swick, Ashley, & Turken, 2011; Verbruggen, Best, Bowditch, Stevens, & McLaren, 2014). A reversal learning task is also used to a lesser degree to measure inhibition (Eagle, Bari, & Robbins, 2008; Izquierdo & Jentsch, 2012). In the antisaccade task, participants must avoid moving their eyes toward a target and instead they must move their eyes in the opposite direction (Klein, Rauh, & Biscaldi, 2010). In the Stroop task, participants must read the ink color of a word out loud, even when it conflicts with the word's meaning (i.e., the word "red" printed in yellow ink; Stroop, 1935). In the GNG and stop signal tasks, participants must make a motor response when they see one type of stimulus and withhold the motor response when they see or hear other types of stimuli. For the GNG task, participants receive successive presentations of two stimuli intermingled within the session. Responses to the positive discriminative stimulus (S+) are rewarded while responses to the negative discriminative stimulus (S−) are not rewarded. Initially participants typically make responses to both stimuli, but with further training learn to inhibit responses to the S−. The stop-signal task is similar to a GNG task. On some trials, only the S+ is presented, and participants are rewarded for responding to the S+. Occasionally a trial will initially present the S+, and after a short delay the S− is also presented. The participant is instructed to withhold responses when the S− is presented. Thus, the stop-signal task measures the ability of the participant to suppress behavior in the midst of preparing or making a response. (Swick et al., 2011; Verbruggen et al., 2014). Finally, in the reversal-learning procedure, the first phase consists of a GNG procedure in which participants learn that one cue (S+) is associated with a reward, while the other is not (S−; note, the S+ and S− may be presented simultaneously rather than successively). Once discrimination performance stabilizes, the stimulus-outcome assignments are reversed (Eagle et al., 2008; Izquierdo & Jentsch, 2012). For example, after learning to respond to a blue circle (S+) and withhold responding to a yellow circle (S−), the blue circle becomes the S− and the yellow circle becomes the S+. Response inhibition influences how quickly the participant can inhibit the original learned responses, and replace them with new responses.

For reversal learning, no significant relationship has been found between intelligence and the number of trials needed to reverse the initial discrimination in children (Plenderleith, 1956) or adults (Stevenson & Zigler, 1957). Using the antisaccade task, the total number of errors (Friedman et al., 2006) and the errors with a

regular latency (Klein et al., 2010) showed a modest correlation with intelligence. A similar result was found with the GNG (Horn, Dolan, Elliott, Deakin, & Woodruff, 2003) and the stop-signal tasks (Friedman et al., 2006). The Stroop task had low (Friedman et al., 2006) to nonsignificant (Polderman et al., 2009) correlations with measures of intelligence. When intelligence was broken down into *Gf* and *Gc*, only the antisaccade task had a significant correlation with *Gf*, ranging from .19–.23, while the stop signal and Stroop task were not significantly correlated, with a correlation ranging from .03–.12. For *Gc* the correlations for these three inhibitory tasks were low, .12–.19, though 4 out of 6 were significant (Friedman et al., 2006). These three tasks loaded significantly onto the same "inhibition" factor, but *SEM* showed that it did not explain any unique variance for *Gf* or *Gc* when other cognitive abilities were in the model (Friedman et al., 2006; Figure 1). This indicates that the modest correlations were the result of task impurity. Inhibition was not the cause of the correlations, but a correlation was found because of the other cognitive factors that were also being used, potentially WM. Yet, it is not possible to draw a strong conclusion about the relationship between inhibition and *g* because of relatively small amount of research that has been conducted.

Associative Learning

Associative learning is the ability to mentally link or associate specific stimuli together. One measure of associative learning is a simple discrimination task, where the participant must select between two stimuli. One of the stimuli is paired with a reward while the other stimulus is not. Selecting the rewarded stimulus does not seem to be related to intelligence in children or adults, but this is underexplored (Plenderleith, 1956; Stevenson & Zigler, 1957). The paired associates task and the three-term contingency task, however, show a more promising relationship with intelligence. In the paired associates task, participants are told to remember pairs of unrelated one-syllable words. During training, participants see the first word of the pair, then press a key to reveal the second word. A test usually follows immediately after the training phase, where the first word of the pair is given and the participant must type the second word (Alexander & Smales, 1997). A variation on this is the three-term contingency task. During training, one word serves as the stimulus and there are three response keys. When the participant presses the response key, a word is revealed. At test, the participant is shown the stimulus word and must type the correct word for each response key (Williams & Pearlberg, 2006). The paired associates and three-term contingency tasks are significantly correlated with each other, .43–.64, and with *g*, .31–.52 (Alexander & Smales, 1997; Kaufman et al., 2009; Tamez, Myerson, & Hale, 2008; Williams & Pearlberg, 2006). The paired associates and three-term contingency tasks are not pure measures of associative learning considering how much information needs to be stored and retrieved, which clearly relies on memory processes. As discussed earlier, WM and possibly STM are related to *g*. Nevertheless, it has been found using *SEM* that associative learning tasks are uniquely related to *g* independent of the memory and retrieval requirements (Kaufman et al., 2009; Figure 1). This suggests that associative learning is another potential underlying cognitive mechanism of *g*, but the task needs to be difficult or complex to reveal such a relationship.

Related Cognitive Factors–Summary

Four of the cognitive mechanisms discussed, WM, STM, processing speed, and associative learning are all related to *g* to varying degrees (see Figure 1). Why these factors are related is still being explored, and the relative importance of each factor is debated. The relationship between response inhibition and *g* is underexplored, but so far response inhibition does not seem to be related to *g* in any significant way. This is difficult to understand in the context of the relationship *g* has with task complexity since some measures of inhibition, like the Stroop task, appear more complicated than measures of processing speed, yet processing speed has consistent correlations with intelligence (Deary, 2000). Recently, how much unique variance processing speed can explain was called into question (Conway et al., 2002), but the consistency of the correlation is undeniable. Nevertheless, merely knowing which cognitive mechanisms are related to *g* does not provide much insight into what, exactly, *g* actually is. We know how to measure *g* and its validity as a predictor of many life outcomes, but over 100 years of research has yet to elucidate its exact nature. Investigating *g* and its psychological correlates in nonhuman animals (the focus of the next part of our review) would open up new avenues of research into the biological and empirical nature of *g*, and perhaps break through the current impasse in human research on the subject.

g in Nonhuman Animals

The *g* factor has been found consistently in human samples with a variety of measures, but what about other species? Finding a *g* factor in nonhuman animals would enable the study of many important questions about *g*, such as its evolutionary origins, its effects on biological fitness, and questions about mechanism that are challenging or even impossible to study in humans, such as the role of genes, its neural underpinnings, and environmental determinants during development. Useful animal models for studying *g* would allow the latest tools to be applied, such as optogenetics, chemogenetics, other gene-modification techniques (e.g., CRISPR), powerful control of the individual's environment from conception to adulthood, and other forms of neural manipulation. Application of these tools would allow for unprecedented insights into the causal role of genetic, neural, and environmental factors in *g* and intelligence. These insights could, in turn, provide translational significance to understanding *g* in humans. Working with animals, either in a lab or other setting, it is easy to see individual differences in task performance, but it is not clear if these differences would be consistent across a variety of tasks, like what we see in humans with the *g* factor (Macphail, 1987).

Research on nonhuman animals has shown they are capable of extraordinary cognitive feats, like tool use in New Caledonian crows (Auersperg, von Bayern, Gajdon, Huber, & Kacelnik, 2011), the range of abilities demonstrated by Alex the African Gray parrot (Pepperberg, 2018), and many more than what can be listed here. While impressive in their own right, extraordinary performance by particular species, be it in a single, specialized task (e.g., the spatial memory of the Clark's nutcracker, Balda & Kamil, 1992; or tool use by the New Caledonian Crow), or by only a few subjects across many tasks (e.g., by the African Gray parrot, the Bottlenose dolphin; Herman, 2010) does not provide insights into psychometric *g* that would be provided by consistent performances across many tasks whereby stable individual differences

replicate key aspects of psychometric *g* in humans (Macphail, 1987). These exceptional animals cannot be meaningfully discussed in a review about *g* since they have not been given test batteries designed to determine cross-task consistency in performance. The purpose of exploring the potential for psychometric *g* in other species is not to rank species in their intelligence. Rather, developing test batteries that can be applied across species can help illuminate the conditions under which cognitive abilities will show a pattern of positive correlations. Thus, an animal model for measuring *g* would open up new avenues of research into the environmental and neural contributions to psychometric *g*, which can inform on theories of the causes for the correlations that determine *g*.

To meaningfully relate task performance to *g* in nonhuman animals requires reliable measures of performance in standardized behavioral tasks. Recently, researchers have been investigating individual differences in cognition in nonhuman primates, mice, and birds using test batteries (Burkart, Schubiger, & van Schaik, 2017; Shaw & Schmelz, 2017). Some of these test batteries, however, are inadequate. Some suffer from including too few tasks (Anderson, 1993), or the included tasks lack sufficient variety to derive meaningful individual differences (Locurto & Scanlon, 1998). Other batteries include tasks that are ill-defined and, therefore, obscure the underlying constructs (Keagy, Savard, & Borgia, 2011). Finally, some test batteries do not adequately control for the way in which a particular species interacts with their environment, or noncognitive differences between species, such as motivation (Bitterman, 1965; Macphail, 1987). For example, it may be difficult for a subject to use a tool with their beak when the tool was designed to be used with a hand (Krasheninnikova, Berardi, Lind, O'Neill, & von Bayern, 2019). Thus, while many studies of *g* in animals have found positive correlations across tasks, these deficiencies make it difficult to relate these studies to the *general* cognitive ability found in humans.

Nevertheless, some test batteries used in nonhuman animal research have enabled assessment of the underlying cognitive abilities (see Table 1). We focus the remainder of this review on these stronger test batteries that provide evidence for a general factor of intelligence. This necessarily restricts our discussion to species for which sufficiently strong data are available. As a reminder, we are also not focusing on species differences in intelligence, but rather individual differences in psychometric *g* for various species. Thus, relatively *smart* species, such as crows, parrots, and dolphins, are not included, while cognitively humble species, such as mice, are. It is beside the point whether parrots are deemed *smarter* than pigeons, or that apes are *smarter* than mice, as we are not concerned with ranking species intelligence against each other, but rather finding in nonhuman populations, similar individual differences as have been consistently found between individual people. Indeed, even demonstrations of differences in cognitive prowess of various species are not sufficient evidence for true species differences in *general* cognitive abilities (Burkart et al., 2017; Macphail, 1987). Until these species are given a diverse battery of tests, it is impossible to comment on the consistency of performance across tasks, which is at the center of *g* in research on humans.

Using appropriate test batteries, evidence for correlations within subject have been found in chimpanzees (Hopkins, Russell, & Schaeffer, 2014; Woodley of Menie et al., 2015), cotton-top tama-

rin monkeys (Banerjee et al., 2009), rhesus macaques (Herndon, Moss, Rosene, & Killiany, 1997), orangutans (Damerius et al., 2019), mice (Galsworthy, Paya-Cano, Monleón, & Plomin, 2002, 2005; Kolata et al., 2005; Kolata et al., 2007; Matzel et al., 2003; Matzel et al., 2006), robins (Shaw, Boogert, Clayton, & Burns, 2015), bowerbirds (Isden, Panayi, Dingle, & Madden, 2013), and magpies (Ashton, Ridley, Edwards, & Thornton, 2018). The general factor found in these studies can explain from 18 to 64% of the variance in individual performance. Performance has been related to WM (Kolata et al., 2005) and is stable over long periods of time (Ashton et al., 2018; Hopkins et al., 2014). Despite this, these test batteries sometimes fail to extract a general factor, such as studies in chimpanzees (Herrmann, Call, Hernàndez-Lloreda, Hare, & Tomasello, 2007, 2010), mice (Locurto, Fortin, & Sullivan, 2003; Locurto, Benoit, Crowley, & Miele, 2006), and song sparrows (Anderson et al., 2017; Boogert, Anderson, Peters, Searcy, & Nowicki, 2011). We next explore why evidence for a *g*-like factor in nonhuman species is not as reliable as in the human literature. We also discuss the value of *g* in nonhuman species in predicting fitness related outcomes.

Nonhuman Primates

Herrmann et al. (2007) developed the primate cognitive test battery (PCTB) to assess performance across human children (age 2.5 years) and adult nonhuman primates. The PCTB includes 15 tasks from social and physical cognitive domains, such as the understanding of physical objects, social cues, and causal relationships (see Table 1). This test battery was specifically created to test different evolutionary theories on why humans seem to show more advanced cognitive abilities compared with nonhuman primates, which is why tests of social abilities are included (Herrmann et al., 2007). Approximately the same test battery was given to children, chimpanzees, and orangutans, but the analyses conducted did not allow for the examination of how individuals performed across all tasks. A follow up article examined the results from the children and chimpanzees to determine the structure of these cognitive abilities (Herrmann, Hernández-Lloreda, Call, Hare, & Tomasello, 2010). Using CFA, for children they found evidence for three factors, physical, social, and spatial, which is surprising considering other research has shown evidence for a general factor for children in this age range (Spinath et al., 2003). For chimpanzees, they found evidence for two factors, spatial and physical-social that account for individual differences in performance. While the initial research indicated there may be a relationship in chimpanzees between boldness and performance on the physical tasks, where bolder chimpanzees had better performance (Herrmann et al., 2007), this relationship was not further elaborated (Herrmann et al., 2010). The lack of a *g* factor for either species is surprising, though this could be because of a number of factors, particularly when examining the results for the chimpanzees. The authors acknowledge that their test battery contains a much higher proportion of social tasks compared with what is typically found in the literature (Spinath et al., 2003). They also acknowledge that a number of their test items had low variabilities, though they do not specifically state which tasks (p. 108). Finally, the reliabilities of the tasks for the chimpanzee sample ranged from .05–.66, which as discussed earlier, can weaken subsequent correlations. It is not

Table 1

Summary of Cognitive Test Battery Research in Nonhuman Primates, Rodents, and Birds

Species (n)	Tasks	Correlational matrix	Analysis	g?	Other notable findings	Reference
			Primates			
Rhesus monkeys (30)	Delayed nonmatch to sample (10 and 120 s delay), delayed recognition span task—spatial and color, reversal learning—spatial and object	Not shown	PCA	Yes 48%	Performance declined with age	Herndon, Moss, Rosene, and Killiany (1997)
Cotton-top tamarin monkeys (22)	Occluded reach, targeted reach, A-not-B, reversal learning, exploration, numerical discrimination, acoustic discrimination, object tracking social tracking, hidden reward retrieval, food extraction puzzle	Not shown	Bayesian latent variable analysis	Yes, unclear how much the g factor accounted for		Banerjee et al. (2009)
Chimpanzees (106)	Spatial memory, object permanence, rotation, transposition, relative numbers, addition numbers, causal noise, causal shape, tool properties, social learning, comprehension, pointing cups, attentional state, gaze following, intentions	Not shown	EFA and CFA	No; two factors spatial and physical/social		Herrmann, Hernández-Lloreda, Call, Hare, and Tomasello (2010)
Chimpanzees (99)	Spatial memory, object permanence, rotation, transposition, relative numbers, causal noise, causal visual, tool use, tool properties, comprehension, production, attention state, gaze following	Not shown	PCA	Yes 54%	The g factor and two of the other components were highly heritable; a retest 2 years later showed consistent performance	Hopkins, Russell, and Schaeffer (2014)
Orangutans (53)	Box task, detour tube task, tube trap task, honey tool task, associative and reversal learning	Majority positive, none reached significance	PCA (confirmed with EFA)	Yes 36%	Curiosity correlated with g in captive only, not wild types	Damerius et al. (2019)
			Rodents			
Mice (40)	Open field, spontaneous alteration in a T maze, Hebb-Williams, MWM, burrowing task, contextual memory, plug puzzle	Majority positive, 8/28 significant	PCA	Yes 31%		Galsworthy, Paya-Cano, Monleón, and Plomin (2002)
Mice (56)	Lashley maze, passive avoidance, MWM, odor discrimination, fear conditioning, control: open-field exploration, defecation in water/ novel environments	Uniformly positive, 2/10 significant	PCA	Yes 38%	Propensity to explore was correlated with 4/5 tasks	Matzel et al. (2003)

(table continues)

Table 1 (*continued*)

Species (*n*)	Tasks	Correlational matrix	Analysis	*g*?	Other notable findings	Reference
Mice (60)	Hebb-Williams, plus maze, radial arm maze, visual nonmatch to sample, detour problems, control: light dark, activity measures in land and water	Majority positive, 3/15 significant (excluding controls)	PCA	No, four components that all explained the same amount of variance	Included the control measures in the PCA analysis; counterbalanced the task order	Locurto, Fortin, and Sullivan (2003)
Mice (21)	Lashley maze, passive avoidance, MWM, odor discrimination, fear conditioning, radial arm maze with delay, dual radial arm maze, control: open field exploration	All in the appropriate direction, 9/28 significant	PCFA	Yes 40%	Only Working memory capacity was correlated with performance on the learning battery	Kolata et al. (2005)
Mice (Experiment 1 = 84, Experiment 2 = 167)	Experiment 1: spontaneous alt. in T-maze, Hebb-Williams, MWM, burrowing puzzle, plug puzzle; Experiment 2: all tasks as Experiment1 plus MWM reversal, syringe puzzles, water plus maze, object exploration	Experiment 1: uniformly positive, 6/15 significant; Experiment 2: majority positive, 15/55 significant	PCFA (replicated w/ PCA and PFA)	Experiment 1: yes 36%; Experiment 2: yes 22%	Males outperformed females on all tasks	Galsworthy et al. (2005)
Mice (43)	Lashley maze, passive avoidance, MWM, odor discrimination, fear conditioning, plus maze (spatial), control: exploratory behaviors, sensory/motor function, stress, fear, pain reactivity	Not shown	PCFA	Yes 32%	Exploratory behavior loaded onto the general factor	Matzel et al. (2006)
Mice (Experiment 1 = 47, Experiment 2 = 51)	Experiment 1: detour, win-shift, olfactory discrimination, fear conditioning, operant acquisition, control: light dark test, open field; Experiment 2: detour, Hebb-Williams, radial arm maze, olfactory foraging, fear conditioning, control: light dark test, open field	Experiment 1: half positive, 4/10 significant; Experiment 2: majority positive, 3/10 significant (excluding controls)	PCA	Experiment 1: no, two components extracted; Experiment 2: yes 34% (controls excluded)	For Experiment 1 3/4 significant correlations were negative. When control measures were included in the PCA, 3 independent components were extracted	Locurto et al. (2006)
Mice (27)	Lashley maze, passive avoidance, MWM, odor discrimination, fear conditioning, mouse Stroop, nonspatial radial arm maze, delayed reinforced alternation, control: open-field exploration	All in the appropriate direction, 5/15 significant	PCFA	Yes 43%	Selective attention had the strongest correlation, short term memory capacity was modest, and duration had the weakest	Kolata, Light, Grossman, Hale, and Matzel (2007)
Mice (244, combined from prior studies)	Lashley maze, passive avoidance, MWM, odor discrimination, fear conditioning, spatial win-stay (*n* = 98), reinforced alternation (*n* = 78), control: open-field exploration, defecation in water/ novel environments	Uniformly positive, 9/10 significant	PCFA verified with CFA	Yes 38%	Spatial group factor with a hierarchical design	Kolata, Light, and Matzel (2008)

(table continues)

Table 1 (*continued*)

Species (*n*)	Tasks	Correlational matrix	Analysis	*g*?	Other notable findings	Reference
Mice (26)	Mouse Stroop, T-Maze reversal, latent inhibition, dual radial arm maze, odor discrimination, reinforced alternation, radial arm maze	Not shown	EFA and CFA	Yes 37%	External attention was significantly related to the general factor	Sauce, Wass, Smith, Kwan, and Matzel (2014)
Birds						
Song sparrows (52)	Novel motor task, color association, color reversal (2009, 2010), tube task (2010)	Majority positive, none reached significance	PCA	No, two components extracted	Not related to song repertoire size	Boogert, Anderson, Peters, Searcy, and Nowicki (2011)
Spotted bowerbird (14)	Barrier removal, novel motor task, color discrimination, color reversal, shape discrimination, spatial memory	Majority positive, none reached significance	PCA (on 11 subjects)	Yes 44%	Not related to mating success	Isden, Panayi, Dingle, and Madden (2013)
Robins (16)	Motor task, color discrimination, color reversal, spatial memory, tube task, symbol discrimination	Majority positive, none reached significance	PCA	Yes 34%		Shaw, Boogert, Clayton, and Burns (2015)
Song sparrows (41)	Novel motor task, color association, color reversal, tube task, spatial learning	Majority negative, 1/10 significant (positive)	PCA	No, two components extracted		Anderson et al. (2017)
Magpies (56)	Color association, color reversal, tube task, spatial learning	Uniformly positive and significant	PCA	Yes 64%	Larger group size was related to better cognitive performance. Better cognitive performance in females resulted in better offspring success	Ashton, Ridley, Edwards, and Thornton (2018)

Note. MWM = Morris water maze; CFA = confirmatory factor analysis; PCA = principal component analysis; EFA = exploratory factor analysis; PCFA = principal component factor analysis.

entirely clear how or if this was controlled for in their subsequent analyses.

Another group of researchers used a modified version of the PCTB and found evidence of a *g* factor in chimpanzees using principal component analysis (PCA) that was stable across 2 years and was heritable, consistent to what is seen with humans (Hopkins et al., 2014). However, it is unclear how much variance in performance is explained by this *g* like factor in chimpanzees or if the modified version of the PCTB changed the task reliabilities. Additionally, while age and sex were collected as potentially confounding variables, it is not clear if any personality measures, like boldness, were taken. Using the same data set, a follow up study used different statistical techniques and confirmed both the presence of a single factor and its heritability, but it is still unclear how much variance is explained by this factor (Woodley of Menie et al., 2015). In an attempt to resolve the discrepancy in results, a reanalysis combined both data sets (Kaufman, Reynolds, & Kaufman, 2019). Using CFA, they found evidence for a *g* factor and group factors for chimpanzees and children, however, the exact structure of these factors was different between the two species. Additionally, it was unclear how much variance in performance was explained by *g*. For chimpanzees, they confirmed that this was relatively stable over time, though performance tended to improve during the second test. They reported that the stability coefficient for the PCTB was .5 compared with .96 for FSIQ tests given to children.

Evidence of *g* has also been found in cotton-top tamarins and orangutans using different test batteries for each (Banerjee et al., 2009; Damerius et al., 2019). The tamarins were tested on 11 tasks, including social tracking, reaching, and reversal learning (Banerjee et al., 2009). Participation in all tasks was voluntary. Data were collected in the form of ranks and Bayesian latent variable analysis was used. Using this method, they found evidence for a *g* factor, but no evidence for distinct group factors. They acknowledge, however, that the lack of group factors could have been because of low levels of reliability for some of the tasks. The orangutans were tested on five tasks, including response inhibition, causal reasoning, and reversal learning, all showing high levels of variability (Damerius et al., 2019). Using PCA, one factor was extracted that explained 31.28% of the variance in performance and all tasks loaded onto this factor, similar to what is seen in humans. While this research was conducted with orangutans at rehabilitation centers, there was variation in how much of their development occurred in the rehabilitation center versus the wild, which was related to differences in noncognitive factors. For nonwild subjects, there was a positive relationship between curiosity and *g*.

Rhesus macaques have also been given a test battery that included six tasks, including delayed nonmatch to sample (DNMS) and reversal learning, but the goal of this study was to determine if there were age related cognitive declines in this species (Herndon et al., 1997). Using PCA, the first component extracted accounted for 48% of the overall variance, but was significantly negatively correlated with age, indicating that older subjects performed worse on all tasks. While *g* is stable across individuals over time in human populations, there is evidence for age related declines in cognitive abilities, and that these declines are independent of *g* (Gow et al., 2011). The relationship between *g* and age-related cognitive decline is complicated and outside of the scope of this review article, but the research by Herndon et al.

(1997) indicates that it is likely that rhesus macaques have a *g* like factor.

One of the key differences from human research in test batteries for nonhuman primates and other species is the inclusion of social tasks (Banerjee et al., 2009; Herrmann et al., 2010; Hopkins et al., 2014; Table 1). In human research, intelligence and social ability appear to be separable domains and dissociable. People can show an impairment in social ability while performing normally on IQ tests, and vice versa (Adolphs, 1999). When humans with intact and normal brain functioning were tested on both measures of *g* and social knowledge, the correlation between the two measures was quite low (Derksen, Kramer, & Katzko, 2002). Nevertheless, this low correlation could also result from comparing the subjective self-report measure of social knowledge to the more objectively measured *g* (Derksen et al., 2002). Studies with human children and adolescents indicate that general intelligence and "Theory of Mind," or the ability to understand the mental state of another, are independent (Cavojová, Mikušková, & Hanák, 2013; Rajkumar, Yovan, Raveendran, & Russell, 2008), but these populations are older than the participants tested by Herrmann et al. (2007, 2010). For nonhuman primates, inclusion of six social tasks in the PCTB also failed to find a *g* factor (Herrmann et al., 2010). Others suggest, based on reanalysis of these data, that these social tasks could be equivalent to *Gc*, the cultural-knowledge group factor seen in humans (Kaufman et al., 2019). This suggestion is premature, however, given that the operational definition and assessment of *Gc* in humans varies widely across labs (Kan et al., 2011; Keith & Reynolds, 2010; Schipolowski et al., 2014). The relationship between social ability, cultural knowledge, and general cognitive abilities should be tested more thoroughly in humans throughout the life span to better establish their relationship.

Mice

The cognitive abilities for mice have been heavily explored by Locurto, Galsworthy, and Matzel (see Table 1). Test batteries typically include measures of WM, associative fear learning, olfactory discrimination, and spatial memory, though the content and quantity of tasks varies. Unlike primate test batteries, mouse batteries frequently include measures of anxiety and overall activity levels, likely because these emotional responses are frequently studied in mice, especially in connection to fear learning and drug effects. Across a series of experiments, Locurto et al. (2003) and Locurto et al. (2006) devised cognitive test batteries for mice consisting of a visual nonmatch to sample (NMTS) task, spatial NMTS, spatial learning (Hebb-Williams Maze), detour problems, WM, place learning, olfactory learning and discrimination, fear conditioning, and operant acquisition. Briefly, the WM tasks have been a four (Locurto et al., 2006) or eight-arm radial maze and a variation of the radial maze task called the 4×4 task (Locurto et al., 2003). In the radial arm maze, there is a central platform with *n* enclosed arms radiating from it. Each arm contains a food reward and the subject is allowed to freely sample any arm at any time. Subjects entering an arm and failing to obtain the reward or entering an arm again after already obtaining the food reward were counted as WM errors. The 4×4 task also took place in the eight-arm radial maze. In the first phase, four of the arms contained a food reward, while the other four were blocked off. Once the animal had sampled all of the rewards, they were removed

from the maze for 30 s. In the second phase, all of the arms were open, but only the four previously blocked arms were baited. Entering arms that had been rewarded in the first phase, entering the same arm twice in the second phase, and entering an arm for the first time, but failing to obtain the food reward were all counted as errors. The control procedures measured activity levels on land and in water in an open field chamber, and a light-dark preference test. The number of transitions in the light-dark chamber and the distance traveled in the open field chamber was termed activity. The time spent next to the wall in the open field chamber negatively correlated with the number of center crosses in the open field chamber. Together they were counted as an anxiety measure. The same control procedures were used with all cognitive test batteries (Locurto et al., 2003; Locurto et al., 2006).

For the research conducted in 2003, the subjects were trained on the cognitive tasks until their performance reached asymptote. Multiple dependent measures were taken from each task and an aggregate score was used in the analysis, which had a reliability of .88. The average correlation between the cognitive tasks, however, was .12. When the correlational matrix of cognitive tasks and control measures was subjected to PCA, multiple independent factors were extracted (Locurto et al., 2003). In the follow up study of 2006, subjects were given fewer trials on the cognitive tasks, and only one dependent measure was used in subsequent analyses. This reduced reliability to .54 in Experiment 1 and .58 in Experiment 2. The average correlations between the learning tasks for these experiments were −.03 and .15, respectively. The authors state, "The relatively low reliabilities in the present study contributed to the relatively low average correlations observed," yet it does not appear as though these correlations were corrected for measurement unreliability (Locurto et al., 2006 p. 382). PCA, including the control measures, revealed a similar result, where multiple independent factors were extracted.

Other researchers have not had similar results, even when using the some of the same tasks. Galsworthy and colleagues have also tested mice on a diverse battery of cognitive tests, but have found evidence for *g*. In 2002, Galsworthy et al., tested mice with two measures of spatial learning (Hebb-Williams and Morris water maze), spontaneous alternation in a T-shaped maze, a detour task, contextual memory, and a problem-solving task. Multiple dependent measures were used in the correlational matrix for some of these tasks. The reliabilities of these tasks ranged from .68−.84. The control procedures measured anxiety with an open field arena, defecation in testing environments, and latency to swim to a visible platform. A correlation matrix with all of the cognitive tasks and the spontaneous alternation task revealed that a majority of the tasks were positively correlated, and some of the positive correlations were nonsignificant, with an average correlation of .2. When a PCA was conducted, the first component explained 31% of the total variance in performance. A separate PCA was conducted on the measures of anxiety. They found that the first component could explain 46% of the variance in anxiety, but this component did not significantly correlate with any of the cognitive measures or their *g*-like factor.

In a follow-up study, Experiment 1 used essentially the same test battery, but for Experiment 2 it was expanded to include a spatial reversal in the Morris water maze, a water plus maze, novel object exploration, and an additional problem-solving task (Galsworthy et al., 2005). Additionally, in Experiment 2, many of the

tasks were shortened. For these experiments only one dependent measure per task was used in the correlational matrix and an aggregate performance score was used when appropriate. Reliabilities were only reported for each dependent measure, however, not the aggregate. For Experiment 1, reliabilities ranged from .47−.87, and in Experiment 2 they ranged from .03−.78. The mean correlation was .18 and .06, respectively. A principal component factor analysis (PCFA) resulted in one factor that could account for 32% of the variance in Experiment 1, and 19% of the variance in Experiment 2. They acknowledge that the low task reliabilities could have attenuated the subsequent *g* factor, but did not indicate that the correlations had be corrected to compensate for this (Galsworthy et al., 2005 p. 688).

Studies conducted in Matzel's lab used a test battery that consisted of egocentric navigation (Lashley III maze), passive avoidance, spatial learning (Morris water maze), odor discrimination, and fear conditioning (Kolata et al., 2005; Kolata et al., 2007; Kolata et al., 2008; Matzel et al., 2003; Sauce, Wass, Smith, Kwan, & Matzel, 2014). These tasks were administered in such a way to ensure variability between subjects and capture differences in learning (Kolata, Light, & Matzel, 2008) An open-field arena was used to determine anxiety and activity levels. An analysis similar to Galsworthy et al. (2002, 2005) was conducted. Performance on the cognitive tasks showed a uniformly positive correlational matrix and PCA extracted one component that explained 38% of the variance (Matzel et al., 2003). When the behavior in the open field was analyzed, only the amount of time spent away from the walls was significantly related to the general factor. This type of behavior, spending time in the open part of the open-field arena, is thought to reflect novelty seeking. As with humans, subsequent studies found this factor to correlate with WM, which was assessed with two 8-arm radial mazes (Kolata et al., 2005; Sauce et al., 2014). This factor also correlated with performance on a mouse version of the Stroop task (Kolata, Light, Grossman, Hale, & Matzel, 2007). In humans, such a relationship has received only mixed support, however, it is underexplored (Friedman et al., 2006; Polderman et al., 2009). Pooling across prior data sets (*n* = 241) produced a sample size with substantially more power. With this sample, the average correlation was .22, a magnitude similar to what they had found in the individual studies, but they did not report the task reliabilities. PCA confirmed a general factor that accounted for 38% of the variance and identified a potential group factor of spatial ability (Kolata et al., 2008). This strengthens the similarity between humans and mice in the structure of cognitive abilities.

To recap, for mice, one lab has had consistent success in capturing a general factor for cognition using their test battery (Kolata et al., 2005; Kolata et al., 2007; Kolata et al., 2008; Matzel et al., 2003; Sauce et al., 2014), while other labs have had more inconsistent results (Galsworthy et al., 2002; Galsworthy et al., 2005; Locurto et al., 2003; Locurto et al., 2006; Table 1). One key difference comes from how control measures are incorporated into the data analysis. When the control measures are entered into the factor analysis, a *g* factor is not extracted (Locurto et al., 2003; Locurto et al., 2006). When the control measures are subjected to a separate factor analysis, and a correlational analysis is used to determine if the factors are related, typically a *g* factor that can account for approximately 30% of the variance is found (Galsworthy et al., 2005; Kolata et al., 2008; Matzel et al., 2003). The latter

is the method of correlated vectors, and while some human researchers have advocated for its use (Jensen & Weng, 1994) other researchers have identified potential issues with its use (Ashton & Lee, 2005; Wicherts, 2017). It is also not always clear why certain dependent measures are being collected in cognitive tasks with mice (Locurto et al., 2006). Unlike in human intelligence tests where there is one dependent measure for each task, with mouse studies multiple measures are typically collected. If the rodent *g* is as robust as the human *g*, we would expect to see a similar positive correlational matrix in each species, regardless of task battery composition or dependent measures collected. When constructing test batteries for humans, however, tasks are chosen specifically because they are known to load highly onto *g*, and avoided if they do not. This bias could artificially strengthen the correlation between tasks (Locurto et al., 2006). The discrepancy between Galsworthy's, Locurto's, and Matzel's labs in data analysis and success in finding a general factor should be investigated further, possibly by standardizing certain methods to ensure minimum between lab variation.

Avian Species

The structure of cognition has also been explored in a wide variety of avian species, including song sparrows (Anderson et al., 2017; Boogert et al., 2011), robins (Shaw et al., 2015), spotted bower birds (Isden et al., 2013), and Australian magpies (Ashton et al., 2018; Table 1). Given the more distant relationship between birds and mammals (~350 mya), investigation of *g* in birds could provide insight into the phylogenetic depth of general intelligence (see Figure 2). Similar results across birds and mammals could also result from convergent evolution, where a general cognitive factor evolves independently across multiple species because of similar environmental conditions or social structures. Likewise, since research with birds, especially pigeons, often uses similar methods and procedures as used in human cognitive research (e.g., behavioral psychophysics experiments using visual touchscreen operant chambers), birds provide a powerful tool, similar to nonhuman primates, with which to tease apart the relationship between *g* and its underlying cognitive components. For nonpigeon avian research, test batteries typically consist of acquisition of novel operant behavior, discrimination learning, reversal learning, spatial/reference memory, and response inhibition (see Table 1). Re-

sponse inhibition is assessed with a detour tube task. In this task, subjects are presented with a transparent tube with a visible food reward inside. The tube is positioned such that the subject must inhibit the direct approach to the food, and instead move away from the reward to access it from the side of the tube (Kabadayi, Bobrowicz, & Osvath, 2018; van Horik et al., 2018).

Wild male song sparrows were administered the motor learning, color association, color reversal, and the detour task in a laboratory environment. The number of songs in their repertoire was also collected. Song learning is thought to encompass cognitive abilities because of the process of learning songs from other males, directly or through recordings, during the critical period early in life. Once males reach sexual maturity, they produce crystalized song typical of adults of that species. If song learning was influenced by general cognitive ability, it would be a potential mechanism for mate choice for cognition (Boogert et al., 2011). The correlational matrix for the cognitive abilities was not uniformly positive and the average correlation was .248. PCA extracted two components, where the first component accounted for 45% of the variance and the second component accounted for 33% of the variance. The color association and color reversal learning tasks loaded positively onto the first factor, the motor learning task had a weak negative loading, and detour performance had a strong negative loading. This negative loading indicates that the detour tube task is measuring something different compared with the other tasks. Song repertoire size showed a complicated relationship with these cognitive tasks. Larger song repertoires were associated with faster performance on the detour task, but slower performance on the reversal learning task. Song repertoire size was negatively correlated with detour performance, however, meaning that birds with a larger repertoire were faster at the detour task. The researchers acknowledge that differences in noncognitive factors like personality and experience could have influenced performance on these measures and be a potential factor in why a *g* like factor was not found.

A similar test battery, but with the inclusion of a spatial/reference memory task, was given to hand-reared male and female song sparrows (Anderson et al., 2017). Two measures of song accuracy were assessed in addition to repertoire size. The correlational matrix was not uniformly positive and many correlations were weak. The average correlation for males (*n* = 19) was .101, but this actually decreased to .036 when females were added to create a larger sample size (*n* = 38–41). PCA was conducted with the correlational matrix from the male subjects and two components were extracted from this test battery. Similar to the results with the wild population, the color association, reversal, and spatial learning task loaded positively onto the first component, but the detour task loaded negatively (Anderson et al., 2017; Boogert et al., 2011). All measures of song performance were positively correlated but, in contrast to the wild population, better performance on color reversal was associated with higher song quality while better performance on the detour task was associated with *poorer* song quality. This further emphasizes that cognitive abilities in sparrows do not show a uniform relationship (Anderson et al., 2017; Boogert et al., 2011). While these studies did not allude to task reliability, a follow up study investigated the consistency of performance across time (Soha, Peters, Anderson, Searcy, & Nowicki, 2019). Subjects were tested once a year for 2 or 3 years on the test battery used by Anderson et al. (2017). Performance and
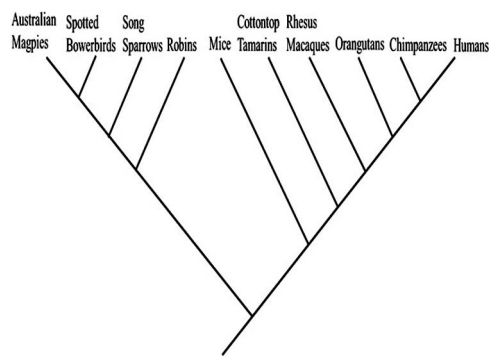

Figure 2. A cladogram of the species reviewed that have been given cognitive test batteries.

relative rank were not consistent across years for males or females, with the average correlation across time being .13. The relationship that cognitive performance had with measures of song accuracy also varied across years. This variance over time makes interpreting the initial studies difficult.

Research with other avian species has produced similar correlational matrices to what was found with song sparrows, though with stronger evidence for a general cognitive factor. Bower birds were given a problem-solving task, where they had to remove a novel barrier, novel motor learning, color discrimination and reversal, shape discrimination, and spatial memory. A majority of the correlations were positive and the average correlation was .26. PCA extracted two components, where all tasks loaded positively on the first component and it accounted for 44% of the variance, indicating a general factor. Whether this general factor was related to mating success was also studied. Male bower birds build elaborate nests (bowers), which appears to be a cognitively demanding task, to attract mates. Similar to song sparrows, if this nest building ability is generally related to performance on other cognitive tasks, it could be used as a signal by females for mate selection. Yet, no consistent relationship between mating success and cognitive measures has been found (Isden et al., 2013). Wild robins were administered the same test battery as described by Anderson et al. (2017). A majority of the correlations were positive, with an average of .158, and PCA extracted two components. All tasks loaded positively onto the first component and it accounted for 34.46% of the total variance. The loadings onto the first component were strengthened after removing potential noncognitive confounds, like innate color preference (Shaw et al., 2015). However, the reliability of these tasks was not given.

A study with Australian magpies administered the same test battery that was given to song sparrows (Anderson et al., 2017) and robins (Shaw et al., 2015), though time to learn a novel motor behavior was not included in the correlational matrix or the PCA. They found uniformly strong positive correlations among cognitive tasks, with an average correlation of .465, and when given similar tasks 2 weeks later, performance was very reliable (.806–.975). PCA extracted one component that explained 64% of the variance in performance. Group size was related to this factor, where subjects living in larger groups performed better on these cognitive tasks. Furthermore, maternal cognitive ability was found to be the best predictor of reproductive success as measured by the number of fledglings produced and the number that survived to adulthood (Ashton et al., 2018). This contrasts with earlier studies that had only looked at the mating performance of males (Anderson et al., 2017; Boogert et al., 2011; Isden et al., 2013).

Thus, as with nonhuman primates and mice, evidence for g in avian species has yielded mixed results. In the song sparrow, performance on the detour task has a negative loading on the first factor extracted (Anderson et al., 2017; Boogert et al., 2011). Negative loadings are not seen in human studies of intelligence unless better performance is measured in the opposite direction of the other tasks (Jensen, 1998). In contrast, for the remaining species, robins, spotted bower birds, and Australian magpies, performance on all tasks showed positive loadings on the first factor and the first component accounted for an average of 47% of variance in performance (Ashton et al., 2018; Isden et al., 2013; Shaw et al., 2015). The detour task itself could be the reason for the different pattern of results. Follow-up studied with robins and

pheasant chicks found that better (i.e., healthier) body condition and experience with transparent objects reduced the number of ineffective pecks to the transparent wall (Shaw, 2017; van Horik et al., 2018). Noncognitive factors could be influencing performance on the detour task and obscuring a general factor in sparrows.

It is also possible that these species are under different evolutionary pressures that has created differences in how cognitive abilities are related. The predictive value of g in humans has been strongly linked to outcomes that are the products of cultural evolution that themselves can vary substantially across individual, such as occupation and education attainment. Thus, exploring g in an ecological/evolutionary context could help illuminate why certain tasks load more highly onto g than others. Avian species that show nonsignificant positive and negative correlations on these cognitive tasks might be under different evolutionary pressures than those showing significantly, uniformly positive correlations. Evolutionary theories are elaborated on later, but briefly, Australia is home to spotted bower birds, which has a weak correlational matrix, and magpies, which has uniform, positive correlations. These species differ in terms of how they interact with conspecifics and humans, with bower birds being more isolated, which could be driving differences in how performance on these tasks are related (Ashton et al., 2018; Isden et al., 2013). The consistency of the test batteries given to these different avian species makes it easier to theorize about which factors are causing the differences in performance, a strength of the studies conducted so far. Testing wild subjects allows for a more nuanced understanding of how cognitive ability can impact reproduction and survival, and how this could interact with environment and social structure.

One species that is conspicuously absent from avian studies of intelligence is the pigeon. This is surprising given their long history as research subjects in psychology. Pigeons show evidence of cognitive processes typically studied in human and nonhuman primates, such as abstract reasoning (Blaisdell & Cook, 2005; Katz & Wright, 2006), rule learning (Garlick, Fountain, & Blaisdell, 2017), WM (Cook & Blaisdell, 2006; Kangas, Berry, & Branch, 2011; Lind, Enquist, & Ghirlanda, 2015), associative learning (Cook, Levison, Gillett, & Blaisdell, 2005), artificial grammar learning (Herbranson & Shimp, 2008), inhibition of return (Cook, Katz, & Blaisdell, 2012), memory interference dynamics (Wright, Santiago, Sands, Kendrick, & Cook, 1985), and choice RT (Vickrey & Neuringer, 2000). Often these cognitive processes are assessed in similar ways in pigeons as they are in human and nonhuman primates facilitating cross species comparisons. These cognitive processes seem to be supported by similar neuroanatomical structures, indicating that there is some restriction on how certain cognitive abilities evolve (Colombo & Broadbent, 2000; Colombo & Scarf, 2012; Divac, Mogensen, & Björklund, 1985; Güntürkün, 2005). These similarities suggest that, if pigeons were given a comprehensive test battery, a g factor would emerge. Frequently, pigeon researchers use the same subjects across multiple experiments, so it is likely that many labs already have assessed subjects on a variety of cognitive tasks, making it even more surprising that no lab has yet correlated their performance on different tasks. We are currently assessing pigeons on a test battery to measure g, with the addition of a novel reasoning task—a modified version of the RPM (mRPM; Flaim & Blaisdell, 2020)—an assessment of Gf in humans (Raven, 2008). This factor loads highly onto g, yet is completely absent from any study in

nonhumans. Briefly, for the mRPM task, subjects must learn an abstract rule to identify the rewarded stimulus, and transfer learning of that rule to novel stimuli. Preliminary results indicate that the mRPM is sensitive to individual differences in rule acquisition and transfer (Flaim & Blaisdell, 2020).

Discussion

Most animal studies have revealed a similar cognitive structure as found in humans. Nevertheless, weaknesses in animal test batteries make it difficult to determine if they extract the same factor across species. Test batteries for nonhuman animals sometimes assess abilities that are underexplored in humans. In primate studies, for example, social tasks are frequently included (Derksen et al., 2002), while avian batteries always include a measure of inhibition—both of which are underexplored in humans (Dempster, 1991). Even when the same cognitive abilities are tested, the methods are vastly different. The eight-arm radial maze is commonly used in animal studies of WM, but this is not the way WM is assessed in human studies of *g* (Conway et al., 2005). WM is theorized to have domain *general* properties that would result in similar performance across specific task stimuli, as long as the tasks had similar demands (Unsworth et al., 2009). The eight-arm radial maze appears as though it has similar task characteristics to the WM tasks given to humans, since the subject has to maintain and update a list of locations within the trial. This would indicate that it is a valid measure to investigate WM across species, even though the particular task format has been designed to take advantage of the rat's species-specific tendencies. Nevertheless, as discussed earlier, WM is unlikely to be a unitary construct. Different aspects of WM have been emphasized in the different tasks used with humans, and each underlying aspect has shown a relationship with *g* (Kane et al., 2007; Unsworth & Engle, 2007; Unsworth et al., 2009). One group of researchers has investigated how these different aspects of WM are related to the *g* factor in mice, providing further evidence that WM and its relationship to *g* is similar across species (Kolata et al., 2007), but more research needs to be conducted before forming strong conclusions. Human performance on an eight-arm radial maze should be compared with more traditional measures of WM and measures of *g* in nonhumans (Astur, Tropp, Sava, Constable, & Markus, 2004). Research with nonhuman animals should investigate WM with a broader array of tasks to determine if it also shows similar domain general properties and specific underlying processes (Kolata et al., 2007; Shaw & Schmelz, 2017).

Associative learning tasks pose a similar issue. In avian studies, with the exception of the study by Cook et al. (2005), studies of associative learning typically involve the acquisition of two associations (Anderson et al., 2017; Ashton et al., 2018; Boogert et al., 2011; Isden et al., 2013; Shaw et al., 2015). While underexplored, this type of associative learning task does not show a significant relationship with intelligence in humans, in children or adults (Plenderleith, 1956; Stevenson & Zigler, 1957). The associative learning tasks that are sensitive to differences in cognitive ability, the paired associates and three-term contingency tasks, have 10–30 unique stimulus pairs, placing more demands on learning, memory, and retrieval systems (Alexander & Smales, 1997; Kaufman et al., 2009; Tamez et al., 2008; Williams & Pearlberg, 2006). Differences in task design are expected when conducting compar-

ative studies to accommodate different physical and sensory capabilities, in addition to other factors like motivation (Macphail, 1987). As mentioned earlier, this means different species will need different parameters to ensure that performance is an adequate reflection of cognitive ability, but greater care should be taken to ensure that the underlying construct is the same (see Wright et al., 1985 for a beautiful demonstration of this using the comparative method of systematic variation; Bitterman, 1975).

Task purity is also a problem, with some tasks included in the batteries unduly influenced by personality, subject experience, experimental conditions, and physical health (Boesch, 2007; Kabadayi et al., 2018; Shaw et al., 2015; Sorato, Zidar, Garnham, Wilson, & Løvlie, 2018; van Horik et al., 2018). Finally, tests included in the battery should show high amounts of variability between subjects, but high reliability within subject. Deficits in either of these elements will hinder detection of a *g* factor (Carroll, 1993; Jensen, 1998). Some tasks in the PCTB have low levels of between-subjects variation, which may contribute to the difficulty in uncovering a general factor (Burkart et al., 2017; Herrmann et al., 2010). In the test batteries for sparrows and mice, low levels of reliability may have attenuated correlations and weakened the general factor found (Cauchoix et al., 2018; Fan, 2003; Galsworthy et al., 2005; Soha et al., 2019). Low reliabilities attenuating the subsequent correlations were often mentioned in these experiments, yet these correlations were not corrected to compensate for this issue. For these species, however, the reliability and variability may not be an issue entirely with the tasks, but with the subjects. Task reliability will be higher in populations with higher variance in their true scores, that is, their scores independent from random error. Populations with higher true variance could produce a stronger *g* factor because there is more variance available to be accounted for. While tasks still need to be carefully constructed to show between-subjects variability on the one hand, and within subject reliability on the other hand, potential differences in true variance across species should be kept in mind.

When a general factor has been found in nonhumans, the correlational matrix across task performance is not as robust as what we see in humans (Banerjee et al., 2009; Carroll, 1993; Galsworthy et al., 2002; Galsworthy et al., 2005; Herndon et al., 1997; Hopkins et al., 2014; Isden et al., 2013; Jensen, 1998; Kolata et al., 2005; Kolata et al., 2007; Kolata et al., 2008; Matzel et al., 2003; Matzel et al., 2006; Shaw et al., 2015; Woodley of Menie et al., 2015; but see Ashton et al., 2018). This is especially problematic when PCA is used to extract a *g* factor. PCA uses the total variance in the extracted components, even unique and error variance. This can result in overestimating the amount of variance the first extracted component can explain (Jensen & Weng, 1994). Some of the studies yielding poor correlational matrices used PCA, which may have overestimated general cognitive ability in animal studies (Galsworthy et al., 2002; Galsworthy et al., 2005; Isden et al., 2013; Matzel et al., 2003; Shaw et al., 2015; Table 1). This is not to dismiss the *g* factors that have emerged from weaker correlational matrices, but we need to understand why the correlations from nonhuman studies tend to be weaker. This could be because of low task reliability, as mentioned earlier. Sample size, however, is another factor impeding strong correlations, as most animal studies are underpowered (Banerjee et al., 2009; Galsworthy et al., 2002; Galsworthy et al., 2005; Herndon et al., 1997; Hopkins et al., 2014; Isden et al., 2013; Matzel et al., 2003; Matzel et al., 2006;

but see Kolata et al., 2008). For PCA and other methods of factor analysis, it is recommended that there should be at least 10 subjects for each measure, but few studies have achieved this ideal (Burkart et al., 2017; Costello & Osborne, 2005; Yong & Pearce, 2013). Some researchers have compensated for this by comparing empirical results to the results of a random bootstrapping procedure or randomly simulated data sets (Ashton et al., 2018; Damerius et al., 2019; Shaw et al., 2015), though this is not common practice. Another potential factor is the subject sample. Because the factors are extracted to explain variance in performance, the subject population must be heterogenous (Burkart et al., 2017; Yong & Pearce, 2013). Animal studies often lack heterogeneity, such as when studies of wild animals only test males (Isden et al., 2013), or bold individuals low in neophobia (Shaw & Schmelz, 2017). In lab environments, although outbred strains of mice are used, they are reared in nearly identical conditions; thereby, diminishing interindividual variance (Galsworthy et al., 2002; Kolata et al., 2005; Matzel et al., 2003). Environmental factors can make important contributions to cognitive abilities (Light et al., 2010; Neumann, Murphy, Gewa, Grillenberger, & Bwibo, 2007; Nisbett, 2009). Thus, the strength of strong environmental control of laboratory populations is also a weakness. A recent study showed that when mice were exposed to an enriched environment for 2 weeks, their performance on a cognitive test battery improved (Sauce et al., 2018). Factors like environmental conditions and population characteristics should be further explored to understand how they could be affecting performance on cognitive tasks.

Given the strong interest in general intelligence in humans, establishing methods for identifying a g factor across diverse species should be a top priority of comparative cognition research. The cognitive abilities of many species are starting to be formally recognized and tested, but understanding how those abilities are related to each other remains a mystery. Under what conditions will species show evidence for general cognitive ability versus distinct and nonoverlapping cognitive abilities? What are the costs and benefits of having a generalized versus specialized system? Social structure/group size, diet, and environmental complexity/variability have all been proposed as determinants of cognitive abilities (Ashton et al., 2018; Herrmann et al., 2007; Mettke-Hofmann, 2014; van Horik & Emery, 2011). Group size, for example, has been theorized to increase cognitive abilities because larger groups put more demands on learning about and remembering more individuals, including their status within the group, and interindividual interactions (van Horik & Emery, 2011). A weak correlational matrix was found in spotted bower birds (Isden et al., 2013), for example, while the correlational matrix found in the Australian magpies was stronger (Ashton et al., 2018). While these species show many behavioral similarities, including vocal imitation, sedentary lifestyle, and diet; they differ in their social interactions, breeding behaviors, and parenting. In bower birds, females select males based on bower attributes and mating display, but males do not assist with parenting (Isden et al., 2013). Additionally, there is evidence to suggest that interaction with conspecifics occurs primarily during breeding and mating in the form of competition, but less research has been published on bower bird behavior outside of bower activities (Madden, 2008). By contrast, Australian magpie groups involve complex social behaviors, where members help to provision nestlings that are not related to themselves (alloparenting), and work together to defend their

territory from predators and out-group members (Farabaugh, Brown, & Hughes, 1992; Finn & Hughes, 2001). The difference in social complexity might contribute to the different strengths of the correlational matrices in these two species. Group size can also explain differences in g within species as well. Within the Australian magpies, Ashton et al. (2018) found that cognitive performance improved as group size increased. This supports the idea that larger, more complex social groups are more cognitively challenging, thereby enhancing cognitive abilities of its members, but its potential explanatory value for magpies does not necessarily mean it will be able to explain differences across species. Enhanced cognition could be general (Ashton et al., 2018), or be restricted to social cognition. Herrmann et al.'s (2010) findings, and the low correlations between g and social ability in humans, suggest little effect of social complexity on g. Nevertheless, nuances within group size and social dynamics could help elucidate why these different results are found (Holekamp, 2007; Shultz & Dunbar, 2006).

Other researchers have argued that diet plays a substantial role in shaping cognitive abilities and brain function (DeCasien, Williams, & Higham, 2017; Holekamp, 2007; Mettke-Hofmann, 2014; but see Allen & Kay, 2012). Having a varied diet (e.g., omnivorous or frugivorous) is associated with larger brains and/or higher cognitive abilities compared with species with specialized diets (e.g., folivorous). This could be because of increased demands on learning and memory systems posed by an omnivorous diet, improved diet quality, or the combination of the two. Nevertheless, research on the role of diet on cognition usually focuses on a single cognitive ability, such as innovation, or uses brain size as a proxy for cognition, rather than measuring g (Chittka & Niven, 2009; Roth & Dicke, 2005; Snodgrass, Leonard, & Robertson, 2009; Sol, Sayol, Ducatez, & Lefebvre, 2016). There is also evidence that habitat complexity can influence brain size and rates of learning (Mettke-Hofmann, 2014; Sayol et al., 2016; Schuck-Paim, Alonso, & Ottoni, 2008). These influences are not necessarily mutually exclusive, and may interact in their contribution to natural behavior (Lefebvre & Sol, 2008; Mettke-Hofmann, 2014). It is possible that these influences will consistently covary. In cichlid fish, for example, environment complexity positively correlates with number of conspecifics (Pollen et al., 2007). A similar result was seen in African Starlings, where cooperative breeding is observed more frequently in complex environments (Rubenstein & Lovette, 2007). The potential for environment, diet, and social structure to covary makes it difficult to determine their independent contributions to brain size or cognitive abilities. Investigating a wider range of species could help answer this question. Noted by Holekamp (2007), spotted hyenas have high quality diets and complex social groups, whereas carnivorous and omnivorous bears also have high quality diets, but are predominately solitary. A better understanding of how diet, environment, and social structure impact specific cognitive abilities and brain size will also facilitate our understanding of how they relate to the underlying cognitive structure.

On the surface, it seems beneficial to have a larger brain and more advanced cognition. Larger brains are more diverse in function and structure (Roth & Dicke, 2005). Yet, brains are metabolically costly and so selection for increased brain size usually requires specific environmental conditions and tradeoffs with other metabolically expensive organ systems (Burkart et al., 2017; Byrne & Bates, 2007; Chittka & Niven, 2009; Isler & van Schaik,

2006; Iwaniuk & Nelson, 2003; Roth & Dicke, 2005). In humans, brain size increased as our digestive tracts shrank (Aiello & Wheeler, 1995). In birds, there is a negative correlation between brain size and pectoral muscle mass (Isler & van Schaik, 2006). Increasing brain size and cognitive ability is not the only solution to meet environmental challenges, however. In birds, there is a negative correlation between migratory distance and brain size, where the birds that traveled the furthest had the smallest brains (Sayol et al., 2016; Vincze, 2016). Yet, migratory birds show better long-term spatial memory compared with nonmigratory birds, indicating that despite having smaller brains specific cognitive abilities can be selected for (Mettke-Hofmann & Gwinner, 2003). Detailed comparative studies can illuminate the conditions that support selection for *general* cognitive abilities versus specific cognitive processes (Chittka & Niven, 2009; Mettke-Hofmann, 2014).

In a comparative analysis, better performance on cognitive tasks may not correlate with measures of fitness. In some species, fitness is increased through the selection of traits that attract mates or defeat rivals, yet with a concomitant decrease in brain size (Lefebvre & Sol, 2008). For other species, while potential mates that perform better on cognitive tasks are preferred (Chen, Zou, Sun, & Ten Cate, 2019; Spritzer, Meikle, & Solomon, 2005), this does not always result in increased fitness for those males or that females will act on that preference (Spritzer, Solomon, & Meikle, 2005).

Survival is another potential correlate with better cognitive performance in animals (Sol, Bacher, Reader, & Lefebvre, 2008; Sol, Székely, Liker, & Lefebvre, 2007), though this correlation is not always found (Kotrschal et al., 2015). There is evidence for a complicated interaction between cognitive abilities and personality that could result in equivalent rates of survival despite differences in cognitive abilities across individuals (Mazza, Eccard, Zaccaroni, Jacob, & Dammhahn, 2018; Mettke-Hofmann, 2014). In great tits (*Parus major*), a species of song bird, individuals who were more competitive in maximizing a particular food resource during winter performed more poorly on a problem-solving task compared with less competitive individuals (Cole & Quinn, 2012). Although intelligence is predictive of health and longevity in humans (Murray, Johnson, Wolf, & Deary, 2011), in modern societies this is more dependent on navigating environments that humans have created (especially schooling), not those created by the natural environment (Flinn, Geary, & Ward, 2005). No other animals have created, then subsequently had to resist, high-fat and high-sugar foods to prevent disease states. Understanding how cognitive abilities are related to survival in nonhumans will require the integration of multiple factors, including how cognitive abilities are interrelated. Species that show evidence for more interrelated cognitive abilities may have different interactions with noncognitive factors, like personality.

Another strength of animal research is that it could inform on different aspects that influence or are correlated with human intelligence. The benefit of using lab animals is they provide more control over biological factors which can be independently manipulated. The ability to closely monitor or manipulate brain function or genetics in animals can help elucidate which genes, brain regions, neural connections, and neurotransmitters are involved or correlated with cognitive functions, including *g* (Plomin, 2001; Matzel, Sauce, & Wass, 2013). Animal models have already identified some neurobiological correlates related to cognitive

ability, such as the importance of dopamine receptor function in mice (Wass et al., 2013; Wass et al., 2018), and cortical thickness and brain size in chimpanzees (Hopkins, Li, & Roberts, 2019). As animal test batteries improve, is it likely that more neurobiological correlates will not only be identified, but manipulated to help determine their causal influence on cognitive performance.

While animal and human research investigating the physical substrate of intelligence is important, there are some misconceptions about how deterministic these neurobiological correlates are. This is partially because some researchers have consistently stated that differences in intelligence are because of differences in inherited genes that are not sensitive to environmental factors (Jensen, 1998; Rushton & Jensen, 2005).[1] This argument is sometimes supported by heritability estimates that state intelligence is 60–80% heritable (Bailey, 1997; Gillborn, 2016). This theory of intelligence being determined by genes has persisted, yet the theory is continuously criticized because of how heritability estimates are calculated, and because more recent findings on the relationship between genes and intelligence fail to support these heritability estimates. Heritability estimates are used to determine how much variance in a characteristic can be attributed to genetic differences at the population level. Heritability estimates are influenced by how variable the environment is and are unique to populations at the time of estimate, meaning the same characteristic can have different estimates depending on who is sampled and when (Nisbett, 2009; Sauce & Matzel, 2018; Tucker-Drob & Bates, 2016; Turkheimer, Haley, Waldron, D'Onofrio, & Gottesman, 2003). These high heritability estimates for intelligence are also difficult to reconcile with current genetic research. For most traits, the number of genes involved in the expression is large, and the effect size of each individual contributing gene is minute on its own (Beauchamp et al., 2011; Chabris et al., 2012; Lango Allen et al., 2010). Furthermore, there is no evidence to suggest that these genes are insensitive to the environment (Bailey, 1997; Chabris et al., 2012; DeYoung & Clark, 2012). Some researchers argue that it is precisely a Gene × Environment interaction that could explain both high heritability estimates and low identification rates for specific gene variants (DeYoung & Clark, 2012; Sauce & Matzel, 2018). Heritability estimates typically over attribute variance in a trait to genes by including the Gene × Environment interaction in the estimate of heritability (Jensen, 1998; Sauce & Matzel, 2018). These results indicate that it is unlikely that differences in intelligence are because of immutable genetic factors. When neurobiological factors related to differences in cognitive performance are found they should not be presented as the sole and universal contributor to differences. Reductionist arguments like these could inadvertently perpetuate racist ideas (Gillborn, 2016; Phelan, Link, & Feldman, 2013). Instead, these findings should be presented in the context of environmental interactions.

Investigating how neurobiological correlates of intelligence are related to the environment is easier with animal research because of the amount of control a researcher has on the environment. As mentioned earlier, short-term interventions that provide environ-

---

[1] Many of these theories were created in attempt to explain differences in IQ scores between races. A discussion of race, IQ, and genes is outside the scope of this review, but please see Frank (2015); Krimsky and Sloan (2011), and Nisbett (2009) for discussion on why it is incorrect and harmful to posit race-based differences as innate.

mental enrichment improve performance on a cognitive test battery in mice (Sauce et al., 2018). Short term, intensive WM training increases dopamine (D1) receptor sensitivity and improves performance on a cognitive test battery in mice, highlighting the importance of even short-term interventions on biological substrates (Wass et al., 2013). Chronic environmental conditions and how that is related to cognitive performance could also be investigated. Animal research has already successfully modeled some of the environmental effects of development in a low socioeconomic status (SES) environment, including its neurobiological consequences (Hackman, Farah, & Meaney, 2010). SES correlates with intelligence, thus integrating these two lines of rodent research (environmental and genetic manipulations) could help uncover the causal direction of this correlation (Brooks-Gunn, Klebanov, & Duncan, 1996; Hackman et al., 2010; Jensen, 1998; Mani, Mullainathan, Shafir, & Zhao, 2013; Schmidt, 2017). Extended environmental manipulations will likely be key to understanding how chronic conditions impact cognitive function and the underlying neurobiological correlates.

Humans are a language-using species, and language enables much greater intelligence in our species than what is found even in other highly intelligent species (Penn, Holyoak, & Povinelli, 2008). Furthermore, verbal fluency correlates positively with FSIQ (Ardila, Pineda, & Rosselli, 2000). Thus, it is difficult to disentangle the contribution of language to g. By studying nonlanguage animal models, we can gain insight into the cognitive processes and capacities that contribute to g that do not require, or that are independent of language (Shaw & Schmelz, 2017; Figure 2).

Well-developed test batteries for use in different animals, including humans, can help validate the neuroscience of g and its related cognitive mechanisms. Finding a general cognitive factor in animals has so far been only partially successful. Correcting methodological issues discussed in the previous sections will improve the search for a g factor in other species. Test batteries across all species, including humans, could be modified to facilitate comparative research. Tasks that have been used with both humans and other species that have not been included in test batteries are ideal targets for development. As discussed earlier, assessments of WM in rodent test batteries typically use an eight-arm radial maze. Humans have been tested on a virtual radial arm maze, but this has not been incorporated into a larger battery or compared with more traditional measures of WM (Astur et al., 2004; Shaw & Schmelz, 2017). Reversal learning is another example of a cognitive task that is commonly included in animal test batteries, and is commonly used in humans to investigate neuropsychiatric disorders (Izquierdo, Brigman, Radke, Rudebeck, & Holmes, 2017); yet, is underexplored in humans in relation to g. Furthermore, nonhuman animals should receive tasks that more closely resemble those used to study g in humans. For example, pigeons have shown similar RT effects on a variation of a human task based on Hick's Law (Vickrey & Neuringer, 2000). Including tasks like this in a test battery for animals would allow for increased correspondence between human and nonhuman animal measures of g.

Test batteries should also include more tasks where animals have to use previously acquired knowledge to solve novel problems (van Horik & Lea, 2017). Understanding how to apply knowledge beyond the trained situation is thought to explain why g is one of the best predictors of job performance (Schmidt, 2014).

In the test batteries given to animals, there is a debate about how ecologically relevant those tasks should be (Burkart et al., 2017; Herrmann et al., 2007). Nevertheless, if the goal is to discover *general* cognitive abilities, then it is not clear how important it is that the tasks in the test battery are ecologically relevant. The more ecologically relevant a task is, the more likely that they will engage highly conserved behavioral processes (those often labeled as "instinctive"), with little interindividual variation (Burkart et al., 2017). Using contrived and standardized tasks, such as in an operant chamber, can actually help control for noncognitive factors, like environmental experience, and facilitate comparisons across species (Clarin, Ruczyński, Page, & Siemers, 2013, but see Shaw, 2017).

Perhaps the most important factor is that test batteries should assess clear and separable domains of cognition as much as possible (Burkart et al., 2017; Shaw & Schmelz, 2017). Many studies, particularly those investigating cognition in the wild, use ill-defined tasks such as problem solving or innovation. This can make it difficult to determine what aspects of cognition are being used to solve the task, whether the same strategy is engaged across subjects, and if the behaviors are related to other cognitive abilities. Ultimately, there should be more communication across labs to determine that test batteries for different species attempt to assess the same underlying constructs, but which constructs should receive the most focus? As reviewed earlier, in humans, WM, processing speed, and associative learning have shown a relationship to g, though the causal nature of this relationship is still debated. These basic cognitive processes have been found in just about all vertebrate orders, ranging from birds and mammals to amphibians and fish. Furthermore, these core cognitive processes reflect basic functions of vertebrate brains, often involving collaboration across multiple circuits, such as hippocampus, frontal cortex, and basal ganglia (Papini, 2008). By focusing on these core cognitive processes, it is reasonable and possible to create a comprehensive cross species test battery that could extract psychometric g should it be present.

If psychometric g is found in a broad range of taxa, the causal factor may reflect a deep homology of the vertebrate brain despite species-specific brain and cognitive specializations (Güntürkün & Bugnyar, 2016; Osvath, Kabadayi, & Jacobs, 2014). This hypothesis requires testing, but such testing in turn requires the development of a test battery that can reliably assess these core cognitive functions across diverse species of vertebrate, from the human to pigeon to fish. Despite our suggestion that g should be assessed with the common set of general core cognitive processes of the vertebrate brain, this does not reject the idea that there are species-specific cognitive specializations found in individual species or groups of species. As an analogy, the five-digit hand is a deep homology found in all tetrapods, and reflects the ancestral state. As a result, there are some common core functions of the five-digit hand. There has also been selection for specialization in hand structure and function, such as the opposable thumb of humans that allows for fine motor precision, and even more extreme specializations for specific forms of locomotion, such as the wings of bats, the fins of whales, and the hooves of horses—each reflecting an adaptive specialization to each species' particular locomotor niche. Nevertheless, independent of these specializations, interindividual variation in hand function within a species should be readily measurable using batteries of functional tasks, such as grip

strength, dexterity, and precision, or locomotor functions. Likewise, as we discussed above, some birds that store seeds to be retrieved weeks or months later show specialized adaptation of spatial memory and the supporting brain systems, in particular the hippocampus. There is likely a complex relationship between specialized cognitive abilities and *g* because of differences in ecological constraints across species. Nevertheless, there ought to be interindividual variation in spatial memory in a species of food-storing birds, just as there are within a species of nonfood storing birds, despite the fact that the food-storing species has an overall greater spatial memory than does the nonstoring species. Appropriate tests that assess general cognitive functions are needed to facilitate assessments of *g* across a diverse array of species. Thus, assessments should be focused on the general cognitive processes, such as WM and associative learning, that are found in all vertebrates.

It is inarguable that one factor explaining half of the variance in performance on cognitive tests has been identified in humans (Lubinski, 2004). This factor is a good predictor of mortality, health, level of education, and SES. Furthermore, it is clear that this factor is most strongly related to WM and processing speed (see Figure 1). What this factor consists of and what underlies its function is still under intense investigation. Better measures of a general factor in humans and animals could be an important effective tool to shed new light on general intelligence. Only then can we more clearly elucidate the evolutionary and environmental contributors to a general cognitive ability.

References

Aben, B., Stapert, S., & Blokland, A. (2012). About the distinction between working memory and short-term memory. *Frontiers in Psychology, 3,* 301. http://dx.doi.org/10.3389/fpsyg.2012.00301

Ackerman, P. L., & Cianciolo, A. T. (2000). Cognitive, perceptual-speed, and psychomotor determinants of individual differences during skill acquisition. *Journal of Experimental Psychology: Applied, 6,* 259–290. http://dx.doi.org/10.1037/1076-898X.6.4.259

Adolphs, R. (1999). Social cognition and the human brain. *Trends in Cognitive Sciences, 3,* 469–479. http://dx.doi.org/10.1016/S1364-6613(99)01399-6

Aiello, L. C., & Wheeler, P. (1995). The expensive-tissue hypothesis: The brain and the digestive system in human and primate evolution. *Current Anthropology, 36,* 199–221. http://dx.doi.org/10.1086/204350

Alexander, J. R. M., & Smales, S. (1997). Intelligence, learning and long-term memory. *Personality and Individual Differences, 23,* 815–825. http://dx.doi.org/10.1016/S0191-8869(97)00054-8

Allen, K. L., & Kay, R. F. (2012). Dietary quality and encephalization in platyrrhine primates. *Proceedings of the Royal Society B: Biological Sciences, 279,* 715–721.

Anderson, B. (1993). Evidence from the rat for a general factor that underlies cognitive performance and that relates to brain size: Intelligence? *Neuroscience Letters, 153,* 98–102. http://dx.doi.org/10.1016/0304-3940(93)90086-Z

Anderson, R. C., Searcy, W. A., Peters, S., Hughes, M., DuBois, A. L., & Nowicki, S. (2017). Song learning and cognitive ability are not consistently related in a songbird. *Animal Cognition, 20,* 309–320. http://dx.doi.org/10.1007/s10071-016-1053-7

Ardila, A., Pineda, D., & Rosselli, M. (2000). Correlation between intelligence test scores and executive function measures. *Archives of Clinical Neuropsychology, 15,* 31–36. http://dx.doi.org/10.1093/arclin/15.1.31

Ashton, B. J., Ridley, A. R., Edwards, E. K., & Thornton, A. (2018). Cognitive performance is linked to group size and affects fitness in Australian magpies. *Nature, 554,* 364–367. http://dx.doi.org/10.1038/nature25503

Ashton, M. C., & Lee, K. (2005). Problems with the method of correlated vectors. *Intelligence, 33,* 431–444. http://dx.doi.org/10.1016/j.intell.2004.12.004

Astur, R. S., Tropp, J., Sava, S., Constable, R. T., & Markus, E. J. (2004). Sex differences and correlations in a virtual Morris water task, a virtual radial arm maze, and mental rotation. *Behavioural Brain Research, 151,* 103–115. http://dx.doi.org/10.1016/j.bbr.2003.08.024

Au, J., Sheehan, E., Tsai, N., Duncan, G. J., Buschkuehl, M., & Jaeggi, S. M. (2015). Improving fluid intelligence with training on working memory: A meta-analysis. *Psychonomic Bulletin & Review, 22,* 366–377. http://dx.doi.org/10.3758/s13423-014-0699-x

Auersperg, A. M., von Bayern, A. M., Gajdon, G. K., Huber, L., & Kacelnik, A. (2011). Flexibility in problem solving and tool use of kea and New Caledonian crows in a multi access box paradigm. *PLoS ONE, 6,* e20231. http://dx.doi.org/10.1371/journal.pone.0020231

Baddeley, A. (2003). Working memory: Looking back and looking forward. *Nature Reviews Neuroscience, 4,* 829–839. http://dx.doi.org/10.1038/nrn1201

Bailey, R. C. (1997). Hereditarian scientific fallacies. *Genetica, 99,* 125–133. http://dx.doi.org/10.1007/BF02259516

Baker, D. P., Eslinger, P. J., Benavides, M., Peters, E., Dieckmann, N. F., & Leon, J. (2015). The cognitive impact of the education revolution: A possible cause of the Flynn Effect on population IQ. *Intelligence, 49,* 144–158. http://dx.doi.org/10.1016/j.intell.2015.01.003

Balda, R. P., & Kamil, A. C. (1992). Long-term spatial memory in Clark's nutcracker. *Nucifraga columbiana Animal Behaviour, 44,* 761–769. http://dx.doi.org/10.1016/S0003-3472(05)80302-1

Banerjee, K., Chabris, C. F., Johnson, V. E., Lee, J. J., Tsao, F., & Hauser, M. D. (2009). General intelligence in another primate: Individual differences across cognitive task performance in a New World monkey (*Saguinus oedipus*). *PLoS ONE, 4,* e5883. http://dx.doi.org/10.1371/journal.pone.0005883

Bartholomew, D. J., Deary, I. J., & Lawn, M. (2009). A new lease of life for Thomson's bonds model of intelligence. *Psychological Review, 116,* 567–579. http://dx.doi.org/10.1037/a0016262

Bayliss, D. M., Jarrold, C., Gunn, D. M., & Baddeley, A. D. (2003). The complexities of complex span: Explaining individual differences in working memory in children and adults. *Journal of Experimental Psychology: General, 132,* 71–92. http://dx.doi.org/10.1037/0096-3445.132.1.71

Beauchamp, J. P., Cesarini, D., Johannesson, M., van der Loos, M. J., Koellinger, P. D., Groenen, P. J., . . . Christakis, N. A. (2011). Molecular genetics and economics. *The Journal of Economic Perspectives, 25,* 57–82. http://dx.doi.org/10.1257/jep.25.4.57

Benson, N., Hulac, D. M., & Kranzler, J. H. (2010). Independent examination of the Wechsler Adult Intelligence Scale-Fourth Edition (WAIS-IV): What does the WAIS-IV measure? *Psychological Assessment, 22,* 121–130. http://dx.doi.org/10.1037/a0017767

Bickley, P. G., Keith, T. Z., & Wolfle, L. M. (1995). The three-stratum theory of cognitive abilities: Test of the structure of intelligence across the life span. *Intelligence, 20,* 309–328. http://dx.doi.org/10.1016/0160-2896(95)90013-6

Bitterman, M. E. (1965). The evolution of intelligence. *Scientific American, 212,* 92–100. http://dx.doi.org/10.1038/scientificamerican0165-92

Bitterman, M. E. (1975). The comparative analysis of learning. *Science, 188,* 699–709. http://dx.doi.org/10.1126/science.188.4189.699

Blaga, O. M., Shaddy, D. J., Anderson, C. J., Kannass, K. N., Little, T. D., & Colombo, J. (2009). Structure and continuity of intellectual development in early childhood. *Intelligence, 37,* 106–113. http://dx.doi.org/10.1016/j.intell.2008.09.003

Blair, C. (2006). How similar are fluid cognition and general intelligence? A developmental neuroscience perspective on fluid cognition as an

aspect of human cognitive ability. *Behavioral and Brain Sciences, 29,* 109–125. http://dx.doi.org/10.1017/S0140525X06009034

Blaisdell, A. P., & Cook, R. G. (2005). Two-item same-different concept learning in pigeons. *Animal Learning & Behavior, 33,* 67–77. http://dx .doi.org/10.3758/BF03196051

Boesch, C. (2007). What makes us human (*Homo sapiens*)? The challenge of cognitive cross-species comparison. *Journal of Comparative Psychology, 121,* 227–240. http://dx.doi.org/10.1037/0735-7036.121.3.227

Boogert, N. J., Anderson, R. C., Peters, S., Searcy, W. A., & Nowicki, S. (2011). Song repertoire size in male song sparrows correlates with detour reaching, but not with other cognitive measures. *Animal Behaviour, 81,* 1209–1216. http://dx.doi.org/10.1016/j.anbehav.2011.03.004

Bornstein, M. H., Hahn, C. S., Bell, C., Haynes, O. M., Slater, A., Golding, J., & Wolke, D., & the ALSPAC Study Team. (2006). Stability in cognition across early childhood. A developmental cascade. *Psychological Science, 17,* 151–158. http://dx.doi.org/10.1111/j.1467-9280.2006 .01678.x

Brodnick, R. J., & Ree, M. J. (1995). A structural model of academic performance, socioeconomic status, and Spearman's g. *Educational and Psychological Measurement, 55,* 583–594. http://dx.doi.org/10.1177/ 0013164495055004006

Brooks-Gunn, J., Klebanov, P. K., & Duncan, G. J. (1996). Ethnic differences in children's intelligence test scores: Role of economic deprivation, home environment, and maternal characteristics. *Child Development, 67,* 396–408. http://dx.doi.org/10.2307/1131822

Brown, K. G., Le, H., & Schmidt, F. L. (2006). Specific aptitude theory revisited: Is there incremental validity for training performance? *International Journal of Selection and Assessment, 14,* 87–100. http://dx.doi .org/10.1111/j.1468-2389.2006.00336.x

Burkart, J. M., Schubiger, M. N., & van Schaik, C. P. (2017). The evolution of general intelligence. *Behavioral and Brain Sciences, 40,* e195.

Byrne, R. W., & Bates, L. A. (2007). Sociality, evolution and cognition. *Current Biology, 17,* R714–R723. http://dx.doi.org/10.1016/j.cub.2007 .05.069

Cahan, S., & Noyman, A. (2001). The Kaufman ability battery for children mental processing scale: A valid measure of "pure" intelligence? *Educational and Psychological Measurement, 61,* 827–840. http://dx.doi .org/10.1177/00131640121971545

Carpenter, P. A., Just, M. A., & Shell, P. (1990). What one intelligence test measures: A theoretical account of the processing in the Raven Progressive Matrices Test. *Psychological Review, 97,* 404–431. http://dx.doi .org/10.1037/0033-295X.97.3.404

Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies.* New York, NY: Cambridge University Press. http://dx.doi.org/ 10.1017/CBO9780511571312

Carroll, J. B. (2003). The higher-stratum structure of cognitive abilities: Current evidence supports g and about ten broad factors. In H. Nyborg, (Ed.), *The Scientific Study of General Intelligence* (pp. 5–22). New York, NY: Pergamon Press. http://dx.doi.org/10.1016/B978-008043793-4/50036-2

Cattell, R. B. (1987). *Intelligence: Its structure, growth and action.* New York, NY: Elsevier.

Cattell, R. B., & Horn, J. L. (1978). A check on the theory of fluid and crystallized intelligence with description of new subtest designs. *Journal of Educational Measurement, 15,* 139–164. http://dx.doi.org/10.1111/j .1745-3984.1978.tb00065.x

Cauchoix, M., Chow, P. K. Y., van Horik, J. O., Atance, C. M., Barbeau, E. J., Barragan-Jason, G., . . . Cauchard, L. (2018). The repeatability of cognitive performance: A meta-analysis. *Philosophical Transactions of the Royal Society B: Biological Sciences, 373,* 20170281. http://dx.doi .org/10.1098/rstb.2017.0281

Cavojová, V., Mikusková, E. B., & Hanák, R. (2013). Do you have to be smart to know what the others are thinking? *Studia Psychologica, 55,* 61–66. http://dx.doi.org/10.21909/sp.2013.01.621

Ceci, S. J. (1991). How much does schooling influence general intelligence and its cognitive components? A reassessment of the evidence. *Developmental Psychology, 27,* 703–722. http://dx.doi.org/10.1037/0012-1649.27.5.703

Ceci, S. J. (1996). *On intelligence: A biological treatise on intellectual development.* Cambridge, MA: Harvard University Press.

Chabris, C. F. (2007). Cognitive and neurobiological mechanisms of the Law of General Intelligence. In M. J. Roberts (Ed.), *Integrating the mind: Domain specific versus domain general processes in higher cognition* (pp. 449–491). Hove: Psychology Press.

Chabris, C. F., Hebert, B. M., Benjamin, D. J., Beauchamp, J., Cesarini, D., van der Loos, M., . . . Laibson, D. (2012). Most reported genetic associations with general intelligence are probably false positives. *Psychological Science, 23,* 1314–1323. http://dx.doi.org/10.1177/ 0956797611435528

Chen, J., Zou, Y., Sun, Y. H., & Ten Cate, C. (2019). Problem-solving males become more attractive to female budgerigars. *Science, 363,* 166–167. http://dx.doi.org/10.1126/science.aau8181

Chittka, L., & Niven, J. (2009). Are bigger brains better? *Current Biology, 19,* R995–R1008. http://dx.doi.org/10.1016/j.cub.2009.08.023

Chooi, W. T., & Thompson, L. A. (2012). Working memory training does not improve intelligence in healthy young adults. *Intelligence, 40,* 531–542. http://dx.doi.org/10.1016/j.intell.2012.07.004

Chuderski, A., Taraday, M., Nęcka, E., & Smoleń, T. (2012). Storage capacity explains fluid intelligence but executive control does not. *Intelligence, 40,* 278–295. http://dx.doi.org/10.1016/j.intell.2012.02.010

Clarin, T. M., Ruczyński, I., Page, R. A., & Siemers, B. M. (2013). Foraging ecology predicts learning performance in insectivorous bats. *PLoS ONE, 8,* e64823. http://dx.doi.org/10.1371/journal.pone.0064823

Cole, E. F., & Quinn, J. L. (2012). Personality and problem-solving performance explain competitive ability in the wild. *Proceedings of the Royal Society B: Biological Sciences, 279,* 1168–1175.

Colom, R., Abad, F. J., García, L. F., & Juan-Espinosa, M. (2002). Education, Wechsler's full scale IQ, and g. *Intelligence, 30,* 449–462. http://dx.doi.org/10.1016/S0160-2896(02)00122-8

Colom, R., Abad, F. J., Quiroga, M. Á., Shih, P. C., & Flores-Mendoza, C. (2008). Working memory and intelligence are highly related constructs, but why? *Intelligence, 36,* 584–606. http://dx.doi.org/10.1016/j.intell .2008.01.002

Colom, R., Román, F. J., Abad, F. J., Shih, P. C., Privado, J., Froufe, M., . . . Jaeggi, S. M. (2013). Adaptive n-back training does not improve fluid intelligence at the construct level: Gains on individual tests suggest that training may enhance visuospatial processing. *Intelligence, 41,* 712–727. http://dx.doi.org/10.1016/j.intell.2013.09.002

Colombo, M., & Broadbent, N. (2000). Is the avian hippocampus a functional homologue of the mammalian hippocampus? *Neuroscience and Biobehavioral Reviews, 24,* 465–484. http://dx.doi.org/10.1016/S0149-7634(00)00016-6

Colombo, M., & Scarf, D. (2012). Neurophysiological studies of learning and memory in pigeons. *Comparative Cognition & Behavior Reviews, 7,* 23–43. http://dx.doi.org/10.3819/ccbr.2012.70002

Conway, A. R., Cowan, N., Bunting, M. F., Therriault, D. J., & Minkoff, S. R. (2002). A latent variable analysis of working memory capacity, short-term memory capacity, processing speed, and general fluid intelligence. *Intelligence, 30,* 163–183. http://dx.doi.org/10.1016/S0160-2896(01)00096-4

Conway, A. R., Kane, M. J., Bunting, M. F., Hambrick, D. Z., Wilhelm, O., & Engle, R. W. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin & Review, 12,* 769–786. http://dx.doi.org/10.3758/BF03196772

Conway, A. R., & Kovacs, K. (2015). New and emerging models of human intelligence. *WIRE: Cognitive Science, 6,* 419–426. http://dx.doi.org/10.1002/wcs.1356

Cook, R. G., & Blaisdell, A. P. (2006). Short-term item memory in successive same-different discriminations. *Behavioural Processes, 72,* 255–264. http://dx.doi.org/10.1016/j.beproc.2006.03.013

Cook, R. G., Katz, J. S., & Blaisdell, A. P. (2012). Temporal properties of visual search in pigeon target localization. *Journal of Experimental Psychology: Animal Behavior Processes, 38,* 209–216. http://dx.doi.org/10.1037/a0026496

Cook, R. G., Levison, D. G., Gillett, S. R., & Blaisdell, A. P. (2005). Capacity and limits of associative memory in pigeons. *Psychonomic Bulletin & Review, 12,* 350–358. http://dx.doi.org/10.3758/BF03196384

Costello, A. B., & Osborne, J. (2005). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical Assessment, Research & Evaluation, 10,* 7.

Cowan, N. (2008). What are the differences between long-term, short-term, and working memory? *Progress in Brain Research, 169,* 323–338. http://dx.doi.org/10.1016/S0079-6123(07)00020-9

Coyle, T. R., Pillow, D. R., Snyder, A. C., & Kochunov, P. (2011). Processing speed mediates the development of general intelligence (g) in adolescence. *Psychological Science, 22,* 1265–1269. http://dx.doi.org/10.1177/0956797611418243

Damerius, L. A., Burkart, J. M., van Noordwijk, M. A., Haun, D. B., Kosonen, Z. K., Galdikas, B. M., . . . van Schaik, C. P. (2019). General cognitive abilities in orangutans (*Pongo abelii* and *Pongo pygmaeus*). *Intelligence, 74,* 3–11. http://dx.doi.org/10.1016/j.intell.2018.10.008

Deary, I. J. (2000). *Looking down on human intelligence: From psychometrics to the brain.* New York, NY: Oxford University Press. http://dx.doi.org/10.1093/acprof:oso/9780198524175.001.0001

Deary, I. J., Der, G., & Ford, G. (2001). Reaction times and intelligence differences: A population-based cohort study. *Intelligence, 29,* 389–399.

Deary, I. J., Pattie, A., & Starr, J. M. (2013). The stability of intelligence from age 11 to age 90 years: The Lothian birth cohort of 1921. *Psychological Science, 24,* 2361–2368. http://dx.doi.org/10.1177/0956797613486487

Deary, I. J., Whiteman, M. C., Starr, J. M., Whalley, L. J., & Fox, H. C. (2004). The impact of childhood intelligence on later life: Following up the Scottish mental surveys of 1932 and 1947. *Journal of Personality and Social Psychology, 86,* 130–147. http://dx.doi.org/10.1037/0022-3514.86.1.130

DeCasien, A. R., Williams, S. A., & Higham, J. P. (2017). Primate brain size is predicted by diet but not sociality. *Nature Ecology & Evolution, 1,* 112. http://dx.doi.org/10.1038/s41559-017-0112

Demetriou, A., Makris, N., Spanoudis, G., Kazi, S., Shayer, M., & Kazali, E. (2018). Mapping the dimensions of general intelligence: An integrated differential-developmental theory. *Human Development, 61,* 4–42. http://dx.doi.org/10.1159/000484450

Dempster, F. N. (1991). Inhibitory processes: A neglected dimension of intelligence. *Intelligence, 15,* 157–173. http://dx.doi.org/10.1016/0160-2896(91)90028-C

Derksen, J., Kramer, I., & Katzko, M. (2002). Does a self-report measure for emotional intelligence assess something different than general intelligence? *Personality and Individual Differences, 32,* 37–48. http://dx.doi.org/10.1016/S0191-8869(01)00004-6

DeYoung, C. G., & Clark, R. (2012). The gene in its natural habitat: The importance of gene-trait interactions. *Development and Psychopathology, 24,* 1307–1318. http://dx.doi.org/10.1017/S0954579412000727

Divac, I., Mogensen, J., & Björklund, A. (1985). The prefrontal 'cortex' in the pigeon. Biochemical evidence. *Brain Research, 332,* 365–368. http://dx.doi.org/10.1016/0006-8993(85)90606-7

Doebler, P., & Scheffler, B. (2016). The relationship of choice reaction time variability and intelligence: A meta-analysis. *Learning and Individual Differences, 52,* 157–166. http://dx.doi.org/10.1016/j.lindif.2015.02.009

Duncan, J., Seitz, R. J., Kolodny, J., Bor, D., Herzog, H., Ahmed, A., . . . Emslie, H. (2000). A neural basis for general intelligence. *Science, 289,* 457–460. http://dx.doi.org/10.1126/science.289.5478.457

Eagle, D. M., Bari, A., & Robbins, T. W. (2008). The neuropsychopharmacology of action inhibition: Cross-species translation of the stop-signal and go/no-go tasks. *Psychopharmacology, 199,* 439–456. http://dx.doi.org/10.1007/s00213-008-1127-6

Engle, R. W., Tuholski, S. W., Laughlin, J. E., & Conway, A. R. A. (1999). Working memory, short-term memory, and general fluid intelligence: A latent-variable approach. *Journal of Experimental Psychology: General, 128,* 309–331. http://dx.doi.org/10.1037/0096-3445.128.3.309

Estrada, E., Ferrer, E., Abad, F. J., Román, F. J., & Colom, R. (2015). A general factor of intelligence fails to account for changes in tests' scores after cognitive practice: A longitudinal multi-group latent-variable study. *Intelligence, 50,* 93–99. http://dx.doi.org/10.1016/j.intell.2015.02.004

Fagan, J. F., Holland, C. R., & Wheeler, K. (2007). The prediction, from infancy, of adult IQ and achievement. *Intelligence, 35,* 225–231. http://dx.doi.org/10.1016/j.intell.2006.07.007

Fan, X. (2003). Two approaches for correcting correlation attenuation caused by measurement error: Implications for research practice. *Educational and Psychological Measurement, 63,* 915–930. http://dx.doi.org/10.1177/0013164403251319

Farabaugh, S. M., Brown, E. D., & Hughes, J. M. (1992). Cooperative territorial defense in the Australian Magpie, Gymnorhina tibicen (Passeriformes, Cracticidae), a group-living songbird. *Ethology, 92,* 283–292. http://dx.doi.org/10.1111/j.1439-0310.1992.tb00966.x

Finn, P. G., & Hughes, J. M. (2001). Helping behaviour in Australian magpies. *Gymnorhina tibicen. Emu - Austral Ornithology, 101,* 57–63. http://dx.doi.org/10.1071/MU00066

Flaim, M., & Blaisdell, A. P. (2020). *A procedure for the assessment of individual differences in relational discrimination behavior in the Pigeon (Columba livia).* Unpublished manuscript, Department is Psychology, University of California, Los Angeles, CA.

Flinn, M. V., Geary, D. C., & Ward, C. V. (2005). Ecological dominance, social competition, and coalitionary arms races: Why humans evolved extraordinary intelligence. *Evolution and Human Behavior, 26,* 10–46. http://dx.doi.org/10.1016/j.evolhumbehav.2004.08.005

Frank, R. (2015). Back to the future? The emergence of a geneticized conceptualization of race in sociology. *The Annals of the American Academy of Political and Social Science, 661,* 51–64. http://dx.doi.org/10.1177/0002716215590775

Friedman, N. P., Miyake, A., Corley, R. P., Young, S. E., Defries, J. C., & Hewitt, J. K. (2006). Not all executive functions are related to intelligence. *Psychological Science, 17,* 172–179. http://dx.doi.org/10.1111/j.1467-9280.2006.01681.x

Galsworthy, M. J., Paya-Cano, J. L., Liu, L., Monleón, S., Gregoryan, G., Fernandes, C., . . . Plomin, R. (2005). Assessing reliability, heritability and general cognitive ability in a battery of cognitive tasks for laboratory mice. *Behavior Genetics, 35,* 675–692. http://dx.doi.org/10.1007/s10519-005-3423-9

Galsworthy, M. J., Paya-Cano, J. L., Monleón, S., & Plomin, R. (2002). Evidence for general cognitive ability (g) in heterogeneous stock mice and an analysis of potential confounds. *Genes, Brain and Behavior, 1,* 88–95. http://dx.doi.org/10.1034/j.1601-183X.2002.10204.x

Garlick, D. (2002). Understanding the nature of the general factor of intelligence: The role of individual differences in neural plasticity as an explanatory mechanism. *Psychological Review, 109,* 116–136. http://dx.doi.org/10.1037/0033-295X.109.1.116

Garlick, D., Fountain, S. B., & Blaisdell, A. P. (2017). Serial pattern learning in pigeons: Rule-based or associative? *Journal of Experimental*

Psychology: Animal Learning and Cognition, 43, 30–47. http://dx.doi .org/10.1037/xan0000109

Gignac, G. E. (2014). Dynamic mutualism versus g factor theory: An empirical test. *Intelligence, 42,* 89–97. http://dx.doi.org/10.1016/j.intell .2013.11.004

Gignac, G. E. (2015). Raven's is not a pure measure of general intelligence: Implications for g factor theory and the brief measurement of g. *Intelligence, 52,* 71–79. http://dx.doi.org/10.1016/j.intell.2015.07.006

Gillborn, D. (2016). Softly, softly: Genetics, intelligence and the hidden racism of the new geneism. *Journal of Education Policy, 31,* 365–388.

Gottfredson, L. S. (2002). Where and why g matters: Not a mystery. *Human Performance, 15,* 25–46. http://dx.doi.org/10.1080/08959285 .2002.9668082

Gottfredson, L. S., & Deary, I. J. (2004). Intelligence predicts health and longevity, but why? *Current Directions in Psychological Science, 13,* 1–4. http://dx.doi.org/10.1111/j.0963-7214.2004.01301001.x

Gow, A. J., Johnson, W., Pattie, A., Brett, C. E., Roberts, B., Starr, J. M., & Deary, I. J. (2011). Stability and change in intelligence from age 11 to ages 70, 79, and 87: The Lothian Birth Cohorts of 1921 and 1936. *Psychology and Aging, 26,* 232–240. http://dx.doi.org/10.1037/ a0021072

Güntürkün, O. (2005). The avian 'prefrontal cortex' and cognition. *Current Opinion in Neurobiology, 15,* 686–693. http://dx.doi.org/10.1016/j.conb .2005.10.003

Güntürkün, O., & Bugnyar, T. (2016). Cognition without Cortex. *Trends in Cognitive Sciences, 20,* 291–303. http://dx.doi.org/10.1016/j.tics.2016 .02.001

Gustafsson, J. E. (1984). A unifying model for the structure of intellectual abilities. *Intelligence, 8,* 179–203. http://dx.doi.org/10.1016/0160-2896(84)90008-4

Gustafsson, J. E. (2003). Measurement from a hierarchical point of view. In H. I. Braun, D. N. Jackson, & D. E. Wiley (Eds.), *The role of constructs in psychological and educational measurement.* Mahwah, NJ: L. Erlbaum Associates.

Gutman, L. M., & Schoon, I. (2013). The impact of non-cognitive skills on outcomes for young people. *Education Endowment Foundation, 59,* 4–57.

Hackman, D. A., Farah, M. J., & Meaney, M. J. (2010). Socioeconomic status and the brain: Mechanistic insights from human and animal research. *Nature Reviews Neuroscience, 11,* 651–659. http://dx.doi.org/ 10.1038/nrn2897

Hakstian, A. R., & Cattell, R. B. (1978). Higher-stratum ability structures on a basis of twenty primary abilities. *Journal of Educational Psychology, 70,* 657–669. http://dx.doi.org/10.1037/0022-0663.70.5.657

Harrison, T. L., Shipstead, Z., Hicks, K. L., Hambrick, D. Z., Redick, T. S., & Engle, R. W. (2013). Working memory training may increase working memory capacity but not fluid intelligence. *Psychological Science, 24,* 2409–2419. http://dx.doi.org/10.1177/0956797613492984

Heckman, J., Pinto, R., & Savelyev, P. (2013). Understanding the mechanisms through which an influential early childhood program boosted adult outcomes. *The American Economic Review, 103,* 2052–2086. http://dx.doi.org/10.1257/aer.103.6.2052

Herbranson, W. T., & Shimp, C. P. (2008). Artificial grammar learning in pigeons. *Learning & Behavior, 36,* 116–137. http://dx.doi.org/10.3758/ LB.36.2.116

Herman, L. M. (2010). What laboratory research has told us about dolphin cognition. *International Journal of Comparative Psychology, 23,* 310–330.

Herndon, J. G., Moss, M. B., Rosene, D. L., & Killiany, R. J. (1997). Patterns of cognitive decline in aged rhesus monkeys. *Behavioural Brain Research, 87,* 25–34. http://dx.doi.org/10.1016/S0166-4328(96)02256-5

Herrmann, E., Call, J., Hernàndez-Lloreda, M. V., Hare, B., & Tomasello, M. (2007). Humans have evolved specialized skills of social cognition:

The cultural intelligence hypothesis. *Science, 317,* 1360–1366. http://dx .doi.org/10.1126/science.1146282

Herrmann, E., Hernández-Lloreda, M. V., Call, J., Hare, B., & Tomasello, M. (2010). The structure of individual differences in the cognitive abilities of children and chimpanzees. *Psychological Science, 21,* 102–110. http://dx.doi.org/10.1177/0956797609356511

Hick, W. E. (1952). On the rate of gain of information. *Quarterly Journal of Experimental Psychology, 4,* 11–26. http://dx.doi.org/10.1080/ 17470215208416600

Holekamp, K. E. (2007). Questioning the social intelligence hypothesis. *Trends in Cognitive Sciences, 11,* 65–69. http://dx.doi.org/10.1016/j.tics .2006.11.003

Hopkins, W. D., Li, X., & Roberts, N. (2019). More intelligent chimpanzees (*Pan troglodytes*) have larger brains and increased cortical thickness. *Intelligence, 74,* 18–24. http://dx.doi.org/10.1016/j.intell.2018.11 .002

Hopkins, W. D., Russell, J. L., & Schaeffer, J. (2014). Chimpanzee intelligence is heritable. *Current Biology, 24,* 1649–1652. http://dx.doi .org/10.1016/j.cub.2014.05.076

Horn, N. R., Dolan, M., Elliott, R., Deakin, J. F. W., & Woodruff, P. W. R. (2003). Response inhibition and impulsivity: An fMRI study. *Neuropsychologia, 41,* 1959–1966. http://dx.doi.org/10.1016/S0028-3932 (03)00077-0

Hoyer, W. J., Stawski, R. S., Wasylyshyn, C., & Verhaeghen, P. (2004). Adult age and digit symbol substitution performance: A meta-analysis. *Psychology and Aging, 19,* 211–214. http://dx.doi.org/10.1037/0882-7974.19.1.211

Isden, J., Panayi, C., Dingle, C., & Madden, J. (2013). Performance in cognitive and problem-solving tasks in male spotted bowerbirds does not correlate with mating success. *Animal Behaviour, 86,* 829–838. http:// dx.doi.org/10.1016/j.anbehav.2013.07.024

Isler, K., & van Schaik, C. (2006). Costs of encephalization: The energy trade-off hypothesis tested on birds. *Journal of Human Evolution, 51,* 228–243. http://dx.doi.org/10.1016/j.jhevol.2006.03.006

Iwaniuk, A. N., & Nelson, J. E. (2003). Developmental differences are correlated with relative brain size in birds: A comparative analysis. *Canadian Journal of Zoology, 81,* 1913–1928. http://dx.doi.org/10.1139/ z03-190

Izquierdo, A., Brigman, J. L., Radke, A. K., Rudebeck, P. H., & Holmes, A. (2017). The neural basis of reversal learning: An updated perspective. *Neuroscience, 345,* 12–26. http://dx.doi.org/10.1016/j.neuroscience .2016.03.021

Izquierdo, A., & Jentsch, J. D. (2012). Reversal learning as a measure of impulsive and compulsive behavior in addictions. *Psychopharmacology, 219,* 607–620. http://dx.doi.org/10.1007/s00213-011-2579-7

Jaeggi, S. M., Buschkuehl, M., Jonides, J., & Perrig, W. J. (2008). Improving fluid intelligence with training on working memory. *Proceedings of the National Academy of Sciences of the United States of America, 105,* 6829–6833. http://dx.doi.org/10.1073/pnas.0801268105

Jarrold, C., Tam, H., Baddeley, A. D., & Harvey, C. E. (2011). How does processing affect storage in working memory tasks? Evidence for both domain-general and domain-specific effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37,* 688–705. http://dx .doi.org/10.1037/a0022527

Jensen, A. R. (1982). Reaction time and psychometric g. In H. J. Eysenck (Ed.), *A model for intelligence* (pp. 93–132). Berlin, Heidelberg: Springer.

Jensen, A. R. (1998). *The g factor: The science of mental ability.* Westport, CT: Praeger.

Jensen, A. R., & Munro, E. (1979). Reaction time, movement time, and intelligence. *Intelligence, 3,* 121–126.

Jensen, A. R., & Weng, L. J. (1994). What is a good g? *Intelligence, 18,* 231–258. http://dx.doi.org/10.1016/0160-2896(94)90029-9

Johnson, W., & Bouchard, T. J., Jr. (2005). The structure of human intelligence: It is verbal, perceptual, and image rotation (VPR), not fluid and crystallized. *Intelligence, 33,* 393–416. http://dx.doi.org/10.1016/j.intell.2004.12.002

Johnson, W., Bouchard, T. J., Jr., Krueger, R. F., McGue, M., & Gottesman, I. I. (2004). Just one g: Consistent results from three test batteries. *Intelligence, 32,* 95–107. http://dx.doi.org/10.1016/S0160-2896(03)00062-X

Kabadayi, C., Bobrowicz, K., & Osvath, M. (2018). The detour paradigm in animal cognition. *Animal Cognition, 21,* 21–35. http://dx.doi.org/10.1007/s10071-017-1152-0

Kan, K. J., Kievit, R. A., Dolan, C., & van der Maas, H. (2011). On the interpretation of the CHC factor Gc. *Intelligence, 39,* 292–302. http://dx.doi.org/10.1016/j.intell.2011.05.003

Kan, K. J., Wicherts, J. M., Dolan, C. V., & van der Maas, H. L. (2013). On the nature and nurture of intelligence and specific cognitive abilities: The more heritable, the more culture dependent. *Psychological Science, 24,* 2420–2428. http://dx.doi.org/10.1177/0956797613493292

Kane, M. J., Conway, A. R. A., Miura, T. K., & Colflesh, G. J. H. (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. http://dx.doi.org/10.1037/0278-7393.33.3.615

Kangas, B. D., Berry, M. S., & Branch, M. N. (2011). On the development and mechanics of delayed matching-to-sample performance. *Journal of the Experimental Analysis of Behavior, 95,* 221–236. http://dx.doi.org/10.1901/jeab.2011.95-221

Katz, J. S., & Wright, A. A. (2006). Same/different abstract-concept learning by pigeons. *Journal of Experimental Psychology: Animal Behavior Processes, 32,* 80–86. http://dx.doi.org/10.1037/0097-7403.32.1.80

Kaufman, A. B., Reynolds, M. R., & Kaufman, A. S. (2019). The structure of ape (*hominoidea*) intelligence. *Journal of Comparative Psychology, 133,* 92–105. http://dx.doi.org/10.1037/com0000136

Kaufman, S. B., DeYoung, C. G., Gray, J. R., Brown, J., & Mackintosh, N. (2009). Associative learning predicts intelligence above and beyond working memory and processing speed. *Intelligence, 37,* 374–382. http://dx.doi.org/10.1016/j.intell.2009.03.004

Keagy, J., Savard, J. F., & Borgia, G. (2011). Complex relationship between multiple measures of cognitive ability and male mating success in satin bowerbirds, *Ptilonorhynchus violaceus Animal Behaviour, 81,* 1063–1070. http://dx.doi.org/10.1016/j.anbehav.2011.02.018

Keith, T. Z., & Reynolds, M. R. (2010). Cattell–Horn–Carroll abilities and cognitive tests: What we've learned from 20 years of research. *Psychology in the Schools, 47,* 635–650. http://dx.doi.org/10.1002/pits.20496

Kent, P. (2017). Fluid intelligence: A brief history. *Applied Neuropsychology: Child, 6,* 193–203. http://dx.doi.org/10.1080/21622965.2017.1317480

Klein, C., Rauh, R., & Biscaldi, M. (2010). Cognitive correlates of antisaccade task performance. *Experimental Brain Research, 203,* 759–764. http://dx.doi.org/10.1007/s00221-010-2276-5

Kolata, S., Light, K., Grossman, H. C., Hale, G., & Matzel, L. D. (2007). Selective attention is a primary determinant of the relationship between working memory and general learning ability in outbred mice. *Learning & Memory, 14,* 22–28. http://dx.doi.org/10.1101/lm.408507

Kolata, S., Light, K., & Matzel, L. D. (2008). Domain-specific and domain-general learning factors are expressed in genetically heterogeneous CD-1 mice. *Intelligence, 36,* 619–629. http://dx.doi.org/10.1016/j.intell.2007.12.001

Kolata, S., Light, K., Townsend, D. A., Hale, G., Grossman, H. C., & Matzel, L. D. (2005). Variations in working memory capacity predict individual differences in general learning abilities among genetically diverse mice. *Neurobiology of Learning and Memory, 84,* 241–246. http://dx.doi.org/10.1016/j.nlm.2005.07.006

Kotrschal, A., Buechel, S. D., Zala, S. M., Corral-Lopez, A., Penn, D. J., & Kolm, N. (2015). Brain size affects female but not male survival under predation threat. *Ecology Letters, 18,* 646–652. http://dx.doi.org/10.1111/ele.12441

Kovacs, K., & Conway, A. R. (2016). Process overlap theory: A unified account of the general factor of intelligence. *Psychological Inquiry, 27,* 151–177. http://dx.doi.org/10.1080/1047840X.2016.1153946

Krasheninnikova, A., Berardi, R., Lind, M. A., O'Neill, L., & von Bayern, A. M. (2019). Primate cognition test battery in parrots. *Behaviour, 156,* 721–761. http://dx.doi.org/10.1163/1568539X-0003549

Krimsky, S., & Sloan, K. (Eds.). (2011). *Race and the genetic revolution: Science, myth, and culture.* New York, NY: Columbia University Press.

Kvist, A. V., & Gustafsson, J. E. (2008). The relation between fluid intelligence and the general factor as a function of cultural background: A test of Cattell's Investment theory. *Intelligence, 36,* 422–436. http://dx.doi.org/10.1016/j.intell.2007.08.004

Lango Allen, H., Estrada, K., Lettre, G., Berndt, S. I., Weedon, M. N., Rivadeneira, F., . . . Hirschhorn, J. N. (2010). Hundreds of variants clustered in genomic loci and biological pathways affect human height. *Nature, 467,* 832–838. http://dx.doi.org/10.1038/nature09410

Lefebvre, L., & Sol, D. (2008). Brains, lifestyles and cognition: Are there general trends? *Brain, Behavior and Evolution, 72,* 135–144. http://dx.doi.org/10.1159/000151473

Light, K. R., Kolata, S., Wass, C., Denman-Brice, A., Zagalsky, R., & Matzel, L. D. (2010). Working memory training promotes general cognitive abilities in genetically heterogeneous mice. *Current Biology, 20,* 777–782. http://dx.doi.org/10.1016/j.cub.2010.02.034

Lind, J., Enquist, M., & Ghirlanda, S. (2015). Animal memory: A review of delayed matching-to-sample data. *Behavioural Processes, 117,* 52–58. http://dx.doi.org/10.1016/j.beproc.2014.11.019

Locurto, C., Fortin, E., & Sullivan, R. (2003). The structure of individual differences in heterogeneous stock mice across problem types and motivational systems. *Genes, Brain & Behavior, 2,* 40–55. http://dx.doi.org/10.1034/j.1601-183X.2003.00006.x

Locurto, C., & Scanlon, C. (1998). Individual differences and a spatial learning factor in two strains of mice (*Mus musculus*). *Journal of Comparative Psychology, 112,* 344–352. http://dx.doi.org/10.1037/0735-7036.112.4.344

Locurto, C., Benoit, A., Crowley, C., & Miele, A. (2006). The structure of individual differences in batteries of rapid acquisition tasks in mice. *Journal of Comparative Psychology, 120,* 378–388. http://dx.doi.org/10.1037/0735-7036.120.4.378

Lubinski, D. (2004). Introduction to the special section on cognitive abilities: 100 years after Spearman's (1904). "'General intelligence,' objectively determined and measured". *Journal of Personality and Social Psychology, 86,* 96–111. http://dx.doi.org/10.1037/0022-3514.86.1.96

Macphail, E. M. (1987). The comparative psychology of intelligence. *Behavioral and Brain Sciences, 10,* 645–656. http://dx.doi.org/10.1017/S0140525X00054984

Madden, J. R. (2008). Do bowerbirds exhibit cultures? *Animal Cognition, 11,* 1–12. http://dx.doi.org/10.1007/s10071-007-0092-5

Mani, A., Mullainathan, S., Shafir, E., & Zhao, J. (2013). Poverty impedes cognitive function. *Science, 341,* 976–980. http://dx.doi.org/10.1126/science.1238041

Martínez, K., Burgaleta, M., Román, F. J., Escorial, S., Shih, P. C., Quiroga, M. Á., & Colom, R. (2011). Can fluid intelligence be reduced to 'simple' short-term storage? *Intelligence, 39,* 473–480. http://dx.doi.org/10.1016/j.intell.2011.09.001

Matzel, L. D., Han, Y. R., Grossman, H., Karnik, M. S., Patel, D., Scott, N., . . . Gandhi, C. C. (2003). Individual differences in the expression of a "general" learning ability in mice. *The Journal of Neuroscience, 23,* 6423–6433. http://dx.doi.org/10.1523/JNEUROSCI.23-16-06423.2003

Matzel, L. D., Sauce, B., & Wass, C. (2013). The architecture of intelligence: Converging evidence from studies of humans and animals. *Current Directions in Psychological Science, 22,* 342–348. http://dx.doi.org/10.1177/0963721413491764

Matzel, L. D., Townsend, D. A., Grossman, H., Han, Y. R., Hale, G., Zappulla, M., . . . Kolata, S. (2006). Exploration in outbred mice covaries with general learning abilities irrespective of stress reactivity, emotionality, and physical attributes. *Neurobiology of Learning and Memory, 86,* 228–240. http://dx.doi.org/10.1016/j.nlm.2006.03.004

Mazza, V., Eccard, J. A., Zaccaroni, M., Jacob, J., & Dammhahn, M. (2018). The fast and the flexible: Cognitive style drives individual variation in cognition in a small mammal. *Animal Behaviour, 137,* 119–132. http://dx.doi.org/10.1016/j.anbehav.2018.01.011

Mettke-Hofmann, C. (2014). Cognitive ecology: Ecological factors, lifestyles, and cognition. *WIREs: Cognitive Science, 5,* 345–360. http://dx.doi.org/10.1002/wcs.1289

Mettke-Hofmann, C., & Gwinner, E. (2003). Long-term memory for a life on the move. *Proceedings of the National Academy of Sciences of the United States of America, 100,* 5863–5866. http://dx.doi.org/10.1073/pnas.1037505100

Murray, C., Johnson, W., Wolf, M. S., & Deary, I. J. (2011). The association between cognitive ability across the lifespan and health literacy in old age: The Lothian Birth Cohort 1936. *Intelligence, 39,* 178–187. http://dx.doi.org/10.1016/j.intell.2011.04.001

Neumann, C. G., Murphy, S. P., Gewa, C., Grillenberger, M., & Bwibo, N. O. (2007). Meat supplementation improves growth, cognitive, and behavioral outcomes in Kenyan children. *the Journal of Nutrition, 137,* 1119–1123.

Nisbett, R. E. (2009). *Intelligence and how to get it: Why schools and cultures count*. New York, NY: WW Norton & Company.

Oberauer, K., Süß, H. M., Schulze, R., Wilhelm, O., & Wittmann, W. W. (2000). Working memory capacity—Facets of a cognitive ability construct. *Personality and Individual Differences, 29,* 1017–1045. http://dx.doi.org/10.1016/S0191-8869(99)00251-2

Osvath, M., Kabadayi, C., & Jacobs, I. (2014). Independent evolution of similar complex cognitive skills: The importance of embodied degrees of freedom. *Animal Behavior and Cognition, 1,* 249–264. http://dx.doi.org/10.12966/abc.08.03.2014

Papini, M. R. (2008). *Comparative psychology: Evolution and development of behavior*. New York, NY: Psychology Press.

Penn, D. C., Holyoak, K. J., & Povinelli, D. J. (2008). Darwin's mistake: Explaining the discontinuity between human and nonhuman minds. *Behavioral and Brain Sciences, 31,* 109–130. http://dx.doi.org/10.1017/S0140525X08003543

Pepperberg, I. M. (2018). Alex the parrot. In J. Vonk & T. Shackelford (Eds.), *Encyclopedia of animal cognition and behavior* (pp. 1–10). Cham: Springer. http://dx.doi.org/10.1007/978-3-319-47829-6_80-1

Phelan, J. C., Link, B. G., & Feldman, N. M. (2013). The genomic revolution and beliefs about essential racial differences: A backdoor to eugenics? *American Sociological Review, 78,* 167–191. http://dx.doi.org/10.1177/0003122413476034

Pietschnig, J., & Voracek, M. (2015). One century of global IQ gains: A formal meta-analysis of the Flynn effect (1909–2013). *Perspectives on Psychological Science, 10,* 282–306. http://dx.doi.org/10.1177/1745691615577701

Plenderleith, M. (1956). Discrimination learning and discrimination reversal learning in normal and feebleminded children. *The Journal of Genetic Psychology, 88,* 107–112. http://dx.doi.org/10.1080/00221325.1956.10532958

Plomin, R. (2001). The genetics of g in human and mouse. *Nature Reviews Neuroscience, 2,* 136–141. http://dx.doi.org/10.1038/35053584

Plomin, R., & von Stumm, S. (2018). The new genetics of intelligence. *Nature Reviews Genetics, 19,* 148–159. http://dx.doi.org/10.1038/nrg.2017.104

Polderman, T. J. C., de Geus, E. J. C., Hoekstra, R. A., Bartels, M., van Leeuwen, M., Verhulst, F. C., . . . Boomsma, D. I. (2009). Attention problems, inhibitory control, and intelligence index overlapping genetic factors: A study in 9-, 12-, and 18-year-old twins. *Neuropsychology, 23,* 381–391. http://dx.doi.org/10.1037/a0014915

Pollen, A. A., Dobberfuhl, A. P., Scace, J., Igulu, M. M., Renn, S. C., Shumway, C. A., & Hofmann, H. A. (2007). Environmental complexity and social organization sculpt the brain in Lake Tanganyikan cichlid fish. *Brain, Behavior and Evolution, 70,* 21–39. http://dx.doi.org/10.1159/000101067

Postlethwaite, B. E. (2011). *Fluid ability, crystallized ability, and performance across multiple domains: A meta-analysis* (Unpublished doctoral dissertation). University of Iowa, Iowa City, Iowa.

Rajkumar, A. P., Yovan, S., Raveendran, A. L., & Russell, P. S. S. (2008). Can only intelligent children do mind reading: The relationship between intelligence and theory of mind in 8 to 11 years old. *Behavioral and Brain Functions, 4,* 51. http://dx.doi.org/10.1186/1744-9081-4-51

Raven, J. C. (1941). Standardization of progressive matrices, 1938. *British Journal of Medical Psychology, 19,* 137–150. http://dx.doi.org/10.1111/j.2044-8341.1941.tb00316.x

Raven, J. (2008). The Raven progressive matrices tests: Their theoretical basis and measurement model. In J. Raven & J. Raven (Eds.), *Uses and abuses of intelligence. Studies advancing Spearman and Raven's quest for non-arbitrary metrics* (pp. 17–68). Unionville, NY: Royal Fireworks Press.

Redick, T. S., Shipstead, Z., Harrison, T. L., Hicks, K. L., Fried, D. E., Hambrick, D. Z., . . . Engle, R. W. (2013). No evidence of intelligence improvement after working memory training: A randomized, placebo-controlled study. *Journal of Experimental Psychology: General, 142,* 359–379. http://dx.doi.org/10.1037/a0029082

Ree, M. J., & Earles, J. A. (1992). Intelligence is the best predictor of job performance. *Current Directions in Psychological Science, 1,* 86–89. http://dx.doi.org/10.1111/1467-8721.ep10768746

Reynolds, M. R., Floyd, R. G., & Niileksela, C. R. (2013). How well is psychometric g indexed by global composites? Evidence from three popular intelligence tests. *Psychological Assessment, 25,* 1314–1321. http://dx.doi.org/10.1037/a0034102

Reynolds, M. R., & Turek, J. J. (2012). A dynamic developmental link between verbal comprehension-knowledge (Gc) and reading comprehension: Verbal comprehension-knowledge drives positive change in reading comprehension. *Journal of School Psychology, 50,* 841–863. http://dx.doi.org/10.1016/j.jsp.2012.07.002

Rolfhus, E. L., & Ackerman, P. L. (1999). Assessing individual differences in knowledge: Knowledge, intelligence, and related traits. *Journal of Educational Psychology, 91,* 511–526. http://dx.doi.org/10.1037/0022-0663.91.3.511

Rose, S. A., Feldman, J. F., Jankowski, J. J., & Van Rossem, R. (2008). A cognitive cascade in infancy: Pathways from prematurity to later mental development. *Intelligence, 36,* 367–378. http://dx.doi.org/10.1016/j.intell.2007.07.003

Roth, G., & Dicke, U. (2005). Evolution of the brain and intelligence. *Trends in Cognitive Sciences, 9,* 250–257. http://dx.doi.org/10.1016/j.tics.2005.03.005

Rubenstein, D. R., & Lovette, I. J. (2007). Temporal environmental variability drives the evolution of cooperative breeding in birds. *Current Biology, 17,* 1414–1419. http://dx.doi.org/10.1016/j.cub.2007.07.032

Rushton, J. P., & Jensen, A. R. (2005). Thirty years of research on race differences in cognitive ability. *Psychology, Public Policy, and Law, 11,* 235–294. http://dx.doi.org/10.1037/1076-8971.11.2.235

Sauce, B., Bendrath, S., Herzfeld, M., Siegel, D., Style, C., Rab, S., . . . Matzel, L. D. (2018). The impact of environmental interventions among mouse siblings on the heritability and malleability of general cognitive ability. *Philosophical Transactions of the Royal Society B: Biological Sciences, 373,* 20170289.

Sauce, B., & Matzel, L. D. (2018). The paradox of intelligence: Heritability and malleability coexist in hidden gene-environment interplay. *Psychological Bulletin, 144,* 26–47. http://dx.doi.org/10.1037/bul0000131

Sauce, B., Wass, C., Smith, A., Kwan, S., & Matzel, L. D. (2014). The external-internal loop of interference: Two types of attention and their influence on the learning abilities of mice. *Neurobiology of Learning and Memory, 116,* 181–192. http://dx.doi.org/10.1016/j.nlm.2014.10.005

Sayol, F., Maspons, J., Lapiedra, O., Iwaniuk, A. N., Székely, T., & Sol, D. (2016). Environmental variation and the evolution of large brains in birds. *Nature Communications, 7,* 13971. http://dx.doi.org/10.1038/ncomms13971

Schipolowski, S., Wilhelm, O., & Schroeders, U. (2014). On the nature of crystallized intelligence: The relationship between verbal ability and factual knowledge. *Intelligence, 46,* 156–168. http://dx.doi.org/10.1016/j.intell.2014.05.014

Schmidt, F. L. (2011). A theory of sex differences in technical aptitude and some supporting evidence. *Perspectives on Psychological Science, 6,* 560–573. http://dx.doi.org/10.1177/1745691611419670

Schmidt, F. L. (2014). A general theoretical integrative model of individual differences in interests, abilities, and personality traits, and academic and occupational achievement. *Perspectives on Psychological Science, 9,* 211–218. http://dx.doi.org/10.1177/1745691613518074

Schmidt, F. L. (2017). Beyond questionable research methods: The role of omitted relevant research in the credibility of research. *Archives of Scientific Psychology, 5,* 32–41. http://dx.doi.org/10.1037/arc0000033

Schmiedek, F., Lövdén, M., & Lindenberger, U. (2010). Hundred days of cognitive training enhance broad cognitive abilities in adulthood: Findings from the COGITO study. *Frontiers in Aging Neuroscience, 2,* 27. http://dx.doi.org/10.3389/fnagi.2010.00027

Schmiedek, F., Lövdén, M., & Lindenberger, U. (2014). A task is a task is a task: Putting complex span, n-back, and other working memory indicators in psychometric context. *Frontiers in Psychology, 5,* 1475. http://dx.doi.org/10.3389/fpsyg.2014.01475

Schmitt, J. E., Raznahan, A., Liu, S., & Neale, M. C. (2020). The genetics of cortical myelination in young adults and its relationships to cerebral surface area, cortical thickness, and intelligence: A magnetic resonance imaging study of twins and families. *NeuroImage, 206,* 116319. http://dx.doi.org/10.1016/j.neuroimage.2019.116319

Schrank, F. A., & McGrew, K. S. (2001). *Woodcock-Johnson® III.* Itasca, IL: Riverside.

Schuck-Paim, C., Alonso, W. J., & Ottoni, E. B. (2008). Cognition in an ever-changing world: Climatic variability is associated with brain size in Neotropical parrots. *Brain, Behavior and Evolution, 71,* 200–215. http://dx.doi.org/10.1159/000119710

Schweizer, K. (2007). Investigating the relationship of working memory tasks and fluid intelligence tests by means of the fixed-links model in considering the impurity problem. *Intelligence, 35,* 591–604. http://dx.doi.org/10.1016/j.intell.2006.11.004

Serpell, R. (2000). Intelligence and culture. In R. J. Sternberg (Ed.), *Handbook of intelligence* (pp. 549–578). New York, NY: Cambridge University Press. http://dx.doi.org/10.1017/CBO9780511807947.026

Shah, P., & Miyake, A. (1996). The separability of working memory resources for spatial thinking and language processing: An individual differences approach. *Journal of Experimental Psychology: General, 125,* 4–27. http://dx.doi.org/10.1037/0096-3445.125.1.4

Shaw, R. C. (2017). Testing cognition in the wild: Factors affecting performance and individual consistency in two measures of avian cognition. *Behavioural Processes, 134,* 31–36. http://dx.doi.org/10.1016/j.beproc.2016.06.004

Shaw, R. C., Boogert, N. J., Clayton, N. S., & Burns, K. C. (2015). Wild psychometrics: Evidence for 'general' cognitive performance in wild New Zealand robins (*Petroica longipes*). *Animal Behaviour, 109,* 101–111. http://dx.doi.org/10.1016/j.anbehav.2015.08.001

Shaw, R. C., & Schmelz, M. (2017). Cognitive test batteries in animal cognition research: Evaluating the past, present and future of comparative psychometrics. *Animal Cognition, 20,* 1003–1018. http://dx.doi.org/10.1007/s10071-017-1135-1

Shelton, J. T., Elliott, E. M., Matthews, R. A., Hill, B. D., & Gouvier, W. D. (2010). The relationships of working memory, secondary memory, and general fluid intelligence: Working memory is special. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36,* 813–820. http://dx.doi.org/10.1037/a0019046

Sheppard, L. D., & Vernon, P. A. (2008). Intelligence and speed of information-processing: A review of 50 years of research. *Personality and Individual Differences, 44,* 535–551. http://dx.doi.org/10.1016/j.paid.2007.09.015

Shultz, S., & Dunbar, R. I. M. (2006). Both social and ecological factors predict ungulate brain size. *Proceedings Biological Sciences, 273,* 207–215. http://dx.doi.org/10.1098/rspb.2005.3283

Snodgrass, J. J., Leonard, W. R., & Robertson, M. L. (2009). The energetics of encephalization in early hominids. In J. J. Hublin & M. P. Richards (Eds.), *The evolution of hominin diets* (pp. 15–29). Dordrecht, the Netherlands: Springer. http://dx.doi.org/10.1007/978-1-4020-9699-0_2

Soha, J. A., Peters, S., Anderson, R. C., Searcy, W. A., & Nowicki, S. (2019). Performance on tests of cognitive ability is not repeatable across years in a songbird. *Animal Behaviour, 158,* 281–288. http://dx.doi.org/10.1016/j.anbehav.2019.09.020

Sol, D., Bacher, S., Reader, S. M., & Lefebvre, L. (2008). Brain size predicts the success of mammal species introduced into novel environments. *American Naturalist, 172*(Suppl. 1), S63–S71. http://dx.doi.org/10.1086/588304

Sol, D., Sayol, F., Ducatez, S., & Lefebvre, L. (2016). The life-history basis of behavioural innovations. *Philosophical Transactions of the Royal Society B: Biological Sciences, 371,* 20150187.

Sol, D., Székely, T., Liker, A., & Lefebvre, L. (2007). Big-brained birds survive better in nature. *Proceedings Biological Sciences, 274,* 763–769. http://dx.doi.org/10.1098/rspb.2006.3765

Sorato, E., Zidar, J., Garnham, L., Wilson, A., & Løvlie, H. (2018). Heritabilities and co-variation among cognitive traits in red junglefowl. *Philosophical Transactions of the Royal Society B: Biological Sciences, 373,* 20170285.

Spearman, C. (1904). General Intelligence," objectively determined and measured. *The American Journal of Psychology, 15,* 201–292. http://dx.doi.org/10.2307/1412107

Spinath, F. M., Ronald, A., Harlaar, N., Price, T. S., & Plomin, R. (2003). Phenotypic g early in life: On the etiology of general cognitive ability in a large population sample of twin children aged 2–4 years. *Intelligence, 31,* 195–210. http://dx.doi.org/10.1016/S0160-2896(02)00110-1

Spritzer, M. D., Meikle, D. B., & Solomon, N. G. (2005). Female choice based on male spatial ability and aggressiveness among meadow voles. *Animal Behaviour, 69,* 1121–1130. http://dx.doi.org/10.1016/j.anbehav.2004.06.033

Spritzer, M. D., Solomon, N. G., & Meikle, D. B. (2005). Influence of scramble competition for mates upon the spatial ability of male meadow voles. *Animal Behaviour, 69,* 375–386. http://dx.doi.org/10.1016/j.anbehav.2004.03.015

Stankov, L., & Roberts, R. D. (1997). Mental speed is not the 'basic' process of intelligence. *Personality and Individual Differences, 22,* 69–84. http://dx.doi.org/10.1016/S0191-8869(96)00163-8

St Clair-Thompson, H. L. (2010). Backwards digit recall: A measure of short-term memory or working memory? *European Journal of Cognitive Psychology, 22,* 286–296. http://dx.doi.org/10.1080/09541440902771299

Stephenson, C. L., & Halpern, D. F. (2013). Improved matrix reasoning is limited to training on tasks with a visuospatial component. *Intelligence, 41,* 341–357. http://dx.doi.org/10.1016/j.intell.2013.05.006

Stevenson, H. W., & Zigler, E. F. (1957). Discrimination learning and rigidity in normal and feebleminded individuals. *Journal of Personality, 25,* 699–711. http://dx.doi.org/10.1111/j.1467-6494.1957.tb01558.x

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology, 18,* 643–662. http://dx.doi.org/10.1037/h0054651

Swick, D., Ashley, V., & Turken, U. (2011). Are the neural correlates of stopping and not going identical? Quantitative meta-analysis of two response inhibition tasks. *NeuroImage, 56,* 1655–1665. http://dx.doi.org/10.1016/j.neuroimage.2011.02.070

Tamez, E., Myerson, J., & Hale, S. (2008). Learning, working memory, and intelligence revisited. *Behavioural Processes, 78,* 240–245. http://dx.doi.org/10.1016/j.beproc.2008.01.008

Thorsen, C., Gustafsson, J. E., & Cliffordson, C. (2014). The influence of fluid and crystallized intelligence on the development of knowledge and skills. *British Journal of Educational Psychology, 84,* 556–570. http://dx.doi.org/10.1111/bjep.12041

Tucker-Drob, E. M., & Bates, T. C. (2016). Large cross-national differences in gene × socioeconomic status interaction on intelligence. *Psychological Science, 27,* 138–149. http://dx.doi.org/10.1177/0956797615612727

Turkheimer, E., Haley, A., Waldron, M., D'Onofrio, B., & Gottesman, I. I. (2003). Socioeconomic status modifies heritability of IQ in young children. *Psychological Science, 14,* 623–628. http://dx.doi.org/10.1046/j.0956-7976.2003.psci_1475.x

Unsworth, N., & Engle, R. W. (2007). On the division of short-term and working memory: An examination of simple and complex span and their relation to higher order abilities. *Psychological Bulletin, 133,* 1038–1066. http://dx.doi.org/10.1037/0033-2909.133.6.1038

Unsworth, N., Redick, T. S., Heitz, R. P., Broadway, J. M., & Engle, R. W. (2009). Complex working memory span tasks and higher-order cognition: A latent-variable analysis of the relationship between processing and storage. *Memory, 17,* 635–654. http://dx.doi.org/10.1080/09658210902998047

Van der Maas, H. L., Dolan, C. V., Grasman, R. P., Wicherts, J. M., Huizenga, H. M., & Raijmakers, M. E. (2006). A dynamical model of general intelligence: The positive manifold of intelligence by mutualism. *Psychological Review, 113,* 842–861. http://dx.doi.org/10.1037/0033-295X.113.4.842

van Horik, J., & Emery, N. J. (2011). Evolution of cognition. *WIREs: Cognitive Science, 2,* 621–633. http://dx.doi.org/10.1002/wcs.144

van Horik, J. O., Langley, E. J., Whiteside, M. A., Laker, P. R., Beardsworth, C. E., & Madden, J. R. (2018). Do detour tasks provide accurate assays of inhibitory control? *Proceedings of the Royal Society B: Biological Sciences, 285,* 20180150.

van Horik, J. O., & Lea, S. E. G. (2017). Disentangling learning from knowing: Does associative learning ability underlie performances on cognitive test batteries? *Behavioral and Brain Sciences, 40,* e220. http://dx.doi.org/10.1017/S0140525X16001795

Verbruggen, F., Best, M., Bowditch, W. A., Stevens, T., & McLaren, I. P. (2014). The inhibitory control reflex. *Neuropsychologia, 65,* 263–278. http://dx.doi.org/10.1016/j.neuropsychologia.2014.08.014

Vernon, P. A. (1983). Speed of information processing and general intelligence. *Intelligence, 7,* 53–70. http://dx.doi.org/10.1016/0160-2896(83)90006-5

Vickrey, C., & Neuringer, A. (2000). Pigeon reaction time, Hick's law, and intelligence. *Psychonomic Bulletin & Review, 7,* 284–291. http://dx.doi.org/10.3758/BF03212983

Vincze, O. (2016). Light enough to travel or wise enough to stay? Brain size evolution and migratory behavior in birds. *Evolution, 70,* 2123–2133. http://dx.doi.org/10.1111/evo.13012

Viswesvaran, C., Ones, D. S., Schmidt, F. L., Le, H., & Oh, I. S. (2014). Measurement error obfuscates scientific knowledge: Path to cumulative knowledge requires corrections for unreliability and psychometric meta-analyses. *Industrial and Organizational Psychology: Perspectives on Science and Practice, 7,* 507–518. http://dx.doi.org/10.1017/S1754942600006799

Warne, R. T., & Burningham, C. (2019). Spearman's g found in 31 non-Western nations: Strong evidence that g is a universal phenomenon. *Psychological Bulletin, 145,* 237–272. http://dx.doi.org/10.1037/bul0000184

Wass, C., Pizzo, A., Sauce, B., Kawasumi, Y., Sturzoiu, T., Ree, F., . . . Matzel, L. D. (2013). Dopamine D1 sensitivity in the prefrontal cortex predicts general cognitive abilities and is modulated by working memory training. *Learning & Memory, 20,* 617–627. http://dx.doi.org/10.1101/lm.031971.113

Wass, C., Sauce, B., Pizzo, A., & Matzel, L. D. (2018). Dopamine D1 receptor density in the mPFC responds to cognitive demands and receptor turnover contributes to general cognitive ability in mice. *Scientific Reports, 8,* 4533. http://dx.doi.org/10.1038/s41598-018-22668-0

Weiss, L. G., Keith, T. Z., Zhu, J., & Chen, H. (2013). WAIS-IV and clinical validation of the four-and five-factor interpretative approaches. *Journal of Psychoeducational Assessment, 31,* 94–113. http://dx.doi.org/10.1177/0734282913478030

Westfall, J., & Yarkoni, T. (2016). Statistically controlling for confounding constructs is harder than you think. *PLoS ONE, 11,* e0152719. http://dx.doi.org/10.1371/journal.pone.0152719

Wicherts, J. M. (2017). Psychometric problems with the method of correlated vectors applied to item scores (including some nonsensical results). *Intelligence, 60,* 26–38. http://dx.doi.org/10.1016/j.intell.2016.11.002

Wicherts, J. M., Dolan, C. V., & van der Maas, H. L. (2010). A systematic literature review of the average IQ of sub-Saharan Africans. *Intelligence, 38,* 1–20. http://dx.doi.org/10.1016/j.intell.2009.05.002

Wilhelm, O., Hildebrandt, A., & Oberauer, K. (2013). What is working memory capacity, and how can we measure it? *Frontiers in Psychology, 4,* 433. http://dx.doi.org/10.3389/fpsyg.2013.00433

Williams, B. A., & Pearlberg, S. L. (2006). Learning of three-term contingencies correlates with Raven scores, but not with measures of cognitive processing. *Intelligence, 34,* 177–191. http://dx.doi.org/10.1016/j.intell.2005.03.007

Woodley of Menie, M. A., Fernandes, H. B. F., & Hopkins, W. D. (2015). The more *g*-loaded, the more heritable, evolvable, and phenotypically variable: Homology with humans in chimpanzee cognitive abilities. *Intelligence, 50,* 159–163. http://dx.doi.org/10.1016/j.intell.2015.04.002

Wright, A. A., Santiago, H. C., Sands, S. F., Kendrick, D. F., & Cook, R. G. (1985). Memory processing of serial lists by pigeons, monkeys, and people. *Science, 229,* 287–289. http://dx.doi.org/10.1126/science.9304205

Yong, A. G., & Pearce, S. (2013). A beginner's guide to factor analysis: Focusing on exploratory factor analysis. *Tutorials in Quantitative Methods for Psychology, 9,* 79–94. http://dx.doi.org/10.20982/tqmp.09.2.p079