The Odd–Even Effect in Sudoku Puzzles: Effects of Working Memory, Aging, and Experience

HYE-SANG CHANG and JANET M. GIBSON
Grinnell College

The odd–even effect in numerical processing has been explained as the easier processing of even numbers compared with odd numbers. We investigated this effect in Sudoku puzzles, a reasoning problem that uses numbers but does not require arithmetic operations. Specifically, we asked whether the odd–even effect occurred with Sudoku puzzles and whether individual differences in working memory (WM), aging, and experience with Sudoku modulated this effect. We manipulated the presence of odd and even numbers in Sudoku puzzles, measured WM with the Wisconsin Card Sorting Test and backward digit span task, tested older and younger adults, and collected Sudoku experience frequency. Performance on Sudoku was more accurate for even puzzles than odd ones. Younger, experienced, and higher-WM participants were more accurate on Sudoku, but these individual difference variables did not interact with the odd–even effect. Odd numbers may impose more cognitive load than even numbers, but future research is needed to examine how age, experience, or WM may influence the odd–even effect.

Numbers and math are an everyday part of life (Campbell, 2005). Perhaps that is why cognitive researchers strive to understand how numbers are mentally represented and manipulated, how people solve arithmetic problems, and how numerical processes use mental resources such as executive control processes of working memory (WM) (e.g., Campbell, Parker, & Doetzel, 2004; Dehaene, Bossini, & Giraux, 1993; Imbo, Duverne, & Lemaire, 2007; Shepard, Kilpatric, & Cunningham, 1975). One finding in such research is the odd–even effect, which concerns an advantage for even numbers in arithmetic tasks (Krueger, 1986; Lochy, Seron, Delazer, & Butterworth, 2000; Nuerk, Iversen, & Willmes, 2004; Vandorpe, De Rammelaere, & Vandierendonck, 2005). Examples of the effect include the findings that an odd result is more likely to be rejected in a sum verification task than an even result (Krueger & Hallford, 1984) and that parity judgments are speedier for even numbers than odd numbers (Hines, 1990).

Various explanations try to account for how parity information is related to other numerical properties (e.g., distinct from magnitude, Sudevan & Taylor, 1987) and is used in arithmetic procedures and strategies. Models include the idea that parity is calculated from an abstract representation of the number (McCloskey, 1992), an attribute of the mental representation of numbers for any format (Noel & Seron,

American Journal of Psychology
Fall 2011, Vol. 124, No. 3 pp. 313–324 • © 2011 by the Board of Trustees of the University of Illinois

1993), or a node in the network of processes used in computation (Campbell & Clark, 1988). The role of a mental number line and the spatial relationship of left and right to smaller and larger numbers (called the spatial–numerical association of response codes [SNARC] effect in the literature) implicate learned spatial processes in number processing (Dehaene et al., 1993; Fitousi, Shaki, & Algom, 2009). SNARC effects show that participants respond faster when larger numbers are responded to with the right hand and smaller numbers are responded to with the left hand in choice decision tasks. Similarly, participants respond faster when odd numbers are responded to with the left hand and even numbers with the right hand (called the markedness by association of response codes [MARC] effect in the literature; Willmes & Iversen, 1995).

The odd–even effect is important in understanding how people mentally represent numbers. Of interest to the present study is the hypothesis that the load of mentally representing and processing odd numbers is greater than that for even numbers. The load may arise from the fact that even-numbered results occur three times as often as the odd-numbered results in the multiplication table (Lochy et al., 2000). Hines (1990) argued that the effect may be due to the markedness of the words of the digits, given that the word *even* is used 2.6 times more frequently in the English language than the word *odd;* the unmarked word *even* is easier to process than the marked word *odd*. Findings by Cho and Proctor (2007) of an odd–even effect for word stimuli (e.g., *one*) and Arabic digits (e.g., *1*) and a larger effect for words than digits (Campbell et al., 2004) in parity judgment tasks implicate semantic memory processes (rather than specific numerical arithmetic operations) in the odd–even effect.

One problem in past research on the odd–even effect is that mental representation of numbers was confounded by task demands with the quantitative processing or retrieval of arithmetic rules from semantic memory. Research on the odd–even effect has used numbers as quantitative entities where some form of numerical processing is involved in the task (e.g., magnitude judgments, parity judgments, and arithmetic operations). It is unclear whether the load of odd numbers is in its representation or in its use in numerical operations. Additionally, previous research examining the odd–even effect was confounded by interactions of parity and magnitude, as in whether magnitude information is processed faster than parity (Sudevan & Taylor, 1987), or by issues with rote memory of the multiplication table or the plausibility of the solution (Lemaire & Fayol, 1995).

The present study tested the idea that odd numbers impose a higher load on memory than even numbers by testing performance on a nonarithmetic task, Sudoku puzzles. Sudoku is a deductive reasoning problem (Lee, Goodwin, & Johnson-Laird, 2008) with a unique solution that in popular form requires a 9×9 array to be filled with the numbers 1 through 9 so that no number appears more than once in a row, column, or 3×3 grid. The numbers are merely symbols in Sudoku; letters, words, or pictures are used instead of numbers in nonconventional forms of the game. Parity and the quantitative value of the numbers are irrelevant to the solving strategies of Sudoku. The strategies to solve Sudoku, as reported by Lee et al. (2008), include tactics concerning the exclusion and inclusion of digits that might fit in a cell and constraints on possible numbers that vary in their relational complexity. None of these tactics for solving Sudoku involve numerical operations. Furthermore, Hopfield (2008) presented a neural network model of Sudoku based on semantic memory searches; his model solved Sudoku puzzles without using arithmetic but instead treated the numbers as semantic concepts.

If the odd–even effect results from the ways we mentally represent numbers and not from its involvement in arithmetic operations, then we should find an odd–even effect with Sudoku puzzles. Furthermore, if odd numbers do present a load on memory resources, then individual differences may distinguish the performance of those handling the load well from that of those who might reason less well because of the memory load. We examined WM, age, and experience as three individual difference variables. As the following literature review will show, these variables have been studied to show differences in quantitative and reasoning processes.

Working Memory

WM is a system that stores, maintains, updates, manipulates, and processes information in the short term (Baddeley, 1986, 1996, 2000, 2003), and the central executive component of the model has spurred a great amount of interest in the study of executive processes as an individual difference variable. Individual differences in executive processes of WM have been explained by differences in attentional resources (Barrett, Tugade, & Engle, 2004). Barrett et al. suggested that controlled attention, related to the limited resources available to the central executive, determines the performance in complex cognitive tasks. According to Barrett et al., low-WM people are more affected by the cognitive load of tasks and are less efficient in maintaining goal-oriented processing because they do not have enough attentional resources. Higher-WM people are able to use attention to activate or suppress representations to perform tasks. They show fewer perseveration errors (Hansen & Goldinger, 2009).

WM affects syllogistic reasoning (Gilhooly, Logie, Wetherick, & Wynn, 1993; Gilhooly, Logie, & Wynn, 2002) and spatial reasoning (Klauer, Stegmaier, & Meiser, 1997). Probably all WM components—the central executive, phonological loop, visuospatial sketchpad, and episodic buffer—play a role in thinking about numbers (DeStefano & LeFevre, 2004; Imbo et al., 2007). WM demands of simple arithmetic and complex problem solving are well documented (e.g., DeStefano & LeFevre, 2004; Hecht, 2002; Imbo et al., 2007; Imbo & Vandierendonck, 2007; McLean & Hitch, 1999; Oberauer, Wendland, & Kliegl, 2003; Seitz & Schumann-Hengsteler, 2000). Mathematical reasoning (e.g., equation verification) task performance is associated with neural processes in the dorsolateral prefrontal cortex and the ventrolateral prefrontal cortex, brain areas considered to be involved in WM (Menon, Mackenzie, Rivera, & Reiss, 2002). Simplification strategies, such as rounding down to the nearest decade in multiplication tasks, are used when participants must perform a secondary task (Imbo et al., 2007). Older adults are slower on math when WM load is high (Oberauer et al., 2003). Furthermore, reasoning correctly when the information is contradictory to beliefs is impaired when memory is loaded with a secondary task (De Neys, 2006).

Aging

Age-related decline in cognitive tasks has been associated with reduced activation in the frontal lobe (Stuss, Craik, Sayer, Franchi, & Alexander, 1996) and distraction control (Hasher, Lustig, & Zacks, 2007). In mathematics, such age-related decline has been related to slower processing due to reduced availability of WM resources, whereas selective access to WM is intact in older adults (Oberauer et al., 2003). Fabre and Lemaire (2005) examined reaction time and event-related potentials in younger and older adults during a parity judgment task where primes were either congruent (even–even) or incongruent (even–odd) with targets; they found that both age groups showed priming, but event-related potentials showed that the N400 effect, which revealed the processing of the incongruent prime 400 ms after the stimulus, was delayed and was smaller in older adults than in young adults. These findings contribute to the argument that parity is not automatically processed in older adults. Accordingly, older adults have fewer cognitive resources than younger adults to process numbers.

Experience

Experience of doing Sudoku increases familiarity with the task and efficiency of one's solving strategies. In addition, efficiency decreases with increased task difficulty, when there is demanding use of limited cognitive resources; experts carry less cognitive load than novices during problem solving (Sweller, 1988). Lee et al. (2008) reported data that reveal the difficulty of reasoning from inexperienced Sudoku solvers who were given mild or difficult puzzles without training; they showed that in 15 min the average number of cells filled in correctly was only 2.2 (Experiment 1). However, puzzles especially created to rely on simple tactics (Experiment 2) resulted in 50–83% filled in correctly within 4 min by novices. Thus, Sudoku solving places a heavy load on novices, who must discover and develop their own strategies. A more experienced Sudoku player not only will be able to solve more of the puzzle than an inexperienced player but may be less likely to show an odd–even effect if his or her experience compensates for the load of representing odd numbers.

The Present Study

We controlled the presence of odd and even numbers in Sudoku puzzles by replacing existing values in easy puzzles so that they contained only odd numbers

(11 through 27), only even numbers (10 through 26), or mixed numbers (11 through 19). Pilot work indicated that when puzzles contained both single- and double-digit numbers, participants were biased to working with the single-digit numbers, and therefore our puzzles used only double-digit numbers. Compared with regular Sudoku puzzles that use only the single digits 1 through 9, our double-digit puzzles provided more processing complexity (supported by most participants' ad hoc self report). Performance on our puzzles was meant not as a study of Sudoku but as a study of the odd–even effect (i.e., we made no predictions about how people solved the puzzles). We also collected data on mathematical word problems and calculation problems to establish that our participants showed the odd–even effect in arithmetic tasks.

For WM measures, we used two conventional neuropsychological tests that measure WM: the Wisconsin Card Sorting Test (WCST) and the backward digit span task (BDST). The WCST requires the ability to plan, sequence, form concepts, and regulate goal-directed behaviors (Cinan & Tanör, 2002; Stratta et al., 1997). Perseverative errors on WCST (WCSTPE) increase when executive functioning fails (Ridderinkhof, Span, & van der Molen, 2002) and when WM fails to update (Hartman, Bolton, & Fehnel, 2001). The BDST involves active manipulation of the initially registered information (Prifitera, Saklofske, & Weiss, 2004). The BDST predicts children's solving of mathematical word problems (Pavlin-Bernardic, Vlahovic-Stetic, & Arambasic, 2008). Peters, Jelicic, Verbeek, and Merckelbach (2007) found that lower BDST scores resulted in higher false recognition, which supports the idea that the BDST measures people's ability to monitor information in WM. Backward memory span decreases with age (Myerson, Emery, White, & Hale, 2003).

Researchers have questioned what exactly the BDST and other measures of WM span (e.g., operation span) measure (Unsworth & Engle, 2007; Waters & Caplan, 2003). Waters and Caplan reported that the BDST had a test–retest reliability of .65 (Table 10), was significantly decreased by age (6.2 for adults 18–30 years old, 5.3 for adults older than 70 years; Table 4), was significantly correlated with other measures of WM (that used words or digits; Table 7), and had the highest factor loadings on a single factor of all the WM span tests compared (.88; Table 9), with a Cronbach's alpha of .84 (Table 11). However, Waters and Caplan found that all the WM span tests were unable to classify people reliably as high functioning or low functioning after 1.5 months. One salient criticism of their findings is that they assumed WM resources are stable over time. It is likely that people's available resources shift daily and perhaps within a day (May, Hasher, & Stoltzfus, 1993); we suggest that measures taken in the same hour are good measures of WM and how it is used on other tasks completed around that time. Furthermore, Unsworth and Engle (2007) argued that, based on their meta-analysis of WM measures, including the BDST, there may be no real difference between "simple" or "short-term" measures and "complex" and "working-memory" measures. They posited that short-term memory and WM are the same construct and that WM tasks such as the BDST are a measure of executive functions. We will use the language WM to be consistent with the class of literature studying executive control processes, but we remain neutral on the question of whether WM and short-term memory are the same construct.

Hypotheses

We hypothesized that our participants would show the odd–even effect on arithmetic problems (word and calculations), replicating the common findings in the literature. Specifically, we expected higher accuracy on problems that contained even numbers compared with odd.

The load hypothesis predicts that we would observe an odd–even effect in Sudoku. Specifically, the even-numbered Sudoku puzzle would be more accurately solved than the odd-numbered one. We also tested a mixed puzzle in order to examine whether a mixed puzzle would behave more like an even one or an odd one, fall in between, or be superior to all-even or all-odd puzzles because it is more like the traditional puzzle and has a continuous range of numbers.

We hypothesized that individual differences in WM, age, and experience would affect Sudoku performance in that those with higher WM resources, younger age, and more experience with Sudoku would successfully complete more of the Sudoku puzzles. If the load hypothesis is correct, we hypothesized that our individual difference variables would affect the degree of the observed odd–even effect. We predicted that if odd numbers impose a high load on WM resources, a greater odd–even effect would be found in people with lower WM than those with higher WM, as measured by WCSTPE and span scores on BDST (BDS). Similarly, those with less experience with Sudoku would show a greater odd–even effect than those with more experience. Older adults would also show a greater effect because they have fewer cognitive resources to deal with the load of odd numbers.

EXPERIMENT

METHOD

Participants

Forty-six older adults (age: M = 66.72 years, SD = 8.88 years; education: M = 15.41 years, SD = 2.45 years; 30 women) from the community were recruited via mailings and newspaper ads that advertised a memory, Sudoku, and math study for participants over the age of 50. Thirty-nine younger adults (age: M = 19.69 years, SD = 1.17 years; education: M = 14.18 years, SD = 1.17 years; 23 women) were recruited through campus flyers and newspaper ads. Participants received $10 or $15 for participating in the study.

Materials

The WCST (Heaton, 2005) and the BDST (written in E-Prime) were administered individually on a PC. The Sudoku puzzles (modified versions by Greenspan & Lee, 2005; see Figure 1) and math problems (Figure 2) were administered as paper-and-pencil tasks. A questionnaire collected information on their frequency of play with Sudoku; those who indicated that they played fewer than 10 times or never were classified as "novice," and those who played more frequently were classified as "experienced." A bottle of water was provided, and a hand timer was used for the paper-and-pencil tasks.

Design

As a within-subject design, each participant received a set of 3 Sudoku puzzles (1 even, 1 odd, and 1 mixed), math problems, the WCST, and the BDST, in counterbalanced orders across participants. We used two types of math tasks: word problems (2 with odd and 2 with even numbers) and computations (8 odd, 8

Even

	26			10		18	16	
20		10	12		16	26		22
	24	14		22	26	20	12	
	16	26	24		12		10	18
	10				22		20	
22	12		10			14		24
		18	22	24		12	14	
	14	12	16		20	10		26
	22	16		12			18	20

Mixed

17	19		18			13	12	
		13		17	16	19		15
18	15			13			17	16
			19		15	18		12
19	12	15	14	11	18	17	16	
14			17		13		11	
16	11			19		12		18
	18	12	13	14		16		
13		19			12		15	14

Odd

11	15		25		23	21	19	
23			27	11			13	25
	25			19	15	23		27
25				21			27	
	19	15		17		11	23	
21	17				27	19		13
19		13		27	21		15	23
	27			23	19		21	
15	23	21	11		13	27		19

FIGURE 1. Even, mixed, and odd Sudoku puzzles

Jason and Bob together have 93 marbles. Bob has 17 marbles less than Jason; if Jason gives Bob 5 of them, how many more marbles does Jason have than Bob?

Two numbers add up to 64. The difference between the two numbers is 8. What are the numbers?

53	77	43	13	12	72
–39	÷ 7	+ 12	x 6	+ 38	÷ 8

FIGURE 2. Examples of math problems

even, and 8 mixed values). The strings of the digits used in the BDST were also counterbalanced across trials with all odd, all even, or mixed numbers occurring in the strings.

The independent variables included the number type in the tasks (odd, even, or mixed), age group (younger or older), and experience with Sudoku (novice or experienced). We also measured WCSTPE, BDS, and number of correct trials for the BDST (BDSN), where series of digits were recalled in reverse order within each trial (Turner & Engle, 1989), as measures of WM. Order of the tasks was also examined as a possible confound, as was gender.

The dependent variables were accuracy for Sudoku puzzles (proportion of the 35 empty cells correctly filled in) and math word problems and calculation problems (proportion of problems correctly solved).

Procedure

Initially, all participants provided informed consent and completed the demographic questionnaire. Four participants in the older group reported that they had almost no computer experience; they were provided instructions concerning typing the digits on the keyboard, pressing the "Enter" key, and moving and clicking with the mouse for the BDST and WCST. No participant failed to carry out the tasks because of low familiarity with the computer.

Each task was performed individually after the general instructions were given. Half of the participants began with the WCST and the other half with the BDST. A standard Sudoku puzzle with the numbers 1 through 9 was given for practice before the experimental puzzles were provided. After completing the Sudoku puzzles, participants completed the sheet of math word problems and the sheet of computation problems. The final task was the WM task not completed at the beginning. The WM tasks (WCST and BDST) were self-paced, ranging from 10 min to 20 min for WCST and from 3 min to 10 min for BDST, whereas the Sudoku puzzles and math problems were allotted 5 min for each type of puzzle and math task. Participants were free to take breaks between tasks. The experiment was completed within 1.5 hr.

RESULTS

Odd–Even Effect

We observed the odd–even effect in Sudoku puzzles, as participants accurately completed more even-numbered Sudoku puzzles ($M = .68$, $SD = .27$) than odd-numbered ones ($M = .58$, $SD = .29$), $t(84) = 3.83$, $p < .001$. Even-numbered math word problems ($M = .71$, $SD = .37$) were also solved more accurately than the odd-numbered ones ($M = .35$, $SD = .38$), $t(84) = 8.64$, $p < .001$. Performance in the math computation problems was near ceiling, and no difference between even ($M = .94$, $SD = .10$) and odd problems ($M = .95$, $SD = .09$) was observed, $t(84) = 0.89$, $p > .05$.

Because the order of working the Sudoku puzzles or the math word problems could affect performance, we examined the effect of order and its interaction with the odd–even effect in a 2 (odd–even) × 6 (orders) mixed ANOVA; order did not affect accuracy or interact with the odd–even effect in either the Sudoku puzzles or the math word problems (Fs < 1.60).

Our design included a mixed Sudoku puzzle ($M = .67$, $SD = .28$). A repeated-measures ANOVA with three levels for type of puzzle found that the means of the odd, even, and mixed puzzles differed significantly, $F(2, 168) = 11.07$, $p < .001$, $\eta^2 = .12$, and pairwise comparisons (LSD) revealed that the means of mixed puzzles did not differ from even puzzles, $p > .05$, but did differ from odd puzzles, $p < .001$. Thus, a Sudoku with double-digit mixed numbers was completed as accurately as that of one with even numbers, but both even and mixed puzzles were more accurately solved within 5 min than an odd puzzle.

The mean proportion correct (and standard deviation) for each of our individual difference variables as a function of type of numbers used in Sudoku puzzles is reported in Table 1 and for math problems in Table 2.

Aging

A 2 (odd–even) × 2 (younger–older) mixed ANOVA of the means in the first two columns of Table 1 revealed that younger adults accurately completed more of the Sudoku puzzles, $F(1, 83) = 9.63$, $p < .003$, $\eta^2 = .10$; even puzzles were completed significantly more often than odd puzzles, $F(1, 83) = 13.89$, $p < .001$, $\eta^2 = .14$; and there was no significant interaction, $F(1, 83) = 1.37$, $p > .05$, $\eta^2 = .02$.

The lack of an interaction fails to support our prediction that the odd–even effect for Sudoku solving would be stronger in older adults. When the data were analyzed separately for each age group, we found a significant odd–even effect for older adults,

$F(1, 45) = 11.56$, $p < .001$, $\eta^2 = .20$, but not for younger adults, $F(1, 38) = 3.58$, $p > .05$, $\eta^2 = .09$.

A 2 (odd–even) × 2 (age) mixed ANOVA of the first 2 columns of Table 2 revealed that younger adults accurately completed more of the word problems, $F(1, 83) = 32.60$, $p < .001$, $\eta^2 = .28$; even problems were completed significantly more often than odd problems, $F(1, 83) = 72.99$, $p < .001$, $\eta^2 = .47$; and there was no significant interaction, $F(1, 83) < 1$. As with Sudoku, the lack of an interaction does not support the idea that older adults would find the odd math word problems more difficult than even problems. Separate analyses for the age groups supported the lack of an interaction: An odd–even effect was found for older adults, $F(1, 43) = 45.47$, $p < .001$, $\eta^2 = .50$, and for younger adults, $F(1, 38) = 29.22$, $p = .001$, $\eta^2 = .44$. Performance on word problems did not show any significant effects or interactions with age, Fs < 1.3.

Experience

Forty-three participants (23 older adults and 20 younger adults) had little or no prior experience with Sudoku. Coding for experience, we analyzed the data to see whether there was an effect of experience and whether it interacted with the odd–even effect. Using a 2 × 2 mixed ANOVA, we found that experienced players were more accurate on the puzzles, $F(1, 83) = 36.75$, $p < .001$, $\eta^2 = .31$; and the significant odd–even effect, $F(1, 83) = 14.78$, $p < .001$, $\eta^2 = .15$, did not interact with experience, $F < 1.15$. Separate analyses for each experience level found a significant odd–even effect: For novices, $F(1, 42) = 4.79$, $p < .05$, $\eta^2 = .10$; and for experienced players, $F(1, 41) = 10.01$, $p < .005$, $\eta^2 = .20$.

Working Memory

As a whole, our sample on average made 10.00 WCSTPE ($SD = 7.97$). Their average BDS was 5.32 ($SD = 1.56$) digits, and when only correct trials before meeting quitting criteria (BDSN) were counted, they averaged 10.67 trials ($SD = 3.55$) correctly recalled in reverse order. Because analyses with either BDST measure (BDS or BDSN) were very similar, we report analyses based on BDS. We computed high- and low-WM groups by converting the WCSTPE and BDS into Z scores, reversed the signs for WCTSPE when summing them so that positive Z scores reflected stronger WM, and placed those below 0 into the low-WM group and those above 0 into the high-WM group. The aggregate of WM scores is considered a better measure than one task alone (Waters & Caplan, 2003), although the use of WM measures and categorical splits is controversial (see Miyake, Emerson, & Friedman, 1999, and Waters & Caplan, 2003, for arguments on either side). Analyses with WCSTPE or BDS separately were similar to those of the composite score.

In a 2 (odd, even) × 2 (high WM, low WM) mixed ANOVA on Sudoku puzzles, we found an effect of WM, $F(1, 83) = 21.54$, $p < .001$, $\eta^2 = .21$; an odd–even effect, $F(1, 83) = 14.38$, $p = .001$, $\eta^2 = .15$; and no interaction, $F < 1.15$. The low-WM group showed the odd–even effect, $F(1, 43) = 11.87$, $p < .001$, $\eta^2 = .22$; and the high-WM group was just about at the alpha criterion of significance, $F(1, 40) = 3.71$, $p = .06$, $\eta^2 = .09$.

TABLE 1. Mean Proportion Correct (*SD*) for Even, Odd, and Mixed Sudoku Puzzles as a Function of Age, Experience, and Working Memory

	Age		Experience		Working memory	
	Younger (*n* = 39)	Older (*n* = 46)	Novice (*n* = 43)	Experienced (*n* = 42)	Low (*n* = 44)	High (*n* = 41)
Even	.73 (.25)	.59 (.28)	.50 (.25)	.81 (.20)	.55 (.25)	.87 (.26)
Odd	.68 (.26)	.49 (.28)	.44 (.28)	.72 (.23)	.45 (.24)	.71 (.28)
Mixed	.74 (.24)	.61 (.29)	.53 (.24)	.81 (.23)	.57 (.27)	.78 (.24)

TABLE 2. Mean Proportion Correct (*SD*) for Even, Odd, and Mixed Arithmetic Problems as a Function of Age, Experience, and Working Memory

	Age		Sudoku experience		Working memory	
	Younger (*n* = 39)	Older (*n* = 46)	Novice (*n* = 43)	Experienced (*n* = 42)	Low (*n* = 44)	High (*n* = 41)
			Word problems			
Even	.88 (.27)	.55 (.38)	.72 (.37)	.69 (.38)	.57 (.40)	.85 (.28)
Odd	.54 (.40)	.18 (.27)	.33 (.39)	.37 (.37)	.26 (.35)	.44 (.39)
			Calculation problems			
Even	.97 (.05)	.97 (.07)	.96 (.07)	.98 (.05)	.96 (.06)	.97 (.06)
Odd	.96 (.08)	.93 (.11)	.92 (.12)	.97 (.07)	.94 (.11)	.95 (.07)
Mixed	.96 (.09)	.95 (.10)	.93 (.12)	.98 (.05)	.95 (.10)	.96 (.08)

In looking at math performance (see means in the last two columns of Table 2), for word problems we found an effect of WM, $F(1, 83) = 25.24, p < .001, \eta^2 = .26$; and the significant odd–even effect did not interact with WM, $F(1, 83) = 1.90, p > .05, \eta^2 = .03$. Each WM group showed the odd–even effect: low WM, $F(1, 43) = 26.95, p < .001, \eta^2 = .39$; and high WM, $F(1, 40) = 51.72, p < .001, \eta^2 = .56$.

All Three Individual Difference Variables

Figure 3 shows the distribution of means for the combinations of the variables. A 2 (even–odd) × 2 (age) × 2 (Sudoku experience) × 2 (WM) mixed ANOVA showed that when all three variables were in the same analyses, significant effects were found for odd–even, $F(1, 77) = 11.95, p < .01, \eta^2 = .13$; experience, $F(1, 77) = 43.22, p < .01, \eta^2 = .36$; and WM, $F(1, 77) = 17.23, p < .01, \eta^2 = .18$. No other effects or interactions were significant. We note that the small sample sizes in each cell hindered the ability to find significant interactions.

Multiple regression analyses were conducted with WM entered as a continuous variable. Table 3 shows that variance in total Sudoku performance is significantly accounted for, but when the odd–even effect was the dependent measure (odd/[odd + even]), no individual difference variables or any interactions significantly accounted for the variance.

Gender and Math Performance

As a check on the possible individual difference of gender confounding our findings, we examined the difference between genders. No effects of gender or interactions were found on Sudoku puzzles, Fs < 1.50, or on math calculation problems. For math word problems, an effect of gender, with higher performance for men, approached alpha, $F(1, 69) = 3.41, p = .069, \eta^2 = .05$, and no interactions were significant.

DISCUSSION

After replicating the odd–even effect in math problems, we found an odd–even effect in Sudoku puzzles. Overall, even-numbered Sudoku puzzles were solved significantly more accurately than the odd-numbered ones. Because the numbers in Sudoku are essentially symbols and the digits' numerical properties are irrelevant, the data suggest that the odd–even effect does not require numerical operations. This finding is important because previous research had not explored the effect in tasks where the numbers were used only as symbols.

Our finding lends support to theories that focus on load and the assumption that odd numbers have a different abstract representation than even numbers (McCloskey, 1992). However, our data tell us only

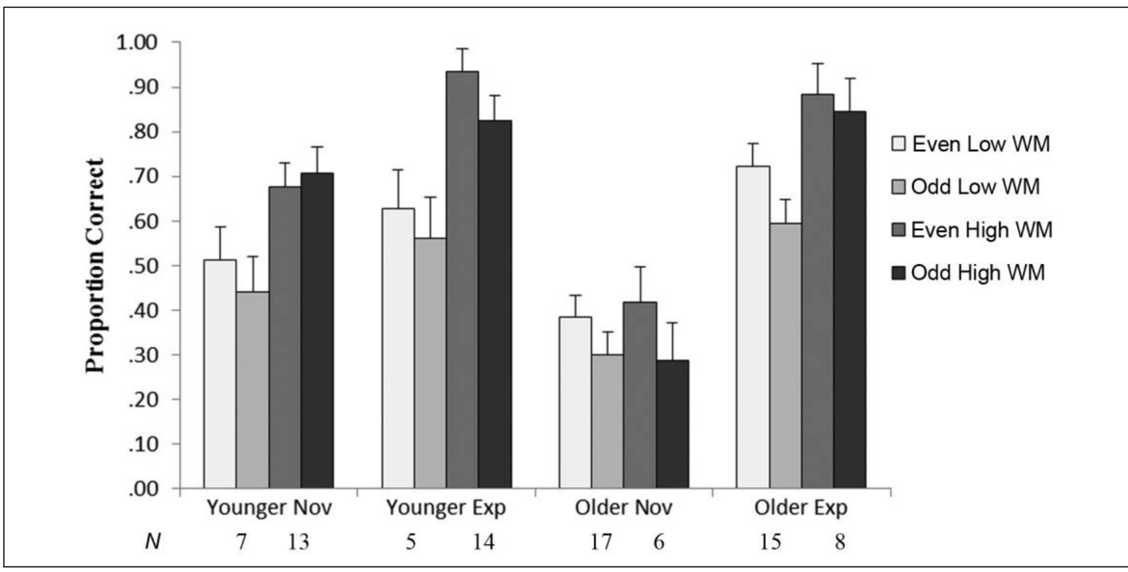

FIGURE 3. Mean proportion correct for even and odd Sudoku puzzles for participants classified in combinations of low and high working memory (WM), younger and older age groups, and novice (Nov) and experienced (Exp) players. The error bars are standard errors. Sample sizes for each WM × age × experience combination appear under each couplet of bars for that combination

that the effect occurs without requiring computational processes acting on the mental representation or a focus on its numerical attributes; it is also possible that the difference in load is observed when other processes acting on the mental representation (in our case, reasoning) are facilitated by other relevant factors. For example, our finding is consistent with models demonstrating that the mental load of odd numbers when processed may be high because of different network weights for computation (e.g., Campbell & Clark, 1988) changing the way in which the numbers are represented as *symbols* not quantities, less experience with odd numbers (Lochy et al., 2000), the markedness of *odd* in language (Hines, 1990), and less efficient retrieval from semantic memory (Campbell et al., 2004). All of these theories suggest that familiarity and frequency of processing for even numbers could ease the load of reasoning in Sudoku with even numbers.

Our data from the mixed puzzle support an important role of familiarity. Mixed puzzles were solved as well as even puzzles and better than odd puzzles. The mixed puzzle was the most familiar to experienced solvers because the numbers were ordered in sequence without alternating (e.g., 11, 12, 13 . . . 19), and it could be treated as the familiar Sudoku puzzle with numbers 1 though 9 by taking away the first digit. Future research is needed to tease apart the role of familiarity in reasoning with the numbers in the construction of the mental representation of them. It will be of interest to extend the study of the odd–even effect with nonarithmetic processes like those used in Sudoku in order to understand whether the load of thinking about odd numbers is better explained by familiarity (Lochy et al., 2000), markedness (Hines, 1990), or semantic memory (Campbell et al., 2000; Hopfield, 2008). Familiarity is also an important variable for researchers to study because we use numbers in everyday life, not just in calculations. Given that the odd–even effect is present when people think with numbers even in a nonmathematical way, an effort should be made to familiarize thinking with odd numbers (perhaps by increasing their frequency to that of even numbers) in current education curricula.

The load hypothesis suggested that participants who have fewer cognitive resources for handling the load would show a greater odd–even effect than those with more. We examined three individual difference variables: age, experience, and WM, three variables known to play a role in availability of cognitive resources for handling load. We did not find significant interactions when the variance of the entire sample was used in the ANOVAs. However, when the data were analyzed separately for each variable, we found a difference between older and younger participants: a significant odd–even effect in older but not younger adults. We also found a difference between lower- and higher-WM groups: a significant odd–even effect for low WM but not (barely) for those with high WM, with the effect sizes .23 and .12 in the low and high WM, respectively. Therefore, some support for the idea that odd numbers consume more WM resources than even numbers can be found in the data. However, individual differences in WM, aging, and experience modulated performance on even-numbered tasks as well (e.g., even-number processing was compromised in people with low cognitive resources or task familiarity). Thus, the degree of the odd–even effect did not vary significantly across participants.

TABLE 3. Multiple Regression Analyses for Total Performance and the Odd–Even Effect for Working Memory (WM), Experience (Exp), and Age

	R^2	F	p
Total performance			
Model 1: WM, Exp, Age	.54	31.70	<.001
Model 2: WM × Exp, WM × Age, Age × Exp	.58	17.93	<.001
Model 3: WM × Exp × Age	.58	15.17	<.001
Odd–even effect [odd/(odd + even)]			
Model 1: WM, Exp, Age	.07	1.96	.13
Model 2: WM × Exp, WM × Age, Age × Exp	.11	1.68	.14
Model 3: WM × Exp × Age	.11	1.42	.21

Age, experience, and WM affected total accuracy on Sudoku. Younger adults were better able to solve the puzzles than older adults, experienced players were better able to solve the puzzles than novices, and those with higher WM were better able to complete the puzzles than those with lower WM (WCSTPE scores, BDS, BDSN successful trials, or a WCSTPE and BDS composite score). But none of these variables interacted with the odd–even effect in clear ways so as to provide strong support that the load of odd compared with even numbers is explained by the processing differences due to age, experience, or WM. Future research with other measures of WM, perhaps those requiring sharing of resources, such as operation span, might capture the executive processes that help people better process difficult material such as odd numbers or those that are used particularly with reasoning (De Neys, 2006; Hansen & Goldinger, 2009; Ridderinkhof et al., 2002). In addition, the roles of the other components of WM, the phonological loop, visuospatial sketchpad, and episodic buffer, could also be explored to help inform researchers on how odd and even numbers differ in their load on resources.

Our study contained several limitations. In the present study, we gave participants 5 min to work on each puzzle. This limit may have been too small for some inexperienced puzzle players and too generous for some experienced players. Time to complete the puzzle is an important variable that future research should examine to determine whether age, experience, or WM predicts the speed of reasoning with even and odd numbers. Furthermore, because of practical limitations, especially when fatigue matters to performance, our study included only one of each type of puzzle (odd, even, and mixed). Future research may find that the use of numerous puzzles strengthens the data and allows manipulations of puzzle difficulty or placement of numbers in the puzzle; for example, spatial representations of numbers, known to matter in numerical processing (e.g., Fischer, 2003; Ito & Hatta, 2004), may interact with the placement of even or odd numbers provided by the experimenter or provided by the participant. Such manipulations may allow researchers to determine whether the odd–even effect can be increased or decreased with manipulations of concurrent task load.

Finding an odd–even effect with Sudoku has interesting theoretical implications for reasoning as well. Lee et al. (2008), who studied the role of deduction in Sudoku, argued that reasoning theories must account for the universal abilities of people around the world to do Sudoku. We believe that in Sudoku, a theory of reasoning would not be complete without consideration that the materials are numerical. Just as familiarity helps with conditional reasoning (Cox & Griggs, 1982; Griggs & Cox, 1982), even numbers may help relative to odd numbers with deductive reasoning in Sudoku. An odd–even effect in Sudoku may complement research done with neural networks and computer simulation models concerning reasoning and symbolic representations of numbers (Hopfield, 2008; Verguts & Fias, 2008).

To conclude, the present study is the first that we know of to show that the odd–even effect occurs in a task that does not require arithmetic operations to be executed or parity or magnitude to be processed. The data support the idea that the odd–even effect may occur because of the nature of the mental representation of the numbers, even without mathematical processes. The data are consistent with theories that suggest that load might be higher for odd numbers because of familiarity, markedness, or semantic memory knowledge of numbers, all of which could be activated when people think about numbers as symbols in Sudoku. Accuracy for each type of Sudoku was affected by each of three individual difference variables that together have not been studied with the odd–even effect: age, experience with Sudoku, and WM. In addition, there was some support from separate group analyses and size effects to suggest that these individual difference variables may affect the degree of the effect, providing partial evidence for the load hypothesis of the odd–even effect. Furthermore, theories of deductive reasoning may converge with theories of mathematical processing to support the interpretation that the odd–even effect in reasoning with numbers occurs because of the load odd numbers present to memory compared with even numbers.

NOTE

Grinnell College provided the funding for this research.

Address correspondence about this article to Janet M. Gibson, Department of Psychology, Grinnell College, Grinnell, IA 50112 (e-mail: gibsonj@grinnell.edu).

REFERENCES

Baddeley, A. D. (1986). *Working memory*. Oxford, England: Oxford University Press.

Baddeley, A. D. (1996). Exploring the central executive. *Quarterly Journal of Experimental Psychology, 49,* 5–28.

Baddeley, A. D. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences, 4,* 417–423.

Baddeley, A. D. (2003). Working memory: Looking back and looking forward. *Nature Reviews: Neuroscience, 4,* 829–839.

Barrett, L. F., Tugade, M. M., & Engle, R. W. (2004). Individual differences in working memory capacity and dual-process theories of mind. *Psychological Bulletin, 130,* 553–573.

Campbell, J. I. D. (Ed.). (2005). *Handbook of mathematical cognition*. New York, NY: Psychology Press.

Campbell, J. I. D., & Clark, J. M. (1988). An encoding-complex view of cognitive numerical processing: Comment on McCloskey, Sokol, and Goodman (1986). *Journal of Experimental Psychology: General, 117,* 204–214.

Campbell, J. I. D., Parker, H. R., & Doetzel, N. L. (2004). Interactive effects of numerical surface form and operand parity in cognitive arithmetic. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30,* 51–64.

Cho, Y. S., & Proctor, R. W. (2007). When is an odd number not odd? Influence of task rule on the MARC effect for numeric classification. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33,* 832–842.

Cinan, S., & Tanör, O. O. (2002). An attempt to discriminate different types of executive functions in the Wisconsin Card Sorting Test. *Memory, 10,* 277–289.

Cox, J. R., & Griggs, R. A. (1982). The effects of experience on performance in Wason's selection task. *Memory & Cognition, 10,* 496–502.

Dehaene, S., Bossini, S., & Giraux, P. (1993). The mental representation of parity and numerical magnitude. *Journal of Experimental Psychology: General, 122,* 371–396.

De Neys, W. (2006). Dual processing in reasoning: Two systems but one reasoned. *Psychological Science, 17,* 428–433.

DeStefano, D., & LeFevre, J. (2004). The role of working memory in mental arithmetic. *European Journal of Cognitive Psychology, 16,* 353–386.

Fabre, L., & Lemaire, P. (2005). Age-related differences in automatic stimulus–response associations: Insights from young and older adults' parity judgments. *Psychonomic Bulletin & Review, 12,* 1100–1105.

Fischer, M. H. (2003). Spatial representations in number processing: Evidence from a pointing task. *Visual Cognition, 10,* 493–508.

Fitousi, D., Shaki, S., & Algom, D. (2009). The role of parity, physical size, and magnitude in numerical cognition: The SNARC effect revisited. *Attention, Perception & Psychophysics, 71,* 143–155.

Gilhooly, K. J., Logie, R. H., Wetherick, N. E., & Wynn, V. (1993). Working memory and strategies in syllogistic reasoning tasks. *Memory & Cognition, 21,* 115–124.

Gilhooly, K. J., Logie, R. H., & Wynn, N. E. (2002). Syllogistic reasoning tasks and working memory: Evidence from sequential presentation of premises. *Current Psychological Research, 21,* 111–120.

Greenspan, G., & Lee, R. (2005). Easy puzzle. *Web Sudoku*. Retrieved from http://www.websudoku.com

Griggs, R. A., & Cox, J. R. (1982). The elusive thematic materials effect in Wason's selection task. *British Journal of Psychology, 73,* 407–429.

Hansen, W. A., & Goldinger, S. D. (2009). Taboo: Working memory and mental control in an interactive task. *American Journal of Psychology, 122,* 283–291.

Hartman, M., Bolton, E., & Fehnel, S. E. (2001). Accounting for age differences on the Wisconsin Card Sorting Test: Decreased working memory, not inflexibility. *Psychology and Aging, 16,* 385–399.

Hasher, L., Lustig, C., & Zacks, R. T. (2007). Inhibitory mechanisms and the control of attention. In A. R. A. Conway, C. Jarrold, M. Kane, A. Miyake, & J. N. Towse (Eds.), *Variation in working memory* (pp. 227–249). New York, NY: Oxford University Press.

Heaton, R. K. (2005). *Wisconsin Card Sorting Test* (Computer Ver. 4, res. ed.). Lutz, FL: Psychological Assessment Resources.

Hecht, S. A. (2002). Counting on working memory in simple arithmetic when counting is used for problem solving. *Memory & Cognition, 30,* 447–455.

Hines, T. M. (1990). An odd effect: Lengthened reaction times for judgments about odd digits. *Memory & Cognition, 18,* 40–46.

Hopfield, J. J. (2008). Searching for memories, Sudoku, implicit check bits, and the iterative use of not-always-correct rapid neural computation. *Neural Communication, 20,* 1119–1164.

Imbo, I., Duverne, S., & Lemaire, P. (2007). Working memory, strategy execution, and strategy selection in mental arithmetic. *Quarterly Journal of Experimental Psychology, 60,* 1246–1264.

Imbo, I., & Vandierendonck, A. (2007). The role of phonological and executive working memory resources in simple arithmetic strategies. *European Journal of Cognitive Psychology, 19,* 910–933.

Ito, Y., & Hatta, T. (2004). Spatial structure of quantitative representation of numbers: Evidence from the SNARC effect. *Memory & Cognition, 32,* 662–673.

Klauer, K. C., Stegmaier, R., & Meiser, T. (1997). Working memory involvement in propositional and spatial reasoning. *Thinking & Reasoning, 3,* 9–48.

Krueger, L. E. (1986). Why $2 \times 2 = 5$ looks so wrong: On the odd–even rule in product verification. *Memory & Cognition, 14,* 141–149.

Krueger, L. E., & Hallford, E. W. (1984). Why $2 + 2 = 5$ looks so wrong: On the odd–even rule in sum verification. *Memory & Cognition, 12,* 171–180.

Lee, N. Y. L., Goodwin, G. P., & Johnson-Laird, P. N. (2008). The psychological puzzle of Sudoku. *Thinking & Reasoning, 14,* 342–364.

Lemaire, P. P., & Fayol, M. (1995). When plausibility judgments supersede fact retrieval: The example of the odd–even effect on product verification. *Memory & Cognition, 23,* 34–48.

Lochy, A., Seron, X., Delazer, M., & Butterworth, B. (2000). The odd–even effect in multiplication: Parity rule or familiarity with even numbers? *Memory & Cognition, 28,* 358–365.
May, C. P., Hasher, L., & Stoltzfus, E. R. (1993). Optimal time of day and the magnitude of age differences in memory. *Psychological Science, 4,* 326–330.
McCloskey, M. (1992). Cognitive mechanisms in numerical processing: Evidence from acquired dyscalculia. *Cognition, 44,* 107–157.
McLean, J. F., & Hitch, G. J. (1999). Working memory impairments in children with specific arithmetic learning difficulties. *Journal of Experimental Child Psychology, 74,* 240–260.
Menon, V., Mackenzie, K., Rivera, S. M., & Reiss, A. L. (2002). Prefrontal cortex involvement in processing incorrect arithmetic equations: Evidence from event-related fMRI. *Human Brain Mapping, 16,* 119–130.
Miyake, A., Emerson, M. J., & Friedman, N. P. (1999). Good interactions are hard to find. *Behavioral and Brain Sciences, 22,* 108–109.
Myerson, J., Emery, L., White, D. A., & Hale, S. (2003). Effects of age, domain, and processing demands on memory span: Evidence for differential decline. *Aging, Neuropsychology, and Cognition, 10,* 20–27.
Noel, M.-P., & Seron, X. (1993). Arabic number reading deficit: A single case study when 236 is read [2306] and judged superior to 1258. *Cognitive Neuropsychology, 10,* 317–339.
Nuerk, H.-C., Iversen, W., & Willmes, K. (2004). Notational modulation of the SNARC and the MARC (linguistic markedness of response codes) effect. *Quarterly Journal of Experimental Psychology Section A, 57,* 835–863.
Oberauer, K., Wendland, M., & Kliegl, R. (2003). Age differences in working memory: The roles of storage and selective access. *Memory & Cognition, 31,* 563–569.
Pavlin-Bernardic, N., Vlahovic-Stetic, V., & Arambasic, L. (2008). Children's solving of mathematical word problems: The contribution of working memory. *Review of Psychology, 15,* 35–43.
Peters, M. J. V., Jelicic, M., Verbeek, H., & Merckelbach, H. (2007). Poor working memory predicts false memories. *European Journal of Cognitive Psychology, 19,* 213–232.
Prifitera, A., Saklofske, D. H., & Weiss, L. G. (Eds.). (2004). *WISC–IV clinical use and interpretation: Scientist-practitioner perspectives.* New York, NY: Elsevier.
Ridderinkhof, K. R., Span, M. M., & van der Molen, M. W. (2002). Perseverative behavior and adaptive control in older adults: Performance monitoring, rule induction, and set shifting. *Brain and Cognition, 49,* 382–401.
Seitz, K., & Schumann-Hengsteler, R. (2000). Mental multiplication and working memory. *European Journal of Cognitive Psychology, 12,* 552–570.
Shepard, R. N., Kilpatric, D. W., & Cunningham, J. P. (1975). The internal representation of numbers. *Cognitive Psychology, 7,* 82–138.
Stratta, P., Daneluzzo, E., Prosperini, P., Bustini, M., Mattei, P., & Rossi, A. (1997). Is Wisconsin Card Sorting Test performance related to "working memory" capacity? *Schizophrenia Research, 27,* 11–19.
Stuss, D. T., Craik, F. I. M., Sayer, L., Franchi, D., & Alexander, M. P. (1996). Comparison of older people and patients with frontal lesions: Evidence from word list learning. *Psychology and Aging, 11,* 387–395.
Sudevan, P., & Taylor, D. A. (1987). The cuing and priming of cognitive operations. *Journal of Experimental Psychology: Human Perception and Performance, 13,* 89–103.
Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Sciences, 12,* 257–285.
Turner, M. L., & Engle, R. W. (1989). Is working memory capacity task dependent? *Journal of Memory and Language, 28,* 127–154.
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.
Vandorpe, S., De Rammelaere, S., & Vandierendonck, A. (2005). The odd–even effect in addition: An analysis per problem type. *Experimental Psychology, 52,* 47–54.
Verguts, T., & Fias, W. (2008). Symbolic and nonsymbolic pathways of number processing. *Philosophical Psychology, 21,* 539–554.
Waters, G. S., & Caplan, D. (2003). The reliability and stability of verbal working memory measures. *Behavior Research Methods, Instruments, & Computers, 35,* 550–564.
Willmes, K., & Iversen, W. (1995, April). *On the internal representation of number parity.* Paper presented at the spring meeting of the British Neuropsychological Society, London.