

Cognitive Reflection and Theory of Mind of Go Players

Marc Oliver Rieger¹ and Mei Wang²

¹ University of Trier

² WHU – Otto Beisheim School of Management

ABSTRACT

KEYWORDS

cognitive reflection test
theory of mind
patience
board games
Go
weiqi
baduk
chess

Go is a classical Chinese mind game and a highly popular intellectual pursuit in East Asia. In a survey at two Go tournaments (one of them the largest in Europe), we measured cognitive reflection and decision in strategic games (using the classical “beauty contest” game) (N = 327). We found that Go players in our survey had outstanding average cognitive reflection test (CRT) scores: 2.51 among all participants and 2.80 among players of high master level (dan). This value easily outperforms previous measurements, for example, of undergraduates at top universities. The CRT score was closely related to the playing strength, but not to the frequency of playing. On the other hand, frequent players tended to have higher theory of mind, regardless of their playing strengths. However, self-reported patience was not statistically significantly correlated with Go strength or playing frequency.

INTRODUCTION

Human behavior is influenced by cognitive ability. Traditionally, these abilities have been summarized in one dimension, intelligence, but subsequent research focused on finding more and more facets of intelligence that are important for specific tasks. In particular, researchers found that performance in computer games (Bonny & Castaneda, 2017; Bonny et al., 2016) and chess (Grabner, 2014a) can be related to different cognitive skills. In this study, we concentrated on two of these facets: cognitive reflection and theory of mind (or mentalizing).

The cognitive reflection test (CRT) is designed to test the tendency to resist intuitive incorrect answers through reflective reasoning (Frederick, 2005). The CRT consists of three questions, each of which has an intuitive but wrong answer, and the correct answers require less impulsive and more reflective thinking. It captures not only thinking ability, but also personal thinking style and characteristics.

The theory of mind is the ability to think from other people’s perspective and predicting their intentions and actions. It is critical for social interaction and learning. This ability is also important when

playing strategic games that require adopting the other person’s point of view, for example, the famous “beauty contest game” or “Keynesian beauty contest” (Nagel, 1995).¹ The name of this game is inspired by comments from the brilliant economist John Maynard Keynes on the similarity of professional investment and a newspaper competition where the prize is awarded to the competitor who chooses the six most beautiful faces from a hundred photographs that are closest to the average preference of all competitors. Keynes (1936, p.156) noticed that “(i) t is not a case of choosing those which are really the prettiest, nor even those which average opinion genuinely thinks the prettiest. We have reached the third degree to anticipating what average opinion expects the average opinion to be. And there are some, I believe, who practice

Corresponding author: Marc Oliver Rieger, University of Trier, Chair of Banking and Finance, 54286 Trier, Germany
Email: mrieger@uni-trier.de

the fourth, fifth, and higher degrees.” This insight can be formally represented by the cognitive hierarchy model which captures the level of reasoning (Camerer et al., 2004). The low-level reasoners do not expect others to behave strategically, whereas high-level reasoners take into account the strategy of others.

The experimental beauty contest game, analogous to the newspaper competition described by Keynes, is a number guessing game. Instead of guessing the most beautiful face chosen by all competitors, each participant is asked to choose a number between 0 and 100 and the winner is the one who chose a number that is closest to 2/3 times the average of all chosen numbers.

The unique Nash equilibrium of this game is zero, which can be derived by eliminating the dominated options iteratively. However, the winning numbers that emerge from experimental and survey studies never reach this rational equilibrium. The reason is that the participants differ in their level of reasoning, so that the winner is one who is good at predicting average responses by other players rather than one who follows the rational equilibrium. The neuroimaging experiments by Coricelli and Nagel (2009) provide clear evidence that a higher level of reasoning in the game is associated with different brain areas when compared to simpler mental models. Higher level answers engage the medial and the ventral prefrontal cortex more while low level reasoning involves the rostral anterior cingulate cortex. In a literature review of neuroimaging studies, Carrington and Bailey (2009) suggest that theory of mind reasoning may activate several distinct brain regions, forming an integrated functional “network,” which is composed of “core” regions (e.g., parts of the prefrontal cortex and superior temporal sulcus) and several “peripheral” regions. In addition, studies also found that subjects with higher cognitive ability give answers closer to the Nash equilibrium in the beauty contest game (Branas-Garza & Teresa Garcia-Munoz, 2012; Burnham et al., 2009). In particular, Branas-Garza and Teresa Garcia-Munoz (2012) found that CRT score predicts better performance on the beauty contest game whereas the Raven test score does not.

Patience, or more precisely, time discounting, has been studied theoretically and empirically for a long time in economics. Recently, the connection between patience and cognitive abilities has been investigated in several studies. Studies on judgment and decision-making reveal two distinctive cognitive processes: “effortless intuition” (System 1) versus “deliberate reasoning” (System 2; Kahneman, 2003; Stanovich & West, 2000). Delay of gratification requires the cognitive ability to resist impulsive instinct and engage in more deliberate thinking. Empirical evidence shows that high cognitive ability is associated with increased patience, consistent with the hypotheses based on the two-system view (Benjamin et al., 2013; Burks et al., 2009; Dohmen et al., 2010; Frederick, 2005). In particular, Frederick (2005) demonstrated that CRT scores have high predictive power for patience. The simplicity and reliability of the three-question CRT makes it an ideal instrument to proxy reflective thinking.

To some extent, future-oriented choice also requires the cognitive ability to imagine future mental states of a different self, an ability that could be associated with theory of mind (Jamison & Wegener, 2010).

Using false-belief task to test theory of mind, researchers found that children with better in theory of mind are more willing to wait longer for larger outcomes (Marchetti et al., 2014; Moore et al., 1998).

It is also interesting that although older children are on average more likely to wait, the age effect disappears after theory of mind is controlled for, implying that individual differences in theory of mind have a stronger influence on patience than the direct effect of the developmental stage.

While the limits of our cognitive abilities are an obvious problem when trying to perform intellectually challenging tasks, the experience of reaching this limit and at least partially mastering such challenging tasks is considered by many as rewarding and stimulating. This explains the relative popularity of spare time activities that involve large amounts of cognitive effort. Examples involve such diverse activities as studying foreign languages, solving sudoku, or playing chess. One such activity that requires particularly intense cognitive efforts is the mind game of Go, known also under the names weiqi (in Chinese) and baduk (in Korean). Go is probably the oldest board game that is still played and its origin dates back around 3000 years (Potter, 1984). It has been invented in China and later spread to Korea and Japan. It came to Europe, America, and the rest of the world only in the 19th century. Estimates for the number of Go players worldwide have been set between 26 (Carlisle, 2009) and 60 million (Britgo, 2016), but the true number is likely much higher, as became clear when in March 2016, more than 100 million people watched the match between AlphaGo and Lee Sedol, see below (Choe, 2016). This number can therefore be seen as a lower bound for the actual number of Go players worldwide. Most players still can be found in East Asia, and most professional players also live in East Asia where Go is considered to be a part of traditional culture, similar to calligraphy, traditional music, or paintings.²

Similar to chess, Go is an abstract strategy board game for two players. The standard Go board has 19 × 19 grid of lines, containing 361 points. The two players (black and white) take turns placing stones on the intersections of the board, with the aim to surround more territory than the opponent. While in Europe and America chess is still much better known than Go, Go is widely considered to be far more complex in its gameplay, although its rules are relatively straightforward. The reason for this is the much larger number of possible game variations and the difficulty to evaluate a given position. For this reason, computer programs have for a long time been notoriously incompetent at Go: the world champion in chess, Gary Kasparov, has been defeated by the IBM program Deep Blue that used brute force computation methods already in 1996, but the first Go professional has been defeated by a computer program only in 2015. This program, AlphaGo, was developed by the DeepMind subsidiary of Google and used a neural network approach that tried to mimic human thought processes in a highly sophisticated way (Silver et al., 2016). AlphaGo has defeated Lee Sedol, a Korean Go world champion, in a match that attracted widespread media attention.

Given that Go is therefore not only one of the most intellectually challenging human activities, but also seems to require human-like

cognitive processes to be played well, Go players are suitable to be tested for their cognitive skills.

Indeed, there are two possible effects that can be expected: on the one hand, Go players might be, by self-selection, different from the average population. To master the game, special cognitive abilities might be required. On the other hand, playing Go might help to develop certain cognitive skills. We know that the brain, even for adults, is flexible, and training effects have been found for certain cognitive skills, for example, spatial orientation (Maguire et al., 2000), working memory performance and function (Söderqvist et al., 2012), and so forth. There are also previous studies on children suggesting that studying Go may improve certain cognitive skills and even facilitate structural brain changes (Jung et al. 2014; Lee et al., 2010). Professional training, at least in China and Korea, often replaces general education which could explain why outcomes of IQ tests of professional Go players in Lee et al. (2010) were slightly below average.

In this paper, we try to find such effects on the aforementioned three specific aspects (cognitive reflection, theory of mind, and patience) in a survey experiment conducted at two Go tournaments in Europe: the annual European Go Congress (EGC) 2015 in Liberec (Czech Republic) and the Mannheimer Aji 2015 in Mannheim (Germany). The two-weeks long EGC is Europe's largest Go tournament (see Figure 1). Including side tournaments, nearly one thousand people participated in this event. On the other hand, the annual Mannheimer Aji, is a small weekend tournament with predominantly German participants. In the year of our study (2015), less than fifty players participated.

We found evidence that cognitive reflection helps to perform better in Go and that, on the other hand, frequent Go players have better theory of mind. However, patience was not related to Go strength or playing frequency.

The paper is structured as follows: first, we describe the survey design and the procedure of the experiment. Then, we present the empirical results. Finally, we summarize the results and discuss potential extensions.

METHODOLOGY

We conducted our survey at the EGC and the Mannheim Aji Go tournaments. For a few participants who took part at both tournaments, data from the second measurement were excluded from the analysis. At both tournaments, the survey was conducted at the beginning of a round. A verbal announcement was made explaining the procedure and that players should fill in the questionnaires independently and truthfully. The questionnaires had been put on the tournament tables before the start of the round and collected during and after the subsequent games. The survey was conducted in collaboration with the European Go Federation and included—besides the questions of primarily academic interest—additional questions on when and where participants had learned Go, what they like most about it, and so forth. This had the side benefit of increasing the participants' motivation and adding credibility to the survey.

The survey was bilingual (English and German) because most participants spoke English, and German was the native language for the highest proportion of participants. As an additional incentive for participation, a total of €200 cash payment was awarded as prize to the winners of the beauty contest question, with the prerequisite that all other questions had to be answered as well. This led to comparably few missing items: on average, only 8.6% in the three main tasks (CRT, patience, and beauty contest).



FIGURE 1.

The annual European Go Congress (EGC) is the largest Go tournament in Europe with several hundred participants. The survey was conducted on participants of this tournament and a smaller one.

We would like to know how good you are in quizzes besides go. Please answer the following three quiz questions:

Wir möchten wissen, wie gut Du in Denkaufgaben außerhalb Go bist. Bitte beantworte die folgenden drei Fragen:

(1) A pen and an eraser cost 1.10€ in total. The pen costs 1.00€ more than the eraser. How much does the eraser cost? _____ cents

(2) If it takes 5 machines 5 minutes to make 5 widgets, how long would it take 100 machines to make 100 widgets? _____ minutes

(3) In a lake, there is a patch of lily pads. Every day, the patch doubles in size. If it takes 48 days for the patch to cover the entire lake, how long would it take for the patch to cover half of the lake? _____ days

(1) Ein Stift und ein Radiergummi kosten zusammen 1,10€.

Der Stift kostet 1,00€ mehr als der Radiergummi.

Wie viel kostet der Radiergummi? _____ cents

(2) Wenn 5 Maschinen für die Produktion von 5 Produkten 5 Minuten brauchen, wie lange brauchen 100 Maschinen für die Produktion von 100 Produkten? _____ Minuten

(3) In einem See gibt es einen Teppich von Seerosenblättern. Der Teppich verdoppelt seine Größe jeden Tag. Wenn es 48 Tage dauert, bis der See vollständig bedeckt ist, wie lange würde es dauern, um den halben See zu bedecken? _____ Tage

FIGURE 2.

Contents of the survey.

The main questions used for our survey are shown in Figure 2.

With these questions, the cognitive reflection score of the participants was measured. The questions were taken more or less verbatim from the original study by Frederick (2005). We took the liberty to change the “bat and ball” question slightly, since outside the US, baseball is uncommon, and pretests showed that Europeans often struggled with the word “bat”, since they were rather thinking of the flying mammal than the sport instrument, causing unnecessary confusion. For specialists in this field, the CRT is well-known and hence there might be concerns that subjects had already taken the CRT at other occasions. To mitigate this concern, we collected feedback from some of the participants. Moreover, Meyer et al. (2018) showed that repeating the task does not significantly diminish performance differences. No participants recalled having taken the CRT before.

For each correct answer, one point was given, so that a total score between zero and three could be reached.

Next, we provided the formulation of the theory of mind measure, the standard beauty contest question, as explained in the previous section (see Figure 3). Six participants came closest to the winning number, thus the prize of €200 was divided between them.

Patience was measured with a question taken from the socio-economic panel (SOEP). The exact wording of this questions was: “Are you generally an impatient person, or someone who always shows great patience?” The answers are coded on an 11-point Likert scale (0-10), with 0 referring to very impatient and 10 to very patient. This measurement instrument has been validated by Vischer et al. (2013) as a simple, but reliable method to elicit patience.

To measure Go playing frequency of the participants, we combined the answers to three Likert scale questions for the frequency of playing Go (in a Go club, with friends, and online) into one variable, each with the possible answers nearly every day, around once per week, around once per month, and less or never (see Figure 4). We defined a value of 3 for all participants who stated that they played Go nearly every day (at a Go club, with friends, or online), 2 for all participants who stated that they played around once per week (but not more) in any given context, 1 for all participants who stated that they played around once per month, and 0 for all who played less frequently or never:

The complete questionnaire and codes can be provided by the authors on request.

Using the names of the participants that they had to fill in for the beauty contest competition, we could find more information about them using the European Go Database (www.europeangodatabase.org).

This publicly available database records all tournament games ever played in Europe since 1996, including more than 10000 tournaments, and tracks all participating Go players, calculates their Go rating score (GoR) as a measure of their playing strength, and provides other useful statistics.

For our study, we collected the GoR³ of all survey participants who stated their name. We also recorded their country of residence, the date of their first tournament participation,⁴ their number of tournaments played (since 1996), and so forth.

RESULTS

Descriptive Results

A total of 327 participants took part in our survey ($N = 296$ at the EGC and $N = 31$ in Mannheim), 75.5% of them were male and 17.7% were female. The average age was 33.5 years ($SD = 16.3$), and 85.9% were adults (i.e., 18 years and older). The age distribution can be seen in Figure 5. Roughly half of the subjects (44.7%) were from a country that had either German or English as native language (the two languages in which the survey was conducted).

The playing strength of the participants who provided their names was taken from the European Go Database and had an average of 1691 points, corresponding to a level of 4 kyu. That is, most participants were weaker than master level (dan). This is representative to the

The next question is deciding about which participant will win the prize of 200€! / Die nächste Frage entscheidet darüber, welcher Teilnehmer den Preis von 200€ gewinnt!

Below please write down a number between 0 and 100. The average of all numbers given by the participants will be multiplied with 2/3. The participant, whose number is closest to this value, wins. (In case of several winners, the 200 Euro are split.)

So please guess, what is 2/3 of the average of all answers!

Bitte schreibe unten eine Zahl zwischen 0 und 100 auf. Der Durchschnitt aller Antworten wird mit 2/3 multipliziert. Derjenige Teilnehmer, dessen Antwort am nächsten an diesem Wert liegt, gewinnt. (Falls es mehrere Gewinner gibt, wird das Preisgeld aufgeteilt.)

Schätze also, was 2/3 des Durchschnitts aller Antworten sein wird!

Your number (0-100) / Deine Zahl (0-100):

To win a prize, please state your full name:

Um einen Preis zu gewinnen, gebe bitte Deinen Vor- und Nachnamen an:

FIGURE 3.

Beauty contest question.

How often do you play go... / Wie oft spielst Du Go...	Nearly every day Fast jeden Tag	Around once per week Etwa 1x pro Woche	Around once per month Etwa 1x pro Monat	Less frequent or never Seltener oder nie
...at a go club? / ...bei einem Go-Spieletreff?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
...elsewhere with friends or relatives? ...anderswo mit Freunden oder Verwandten?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
...online?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

FIGURE 4.

Go playing questions.

average population of Go players in Europe that participate in ranked tournaments. Of course, there are many more people who know the rules of Go, might play a few games with friends and family, but never participate in tournaments. Neither the number of such hobby players nor their playing strength are known, but their ranks typically start with 30 kyu. This was also the lowest GoR level in our survey. The strongest participant in our survey had a GoR of 2704, which corresponds to a (low) professional level (see the last panel of Figure 5).

Correlation Analysis

The CRT, theory of mind, and patience are well-known to be inter-related (see the Introduction section). Table 1 shows the correlation of these variables. It shows that cognitive reflection is significantly correlated to theory of mind, consistent with the results by Branas-Garza & Teresa Garcia-Munoz (2012).

However, we did not find statistically significant correlation between patience and theory of mind.⁵ A further surprising result was that participants with higher CRT scores perceived themselves to be less patient. One reason might be that we measured patience with a self-reported measure, different from the monetary payoff questions in Frederick (2005). The verbal questions reported by Frederick (2005) also gave mixed results. For example, CRT is not related to the tendency to procrastinate, and people with higher CRT think less about the future. Moreover, Frederick (2005) also found no correlation between CRT and time preferences for a longer time horizon. These results raise the question to what extent CRT is related to different dimensions on time preference, for example, the general time preference questions used in the current study and by Frederick (2005) or the time discounting questions with specific payoffs in the short versus long horizon (see Frederick et al., 2002, and Wang et al., 2016, for further discussion on different characteristics of time preferences).

We now relate the above three measures to Go playing strength (GoR rating) and frequency. We found statistically significant correlations of playing strength with cognitive reflection ($r = .22, p < .01$). This was not surprising, as it emphasizes that one needs to be less impulsive in thinking (i.e., having a high cognitive reflection) to play Go well. Further analysis showed that the correlation was $.22$ ($p < .01$) for the top 50% of the players, and $.15$ ($p < .01$) for the bottom 50%, so the correlation was not weaker for strong players.⁶

We also found statistically significant correlations between Go playing frequency, theory of mind, and patience. The most straightforward explanation for this relation is that playing Go improves these skills. If, on the other hand, these skills would merely make people better Go players and, therefore, lead through self-selection to higher skills for frequent players, we would also expect to see a significant

correlation between playing strength and these cognitive skills, which indeed was the case for theory of mind (see Table 1). Additionally, the lack of correlation between playing strength and playing frequency is less surprising than it might seem at first glance: Playing strength can be the result of studying Go intensively without playing often, or the player might have played Go frequently in the past.

In the following subsections, we present these relations in more detail.

Cognitive Reflection

The first remarkable result of our survey was the average CRT score. The scores for the general population are usually below 0.7 (e.g., de Mel, 2010) and the test has been described as “demanding” (Kahan, 2013). Finance professionals score better (1.87, Kirchler et al., 2016) and students of top American universities score around 2 on average (e.g., the average score of MIT students in Frederick, 2005, was 2.18). The mean score among Go players was substantially higher, 2.51 ($SD = 0.05$), even higher than for expert chess players (2.08; Campitelli & Labollita, 2010). To the best of our knowledge, this is larger than all previously measured values for subject groups. Figure 6 summarizes selected scores.⁷

The CRT score among Go players was highest for the strongest Go player (2.80), as was expected given the results of the previous section. Table 2 shows the results for groups of different playing strength: fairly weak players (up to a GoR of 1150 or, in traditional ranks, 10 kyu) that are usually referred to as “double digit kyu” (DDK), average players (up to GoR of 2050 or 1 kyu, called single digit kyu, SDK), low master players (up to GoR 2350 or 3 dan) and high master players (above that level). In the latter category, professionals were also included.

While all of this strongly suggests a relation between cognitive reflection and Go playing strength, this might be a spurious effect of other characteristics. We therefore conduct a regression analysis controlling for various other variables, in particular demographics, Go playing frequency, whether the subject was from a country where one

TABLE 1.

Correlations Between Measured Cognitive Skills, Playing Strength, and Playing Frequency, $N = 246$

	Theory of mind	Patience	Go playing strength	Go playing frequency
CRT score	0.30**	0.08	0.22**	0.01
Theory of mind		-0.05	0.10	0.12*
Patience			0.05	-0.09
Go playing strength				-0.09

* $p < .05$; ** $p < .001$

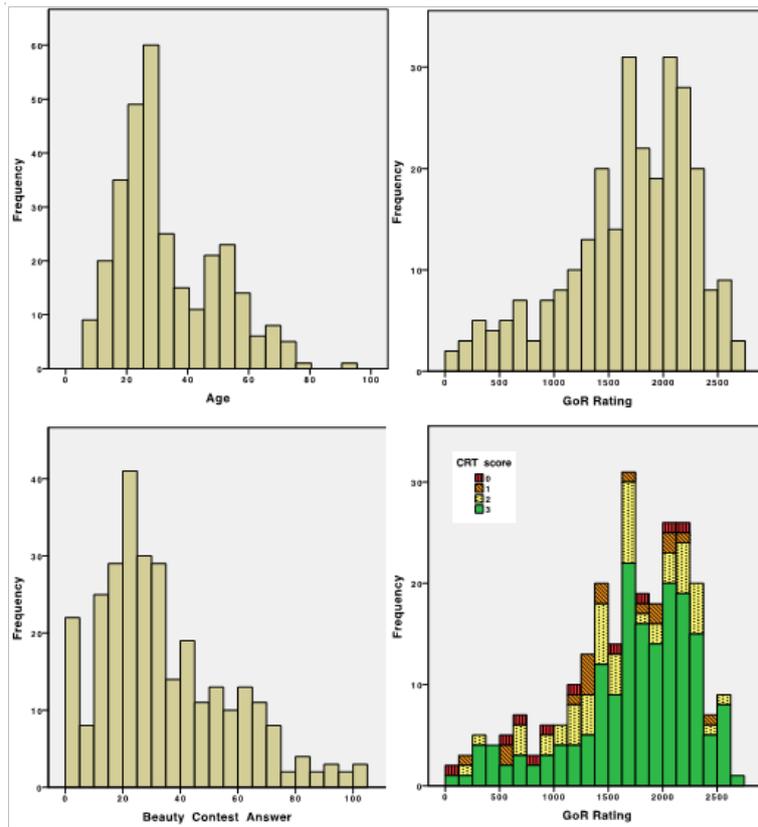


FIGURE 5. Distribution of age, playing strength (GoR), and answers to the beauty contest game among the participants; relation between playing strength and CRT score.

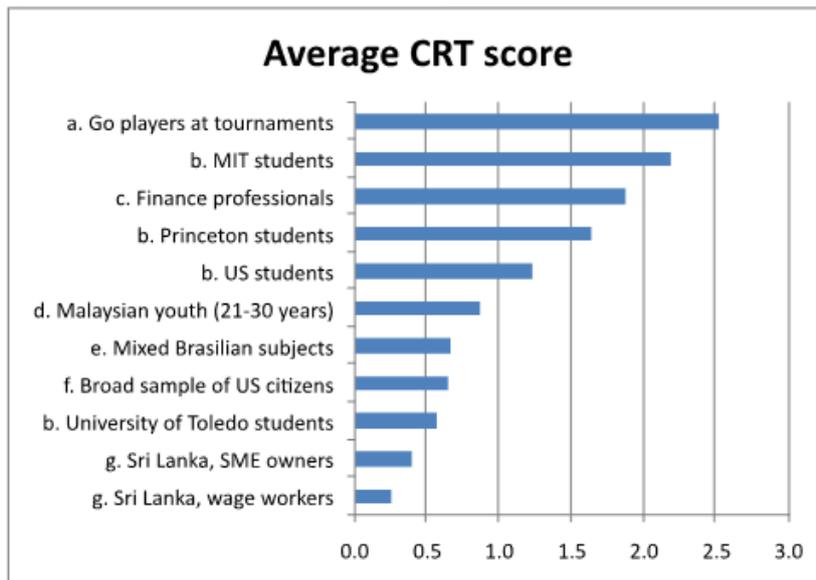


FIGURE 6. Cognitive reflection scores of subject groups from selected previous studies in comparison with the level measured for Go players.

TABLE 2.

Cognitive reflection for different skill levels in Go: from the beginner (20 kyu) to the highest amateur (7 dan) and professional level.

Playing strength	CRT score					μ	N
	0	1	2	3			
20 kyu - 10 kyu (DDK)	14%	7%	21%	57%	2.21	45	
9 kyu - 1 kyu (SDK)	2%	8%	22%	68%	2.56	144	
1 dan - 3 dan (low dan)	2%	5%	22%	71%	2.62	60	
4 dan - 7 dan (high dan) and professional	0%	5%	10%	85%	2.8	23	

of the survey languages is spoken and what tournament he or she took part at.

The results confirm the previous findings: There is a significant relation between cognitive reflection and Go playing strength, but none to playing frequency. In summary, this supports the idea that in order to excel in Go, an exceptionally large degree of cognitive reflection is needed.

Theory of Mind

We now turn our attention to the second cognitive skill, theory of mind, measured in our study by the classic beauty contest question with answers in the interval [0, 100] and the winning number being two thirds of the average of all answers.

The average of the answers was 34.3 ($SD = 22.5$, $SE = 1.3$). The winning number was therefore 22.9. Since the number of participants was large and the majority chose integer numbers as answer, it was natural that more than one person won. Indeed, six participants chose 23 and thus came closest to the winning number and shared the prize of €200. Figure 5, left-bottom panel shows the distribution of the answers.

TABLE 3.

Ordinary least squares regressions on the CRT score, $N = 246$.

	CRT score		
	Model 1	Model 2	Model 3
Go playing frequency	0.042 (0.670)	0.009 (0.132)	0.021 (0.319)
GoR rating	0.212** (3.381)	0.145* (2.207)	0.160* (2.437)
Gender		0.102 (1.548)	0.095 (1.439)
Native language dummy			0.105 (1.544)
EGC dummy			-0.042 (-0.628)
Age		0.713** (2.609)	0.699* (2.588)
Age × Age		-0.644* (-2.383)	-0.653** (-2.440)
Patience quartile			0.136* (2.196)
Adjusted R ² in %	3.762	6.772	8.981

* $p < .05$; ** $p < .001$

Figure 7 shows that the target number for the Go players in the current study was in line with other studies and that is was very similar to the average number for chess players (Bühren & Frank, 2012). Lower numbers have typically been found in newspaper surveys where participants had more time to reflect on their decision, and at the same time, the participation rate was much lower, which implies that people who were not interested in such task and otherwise would have given a more or less random (and therefore relatively high) number did not participate (Bosch-Domenech et al., 2002; Schou, 2005; Selten & Nagel, 1998; Thaler, 1997, 2015). Student samples gave similar, but usually slightly larger numbers than our Go players (Belot et al., 2015; Nagel, 1995; Rubinstein, 2007). The nonstudent sample from Belot et al. (2015) gave by far the highest number among these studies.

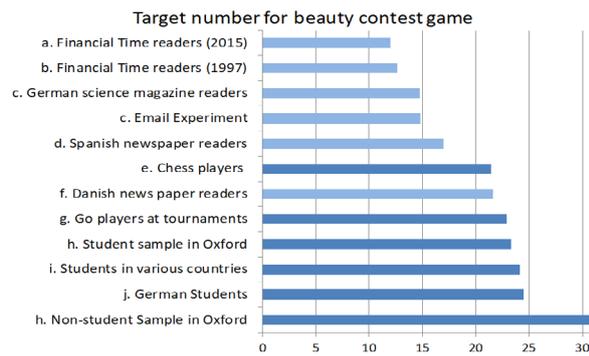
We defined the variable of theory of mind as 100 minus the answer given to the beauty contest question. We were interested in the question whether theory of mind is related to Go playing strength and frequency. Therefore, we conducted an ordinary least squares regression with these variables, adding the same controls as shown in Table 3. The results (see Table 4) confirmed the previous correlation analysis: There is a clear relation between theory of mind and Go playing frequency but not playing strength. We also found a statistically significant relation with cognitive reflection. The predictive power of this (and the following) regression, as expressed by the adjusted R², was fairly low, but it is to be expected that playing Go has a smaller impact on Theory of Mind than many other individual differences that we did not measure in our study.

In summary, the most likely interpretation of these findings are two simultaneous effects:

- theory of mind is enhanced by a high degree of cognitive reflection,
- playing Go frequently may improve theory of mind.

Causality usually cannot be identified in a regression model, but we propose at least some evidence in favor of a causal relation:

Suppose that a higher degree of theory of mind would lead to playing more Go. This might be the case if it increases playing strength, and hence, the wish to experience more success. However, this is ruled out by the lack of relation between playing strength and theory of mind. Alternatively, it might be that high theory of mind increases the fun of playing Go and thus the motivation to play, but the participants of the European Go Congress were highly motivated to play Go, given that they travelled long distances and spent a significant time of their vacation to participate in the event. Moreover, we did not find any statistically significant relation between the frequency of tournament participation and theory of mind, as could be expected if motivational factors were involved.⁸ Therefore, it seems more likely that differences in Go playing frequency (outside of the tournament) can be attributed to nonmotivational factors like differences in workload, family situation, competing hobbies, and so forth. For these reasons, it seems less plausible that high theory of mind increases Go playing frequency, but all of these arguments would also apply if an unknown third factor resulted both in increased theory of mind and in higher Go playing

**FIGURE 7.**

Average answers to the classic beauty contest game in selected previous studies in comparison with the current results of Go players. (Studies with unlimited thinking time, like newspaper games, are marked in light color.)

frequency. Therefore, we consider the results to at least indicate a causal impact of Go playing frequency on theory of mind.

Since the answers to the beauty contest had large skewness, we conducted the same analysis for the logarithm of the answers as a robustness test. The results confirmed the previous analysis (see Table 5).

Patience

The observed distribution of answers to the question measuring patience was highly similar to the patterns found by Vischer et al. (2013) for the SOEP sample (see Figure 8). We found a slightly higher level of patience ($\mu = 6.2$) than in the SOEP sample ($\mu = 6.1$), but since our sample was far from representative, this was not surprising. However, there are two reasons why we expected to see a lower average for our sample. First, patience is particularly high for Germans (Wang et al., 2016) and the SOEP is a German data set, but most participants at the EGC were not German. Second, adults are more patient than children, but in our sample were many children who participated in the tournament. When considering only the adult German participants ($n = 93$), the average value indeed increased to $\mu = 6.3$.

Like for theory of mind, this suggests that frequent Go players develop a higher patience on average, since other possible explanations for this effect are not supported by evidence. In particular, there was no significant relation between frequency of tournament participation and patience.⁹ Therefore, it is again most likely that external constraints prevent some of the players from playing more frequently, but motivational factors are not relevant in this sample. Thus, the most parsimonious explanation is to assume that playing Go frequently enhances average patience.

CONCLUSION

Go is sometimes considered to be the most complex of all classic mind games, which makes Go players an interesting subject pool for studying cognitive skills. Compared to the numerous studies on

the cognitive ability of chess players (see a recent meta-analysis by Burgoyne et al., 2016), the cognitive characteristics of Go players are under-researched. Our empirical study of tournament Go players showed that there is a robust relation between Go playing strength and cognitive reflection for both strong and weaker players. This is in line with recent studies on intelligence and games of strategy, such as chess (Grabner, 2014a; Grabner, 2014b) and videogames (Aung et al., 2018; Bonny et al., 2020; Kokkinakis et al., 2017; Li et al., 2020), but less consistent with the line of research claiming that “individual differences in more ‘basic’ cognitive processes (e.g., intelligence, memory, capacity, and perceptual functioning) have not, to date, been predictive of attained level of skilled performance” (Ericsson & Ward, 2007, p. 348, see Ackerman, 2014, for a review on individual differences in talent and expert performance).

Note that the CRT, although composed of three mathematical questions, does not only measure mathematical ability and intelligence, but also the ability to resist an intuitive response and the disposition to actively open-minded thinking (Campitelli & Gerrans, 2014; Campitelli & Labollita, 2010). Toplak et al. (2011) also found that it is a unique predictor of performance on heuristic-and-biases tasks. Our analysis suggests that a high degree of cognitive reflection helps to achieve a high level of Go playing strength. This explains why the average level of cognitive reflection is the highest recorded so far in a survey, easily surpassing data collected among students from elite universities (Frederick, 2005), as well as expert chess players (Campitelli & Labollita, 2010). Studies on brain imaging also show that Go playing activates brain areas related to problem solving (Chen et al., 2003). Therefore, further studies may disentangle the training and

TABLE 4.

Ordinary least squares regressions on theory of mind (defined as the answer to the beauty contest question subtracted from 100), $N = 246$.

	Theory of mind		
	Model 1	Model 2	Model 3
Go playing frequency	0.199** (3.198)	0.207** (3.166)	0.204** (3.228)
GoR rating	0.101 (1.628)	0.074 (1.107)	0.037 (0.568)
Gender		0.079 (1.173)	0.037 (0.567)
Native language dummy			-0.005 (-0.075)
EGC dummy			-0.073 (-1.103)
Age		0.061 (0.223)	-0.171 (-0.636)
Age × Age		-0.068 (-0.250)	0.150 (0.564)
CRT score			0.296** (4.667)
Adjusted R ² in %	3.919	3.334	11.094

* $p < .05$; ** $p < .001$

TABLE 5.

Robustness test: regressions on the logarithm of the answer to the beauty contest question, $N = 246$.

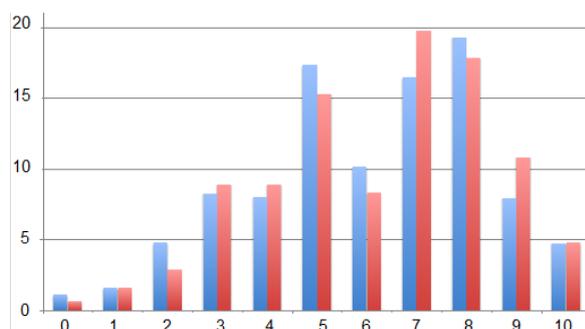
	log(beauty contest)		
	Model 1	Model 2	Model 3
Go playing frequency	-0.143* (-2.273)	-0.140* (-2.119)	-0.138* (-2.131)
GoR rating	-0.100 (-1.591)	-0.076 (-1.128)	-0.039 (-0.585)
Gender		-0.050 (-0.738)	-0.025 (-0.373)
Native language dummy			-0.031 (-0.451)
EGC dummy			-0.014 (-0.211)
Age		-0.162 (-0.583)	0.030 (0.110)
Age × Age		0.150 (0.546)	-0.017 (-0.062)
CRT score			-0.251** (-3.846)
Adjusted R ² in %	2.041	1.253	6.079

* $p < .05$; ** $p < .001$

self-selection effects. The relationship between cognitive reflection and Go playing can be in both directions.

We also found a relation between theory of mind (i.e., the ability to anticipating others' thoughts or intentions, as measured by the classical beauty contest guessing game) and Go playing frequency. However, playing strength had no relation with performance on the beauty contest game, consistent with a previous study on chess players by Bühren and Frank (2012). Our finding is also consistent with the results by Jung et al. (2013), who observed a difference between long-term and novice Go players concerning the connectivity within the default-mode network, which is thought to be related to theory of mind. They suggest that long-term Go playing may change this part of brain structure because it requires the capacity to infer the opponent's intentions when playing Go.

Previous studies on Go players showed that Go training also changes brain areas that are related to self-control, working memory,

**FIGURE 8.**

Comparison of patience between the SOEP data in Vischer et al. (2013) on the left and the current study on the right.

and attention (Lee et al., 2010; Kim et al., 2014). Further neuroimaging research demonstrates that individual differences in time discounting can be caused by differences in working memory (Shamosh & Gray, 2008). A controlled experimental study conducted by Kim et al. (2014) showed that, after only 16 weeks of Go training, children with ADHD had significantly improved executive functions, which are associated with planning, cognitive persistence, and working memory. They further suggest that Go training changes the activity of the prefrontal cortex, which is important for regulation of impulses, controlling, and decision making. This suggests that Go playing may improve patience. However, we did not find a statistically significant relationship between Go playing frequency/strength and patience. It is not clear to what extent such lack of relationship is caused by response errors in the self-reported measurement of patience and playing frequency. Therefore, we encourage future studies to more directly test the effects of game playing (e.g. Go, chess, other board games, and video games) on patience and time discounting.

Our study was based on a relatively large, representative sample of tournament Go players, which allows us to control the effects of age, experience, and other factors (Grabner, 2014b). A slight limitation of our study is that it did not include top level professionals. It would be interesting to see whether the relation between cognitive reflection and playing strength still holds on the highest level. However, it is challenging to find a sufficiently large number of these elite players who would be willing to participate in such a study.

In summary, Go players show remarkable characteristics regarding cognitive reflection. Moreover, Go seems to have positive effects on certain important cognitive characters, such as theory of mind and patience. Given that Go playing involves much richer dimensions of brain activity than chess, we hope that these results can encourage further studies on the relation between playing Go and various aspects of cognition.

FOOTNOTES

¹ Researchers have used different tests to measure theory of mind, such as “reading the mind in the eyes” test (Engel et al., 2014), and “referential communication task” (Dumontheil et al., 2010). We chose the “beauty contest” game because of its more crowd-tapping and strategic potential whereas the other tests may focus more on the individual and their respective emotions. Moreover, other tests are lengthier and harder to set up and they may appear as relatively easy, which leads to a ceiling effect. We thank an anonymous referee for this insight.

² All four are summarized as the four arts “Qin Qi Shu Hua” in classic Chinese education since at least the Tang dynasty.

³ The minimum GoR is 100, professionals' GoR is typically above 2700.

⁴ In case this was before 1996, we used their first recorded tournament as data point. This might affect a few players that were active before 1996, but the first recorded tournament was in 1996 for only 35 participants, this did not have a large effect on the data.

⁵ Given that the distribution of patience was highly skewed (indeed, the Kolmogorov-Smirnov test of normality gave a p value below 0.1%), in further analyses, we used nonparametric (Spearman's) correlations or

transformed the variable to quartiles. In our sample, the first quartile (the most impatient subjects) ranged from 0 to 5 (38%), the second ranged from 6 to 7 (28%), the third was 8 (18%), and the fourth ranged from 9 to 10 (16%).

⁶ This is different from the findings of Burgoyne et al. (2016) on their (related) test of numerical skills, where the correlation became substantially weaker for strong players, a phenomenon referred as range restriction (Detterman, 2014; Deary et al., 2010).

⁷ The value for Go players is statistically significantly higher than that of the second best group, the MIT students (p -value of t -test: 1.4%).

⁸ The correlation between the number of tournaments that a player has participated per year (since first time playing in a Go tournament) and theory of mind was insignificant ($\rho = -0.026$, $p = .677$). Same for his total number of tournaments ($\rho = 0.047$, $p = .442$). Regression analysis did not lead to any other result.

⁹ Correlation with tournaments played per year: $\rho = -0.049$, $p = .429$, correlation with tournaments played in total: $\rho = -0.03$, $p = .630$.

REFERENCES

- Ackerman, P. L. (2013). Nonsense, common sense, and science of expert performance: Talent and individual differences. *Intelligence*, 45, 6–17. <https://doi.org/10.1016/j.intell.2013.04.009>
- Albaity, M., Rahman, M. & Shahidul, I. (2014). Cognitive reflection test and behavioral biases in Malaysia. *Judgment and Decision Making*, 9, 149–151.
- Aung, M., Bonometti, V., Drachen, A., Cowling, P., Kokkinakis, A. V., Yoder, C., & Wade, A. (2018). Predicting skill learning in a large, longitudinal moba dataset. *2018 IEEE conference on computational intelligence and games (CIG)*, 1-7. <https://doi.org/10.1109/CIG.2018.8490431>.
- Belot, M., Duch, R., & Miller, L. (2015). A comprehensive comparison of students and non-students in classic experimental games. *Journal of Economic Behavior & Organization* 113, 26–33. <https://doi.org/10.1016/j.jebo.2015.02.007>
- Benjamin, D. J., Brown, S. A., & Shapiro, J. M. (2013). Who is “behavioral”? Cognitive ability and anomalous preferences. *Journal of European Economic Association*, 11, 1231–1255. <https://doi.org/10.1111/jeea.12055>
- Bonny, J. W., & Castaneda, L. M. (2017). Number processing ability is connected to longitudinal changes in multiplayer online battle arena skill. *Computers in Human Behavior*, 66, 377–387. <https://doi.org/10.1016/j.chb.2016.10.005>
- Bonny, J. W., Castaneda, L. M., & Swanson, T. (2016). Using an international gaming tournament to study individual differences in MOBA expertise and cognitive skills. *Proceedings of the 2016 CHI conference on human factors in computing systems*, 3473–3484. <https://doi.org/10.1145/2858036.285819>
- Bonny, J. W., M. Scanlon, & L. M., Castaneda (2020). Variations in psychological factors and experience-dependent changes in team-based video game performance. *Intelligence*, 80, 101450.
- Bosch-Domenech, A., Montalvo, J. G., Nagel, R. & Satorra, A. (2002). One, two, (three), infinity...: Newspaper and lab beauty-contest games. *American Economic Review* 92, 1687–1701. <https://doi.org/10.1016/j.intell.2020.101450>
- Branas-Garza, P. & Teresa Garcia-Munoz, R. H. G. (2012). Cognitive effort in the beauty contest game. *Journal of Economic Behavior & Organization* 83, 254–260. <https://doi.org/10.1016/j.jebo.2012.05.018>
- Britgo (2016). Frequently asked questions about Go. <http://www.britgo.org/press/faq.html>
- Bühren, C., & Frank, B. (2012). Chess players’ performance beyond 64 squares: A case study on the limitations of cognitive abilities transfer. *Talent Development & Excellence*, 4, 157–169.
- Burgoyne, A. P., Sala, G., Gobet, F., Macnamara, B. N., Campitelli, G., & Hambrick, D. Z. (2016). The relationship between cognitive ability and chess skill: A comprehensive meta-analysis. *Intelligence* 59, 72–83. <https://doi.org/10.1016/j.intell.2016.08.002>
- Burks, S. V., Carpenter, J. P., Goette, L., & Rustichini, A. (2009). Cognitive skills affect economic preferences, strategic behavior, and job attachment. *Proceedings of National Academy of Sciences*, 106, 7745–7750. <https://doi.org/10.1073/pnas.0812360106>
- Burnham, T. C., Cesarini, D., Johannesson, M., Lichtenstein, P., & Wallace, B. (2009). Higher cognitive ability is associated with lower entries in a beauty contest. *Journal of Economic Behavior & Organization*, 72, 171–175. <https://doi.org/10.1016/j.jebo.2009.05.015>
- Camerer, C., Ho, T., & Chong, J. (2004). A cognitive hierarchy model of games. *Quarterly Journal of Economics*, 119, 861–898. <https://doi.org/10.1162/0033553041502225>
- Campitelli, G., & Gerrans, P. (2014). Does the cognitive reflection test measure cognitive reflection? A mathematical modeling approach. *Memory and Cognition*, 42, 434–447. <https://doi.org/10.3758/s13421-013-0367-9>
- Campitelli, G., & Labollita, M. (2010). Correlations of cognitive reflection with judgments and choices. *Judgment and Decision Making*, 5, 182–191.
- Carlisle, R. P. (Ed.). (2009). *Encyclopedia of play in today’s society* (Vol. 1). Sage.
- Carrington, S. J., & Bailey, A. J. (2009). Are there theory of mind regions in the brain? A review of the neuroimaging literature. *Human Brain Mapping*, 30, 2313–2335. <https://doi.org/10.1002/hbm.2067>
- Chen, X., Zhang, D., Zhang, X., Li, Z., Meng, X., He, S., & Hu, X. (2003). A functional MRI study of high-level cognition II. The game of GO. *Cognitive Brain Research* 16, 32–37. [https://doi.org/10.1016/S0926-6410\(02\)00206-9](https://doi.org/10.1016/S0926-6410(02)00206-9)
- Choe, S.-H. (2016, March 16). Machine beats man in go series, winning 4. *New York Times*.
- Coricelli, G., & Nagel, R. (2009). Neural correlates of depth of strategic reasoning in medial prefrontal cortex. *Proceedings of National Academy of Sciences*, 106, 9163–9168. <https://doi.org/10.1073/pnas.0807721106>
- de Mel, S., McKenzie, D., & Woodruff, C. (2010). Who are the micro-enterprise owners? Evidence from Sri Lanka on Tokman versus De Soto. In J. Lerner & A. Schoar (Eds.), *International differences in entrepreneurship* (pp. 63–87). University of Chicago Press.

- Deary, I. J., Penke, L., & Johnson, W. (2010). The neuroscience of human intelligence differences. *Nature Reviews Neuroscience*, *11*, 201–211. <https://doi.org/10.1038/nrn2793>
- Detterman, D. K. (2014). Introduction to the intelligence special issue on the development of expertise: Is ability necessary? *Intelligence*, *45*, 1–5. <https://doi.org/10.1016/j.intell.2014.02.004>
- Dohmen, T., Falk, A., Huffman, D., & Sunde, U. (2010). Are risk aversion and impatience related to cognitive ability? *American Economic Review*, *100*, 1238–1260. <https://doi.org/10.1257/aer.100.3.1238>
- Dumontheil, I., Apperly, I. A., & Blakemore, S.-J. (2010). Online usage of theory of mind continues to develop in late adolescence. *Developmental Science*, *13*, 331–338. <https://doi.org/10.1111/j.1467-7687.2009.00888.x>
- Engel, D., Woolley, A. W., Jing, L. X., Chabris, C. F., & Malone, T. W. (2014). Reading the mind in the eyes or reading between the lines? Theory of mind predicts collective intelligence equally well online and face-to-face. *PLoS One*, *9*, e115212. <https://doi.org/10.1371/journal.pone.0115212>
- Ericsson, K. A., & Ward, P. (2007). Capturing the naturally occurring superior performance of experts in the laboratory toward a science of expert and exceptional performance. *Current Directions in Psychological Science*, *16*, 346–350. <https://doi.org/10.1111/j.1467-8721.2007.00533.x>
- Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic Perspectives*, *19*, 25–42. <https://doi.org/10.1257/089533005775196732>
- Frederick, S., Loewenstein, G., & O'Donoghue, T. (2002). Time discounting and time preference: A critical review. *Journal of Economic Literature*, *40*, 350–401. <https://doi.org/10.1257/002205102320161311>
- Grabner, R. H. (2014a). The role of intelligence for performance in the prototypical expertise domain of chess. *Intelligence*, *45*, 26–33. <https://doi.org/10.1016/j.intell.2013.07.023>
- Grabner, R. H. (2014b). Going beyond the expert-performance framework in the domain of chess. *Intelligence*, *45*, 109–111. <https://doi.org/10.1016/j.intell.2014.02.010>
- Jamison, J., & Wegener, J. (2010). Multiple selves in intertemporal choice. *Journal of Economic Psychology*, *31*, 832–839. <https://doi.org/10.1016/j.joep.2010.03.004>
- Jung, W. H., Kim, S. N., Lee, T. Y., Jang, J. H., Choi, C. H., Kang, D. H., & Kwon, J. S. (2013). Exploring the brains of Baduk (Go) experts: gray matter morphometry, resting-state functional connectivity, and graph theoretical analysis. *Frontiers in Human Neuroscience*, *7*, 633, 1–17. <https://doi.org/10.3389/fnhum.2013.00633>
- Kahan, D. M. (2013). Ideology, motivated reasoning, and cognitive reflection. *Judgment and Decision Making*, *8*, 407–424. <http://dx.doi.org/10.2139/ssrn.2182588>
- Kahneman, D. (2003). Maps of bounded rationality: Psychology for behavioral economics. *American Economic Review*, *93*, 1449–1475. <https://doi.org/10.1257/000282803322655392>
- Keynes, J. M. (1936). *The general theory of employment interest and money*. Macmillan.
- Kim, S. H., Han, D. H., Lee, Y. S., Kim, B.-N., Cheong, J. H., & Han, S. H. (2014). Baduk (the game of Go) improved cognitive function and brain activity in children with attention deficit hyperactivity disorder. *Psychiatry Investigation*, *11*, 143–151. <https://doi.org/10.4306/pi.2014.11.2.143>
- Kirchler, M., Lindner, F. & Weitzel, U. (2016). Rankings and risk-taking in the finance industry. *The Journal of Finance*, *73*, 2271–2302. <https://doi.org/10.1111/jofi.12701>
- Kokkinakis, A. V., Cowling, P. I., Drachen, A., & Wade, A. R. (2017). Exploring the relationship between video game expertise and fluid intelligence. *PLoS One*, *12*, e0186621. <https://doi.org/10.1371/journal.pone.0186621>
- Li, X., Huang, L., Li, B., Wang, H., & Han, C. (2020). Time for a true display of skill: Top players in league of legends have better executive control. *Acta Psychologica*, *204*, 103007. <https://doi.org/10.1016/j.actpsy.2020.103007>
- Lee, B., Park, J.-Y., Jung, W. H., Kim, H. S., Oh, J. S., Choi, C.-H., Jang, J. H., Kang, D.-H., & Kwon, J. S. (2010). White matter neuroplastic changes in long-term trained players of the game of “baduk” (GO): A voxel-based diffusion-tensor imaging study. *NeuroImage*, *52*, 9–19. <https://doi.org/10.1016/j.neuroimage.2010.04.014>
- Maguire, E. A., Gadian, D. G., Johnsrude, I. S., Good, C. D., Ashburner, J., Frackowiak, R. S. J., & Frith, C. D. (2000). Navigation-related structural change in the hippocampi of taxi drivers. *Proceedings of National Academy of Sciences*, *97*, 4398–4403. <https://doi.org/10.1073/pnas.070039597>
- Marchetti, A., Castelli, I., Sanvito, L., & Massaro, D. (2014). Is a bird in the hand worth two in the future? Intertemporal choice, attachment and theory of mind in school-aged children. *Frontiers in Psychology*, *5*, 483, 1–10. <https://doi.org/10.3389/fpsyg.2014.00483>
- Meyer, A., Zhou, E., & Shane, F. (2018). The non-effects of repeated exposure to the Cognitive Reflection Test. *Judgment and Decision Making*, *13*, 246–259.
- Moore, C., Barresi, J., & Thompson, C. (1998). The cognitive basis of future-oriented prosocial behavior. *Social Development*, *7*, 198–218. <https://doi.org/10.1111/1467-9507.00062>
- Nagel, R. (1995). Unraveling in guessing games: An experimental study. *American Economic Review*, *85*, 1313–1326.
- Potter, D. L. (1984). Go in the classics. *Go World*, *37*, 16–18.
- Raven, J., Raven, J. C., & Court, J. H. (2000). *Standard progressive matrices*. Oxford Psychology Press.
- Rubinstein, A. (2007). Instinctive and cognitive reasoning: A study of response times. *Economic Journal*, *117*, 1243–1259. <https://doi.org/10.1111/j.1468-0297.2007.02081.x>
- Schou, A. (2005). Gæt-et-tal konkurrence afslører at vi er irrationelle [Guessing game reveals that we are irrational]. *Politiken*.
- Selten, R. & Nagel, R. (1998). Das Zahlenwahlspiel–Ergebnisse und Hintergrund [The guessing game–results and background]. *Spektrum der Wissenschaft*, *2*, 16–22.
- Shamosh, N. A., & Gray, J. R. (2008). Delay discounting and intelligence: A meta-analysis. *Intelligence*, *36*(4), 289–305. <https://doi.org/10.1016/j.intell.2007.09.004>

- Silver, D., Huang, A., Maddison, C. J., Guez, A., Sifre, L., van den Driessche, G., Nham, J., Kalchbrenner, N., Sutskever, I., Lillicrap, T., Leach, M., Kavukcuoglu, K., Graepel, T., & Hassabis, D. (2016). Mastering the game of go with deep neural networks and tree search. *Nature*, 529, 484–489. <https://doi.org/10.1038/nature16961>
- Söderqvist, S., Nutley, S. B., Peyrard-Janvid, M., Matsson, H., & Humphreys, K. (2012). Dopamine, working memory, and training induced plasticity: Implications for developmental research. *Developmental Psychology*, 48, 836–843. <https://doi.org/10.1037/a0026179>
- Stanovich, K. E. & West, R. F. (2000). Individual differences in reasoning: Implications for the rationality debate? *Behavioral and Brain Sciences* 23, 645–726.
- Thaler, R. H. (1997). Giving markets a human dimension. *Financial Times*, 6.
- Thaler, R. H. (2015). Keynes' s beauty contest. *Financial Times*.
- Toplak, M., West, R. F. & Stanovich, K. E. (2011). The cognitive reflection test as a predictor of performance on heuristics-and-biases tasks. *Memory and Cognition* 39, 1275–1289. <https://doi.org/10.3758/s13421-011-0104-1>
- Vischer, T., Dohmen, T., Falk, A., Huffman, D., Schupp, J., Sunde, U., & Wagner, G. G. (2013). Validating an ultra-short survey measure of patience. *Economics Letters* 120, 142–145. <https://doi.org/10.1016/j.econlet.2013.04.007>
- Wang, M., Rieger, M. O., & Hens, T. (2016). How time preferences differ: Evidence from 53 countries. *Journal of Economic Psychology*, 52, 115–135. <https://doi.org/10.1016/j.joep.2015.12.001>

RECEIVED 28.07.2020 | ACCEPTED 08.02.2021