

Using response times to measure strategic complexity and the value of thinking in games ^{*}

David Gill [†]
Victoria Prowse [‡]

This version: January 20, 2017

First version: January 19, 2017

Abstract

Response times are a simple low-cost indicator of the process of reasoning in strategic games (Rubinstein, 2007; Rubinstein, 2016). We leverage the dynamic nature of response-time data from repeated strategic interactions to measure the strategic complexity of a situation by how long people think on average when they face that situation (where we define situations according to the characteristics of play in the previous round). We find that strategic complexity varies significantly across situations, and we find considerable heterogeneity in how responsive subjects' thinking times are to complexity. We also study how variation in response times at the individual level across rounds affects strategic behavior and success. We find that 'overthinking' is detrimental to performance: when a subject thinks for longer than she would normally do in a particular situation, she wins less frequently and earns less. The behavioral mechanism that drives the reduction in performance is a tendency to move away from Nash equilibrium behavior.

Keywords: Response time; decision time; thinking time; strategic complexity; game theory; strategic games; repeated games; beauty contest; cognitive ability; personality.

JEL Classification: C72; C91.

^{*}We thank the Economic Science Laboratory at the University of Arizona (and, in particular, Andreas Blume, Eric Cardella, Ray Chiu, Martin Dufwenberg, James Fisher, Richard Kiser, Quyen Nguyen, and John Wooders) for hosting our experiment. We also thank Peter Wagner for excellent research assistance. Prowse gratefully acknowledges financial support from a Zvi Meitar – Oxford University Social Sciences Research Grant. A supplementary web appendix provides additional materials.

[†]Department of Economics, Purdue University; gill53@purdue.edu.

[‡]Department of Economics, Purdue University; vprowse@purdue.edu.

1 Introduction

Psychologists have long been interested in choice response or decision times (e.g., Stone, 1960, and Busemeyer and Townsend, 1993). Recently economists have become interested in how the speed of decision making affects behavior and outcomes. Response times are thought to be connected to decision-making style: fast thinking is linked to intuitive or instinctive decision making, while slower thinking is linked to a more deliberate or contemplative mode of thought (e.g., Rubinstein, 2016).¹ Following the psychology literature, one stream of research in economics looks at response times in non-strategic environments: Spiliopoulos and Ortmann (2016) provide a comprehensive survey, while also developing the different benefits and challenges of collecting and analyzing response-time data.²

In contrast to the literature that studies response times in non-strategic settings, our focus is on response times in strategic settings. Speed of decision making is of particular importance in strategic settings because of the complexity of the environment: players need to think about the payoff structure (‘the rules of the game’) and then form beliefs about how others will behave. For a sophisticated agent forming such beliefs involves iterative thinking about how others think about the agent herself, about how others think about how the agent thinks about them, and so forth. The type of reasoning required to perform well in games requires substantial cognitive effort, but standard theory is silent about how much cognitive effort people exert in strategic settings.³ In reality, agents vary in their willingness to exert cognitive effort and in how useful cognitive effort is for their performance in strategic settings. According to Rubinstein (2007), response time is a simple low-cost indicator of the process of reasoning in strategic games, with more ‘cognitive’ choices taking longer than ‘instinctive’ ones.⁴ The speed of decision making in strategic environments has received some recent interest. Experiments on the relationship between response times and strategic behavior have considered how response times relate to: the type of game being played; rates of cooperation in public goods games; behavior in contests; strategies designed to persuade; threshold strategies in global games; and private information in auctions and environments with social learning.⁵

¹As noted by Rubinstein (2016), this distinction between instinctive and contemplative decision making is in some sense consistent with Kahneman (2011)’s distinction between ‘system I’ and ‘system II’ decision processes, although economists generally study response times measured in seconds while psychologists generally study response times measured in fractions of seconds.

²Wilcox (1993) finds that lottery choice response times respond positively to incentives. Chabris et al. (2009) find that response times increase with the similarity in expected values of binary choices. Piovesan and Wengström (2009) and Lohse et al. (2014) find that generosity is associated with longer response times (both across and within-individuals). Achtziger and Alós-Ferrer (2013) link response times to Bayesian updating. Clithero and Rangel (2013) use response times together with choice data to predict choice out of sample. Rubinstein (2013) finds that shorter response times are associated with more mistakes but not with the incidence of behavior inconsistent with standard models of decision making such as expected-utility theory. Cappelen et al. (2015), Chen and Fischbacher (2015), Hutcherson et al. (2015) and Konovalov and Krajbich (2016) show that response times are associated with risk, time and social preferences.

³A small literature in behavioral economics addresses this issue: see, e.g., Alaoui and Penta (2016) who present a cost-benefit analysis of the endogenous depth of reasoning in games.

⁴In the one-shot p -beauty contest, Rubinstein (2007) argues that choosing 33-34 or 22 is more cognitive, and finds that these choices take longer.

⁵Kuo et al. (2009) find faster response times in coordination games than in dominance-solvable games, while Di Guida and Devetag (2013) find shorter response times in games with focal points. Gneezy et al. (2010) find that response times in the Race Game are longer in losing positions. Arad and Rubinstein (2012) study the relationship between response times and behavior in the Colonel Blotto contest. Glazer and Rubinstein (2012) study the association between response times and behavioral types in a game of persuasion. Rand et al. (2012), Lotito et al. (2013) and Nielsen et al. (2014) find that shorter response times are associated with more cooperation in public goods games, although Recalde et al. (2015) argue that the relationship may reflect mistakes and Krajbich et al. (2015) argue that the direction of the correlation is not robust to changing the relative attractiveness of the selfish and cooperative actions. Schotter and Trevino (2014) look at the relationship between response times and threshold strategies in a global game. Agranov et al. (2015) elicit incentivized choices at multiple points in time and find that ‘sophisticated’ players decrease their choice with thinking time in a p -beauty contest variant. Turocy and Cason (2015) consider the relationship between signals and response times in auctions. Frydman and Krajbich (2016) find that in a social learning environment people can learn about others’

In particular, we aim to study response times in *repeated* strategic interactions. Strategic interactions of economic interest are often repeated: examples include repeated rounds of job hiring and searching, markets with repeated price or quantity competition, repeated selling of goods via auction and multiple rounds of competition for promotions within firms. As we explain in more detail below, we leverage the dynamic nature of response-time data from repeated strategic interactions to: (i) measure the strategic complexity of a situation by how long people think on average when they face that situation (where we define situations according to the characteristics of play in the previous round); and (ii) discover whether variation in thinking time across rounds in which an individual faces the same level of strategic complexity affects strategic behavior and success.

To study response times in repeated strategic interactions we use the experimental data collected by us in Gill and Prowse (2016).⁶ In the experiment, 780 student subjects were matched into 260 groups of three players. Each group of three played the p -beauty contest for ten rounds with feedback and no rematching.⁷ In each round, the subject whose chosen number was closest to seventy percent of the average of the three numbers chosen by the group members (i.e., $p = 0.7$) won six dollars. The subjects had ninety seconds to make their choice in each round. Gill and Prowse (2016) also measured the subjects' cognitive ability using the Raven test, and measured the personality of a subset of the subjects (the Big Five, grit and a measure of future orientation).

In the p -beauty contest the incentive to undercut the average of the choices drives the equilibrium to the lower bound of the action set: the unique Nash equilibrium is for all players to choose zero. With repetition choices in experiments move toward the equilibrium (e.g., Nagel, 1995, Ho et al., 1998, Gill and Prowse, 2016). However, the game is well-suited to studying strategic thinking: players who expect others to select non-equilibrium actions often have an incentive to choose away from equilibrium themselves.⁸ Real-world parallels of the p -beauty contest include timing games in financial and labor markets. During a bubble or in a job market, there is an advantage to trading or making job offers a little earlier than competitors, but moving too early is costly (in terms of lost profit on the upward wave of the bubble or missing out on new information about job candidates).⁹

We aim to study behavior while the subjects are learning to play equilibrium. Once groups reach equilibrium response times fall and behavior tends to stay at the equilibrium. Thus, throughout the paper we study response times before groups reach equilibrium, which constitutes the vast majority of our data.¹⁰ We start with between-subject variation in response times. A few papers find a positive between-subject relationship between response times and success in strategic games.¹¹ We replicate this finding for our beauty contest game by showing that subjects who think for longer on average win more rounds and choose lower numbers (that is, numbers closer to equilibrium). We find

private information from observing their response time. Spiliopoulos (2016) studies a repeated 2×2 constant-sum game and finds that the win-stay-lose-shift heuristic is associated with faster choices. Brañas-Garza et al. (forthcoming) study the response times of ultimatum-game proposers.

⁶Gill and Prowse (2016) investigate how cognitive ability and personality influence the evolution of play toward Nash equilibrium. The paper does not report or study response times.

⁷See Nagel et al. (forthcoming) for a history of the beauty contest game.

⁸The group size of three in our data maximizes the number of independent observations, while ensuring that the game remains strategically interesting (when the group size is two, choosing zero is weakly dominant).

⁹Roth and Xing (1994) provide evidence of slow unraveling of the timing of offers in entry-level professional job markets.

¹⁰Specifically, we use subject-round observations for which the subject's three-player group has never played equilibrium in a previous round.

¹¹Arad and Rubinstein (2012) find that longer response times are associated with winning more battles in the Colonel Blotto game. Brañas-Garza et al. (forthcoming) find that ultimatum-game proposers who think longer earn more. With hypothetical payoffs, Rubinstein (2016) finds that in a 2×2 zero-sum game, the more contemplative action (that is, the action associated with more thought on average) yields a higher expected payoff (in non-zero-sum games the relationship between contemplative actions and payoffs is not as clear).

no statistically significant relationship between cognitive ability or personality and average response times.¹²

Our first substantive contribution is to leverage the dynamic nature of repeated-game response times to develop a measure of strategic complexity. In our repeated-game setting with fixed groups, the subjects may perceive that the strategic complexity of the situation that they face varies with the characteristics of play in the previous round. Motivated by this observation, we categorize ten different situations according to the particular subject’s earnings in the previous round and the rank-order of the choices of the three group members in that round.¹³ We then measure the strategic complexity of a situation by how long subjects think on average when they face that particular situation. Using average thinking time to measure the strategic complexity of a situation in our repeated-game setting is related to Rubinstein (2016)’s distinction between ‘contemplative’ and ‘instinctive’ actions in one-shot games according to how long subjects think on average before choosing the action: according to Rubinstein (2016), more contemplative actions requiring more strategic reasoning.¹⁴ Having developed our measure of strategic complexity, we show that strategic complexity varies significantly across situations. We also find considerable between-subject heterogeneity in how responsive subjects’ thinking times are to changes in strategic complexity: we estimate a two-type mixture regression model and find that one type of subject varies her response times substantially with the strategic complexity of the situation that she faces, while the other type hardly varies her response times at all.

Our second substantive contribution is to study how within-subject variation in response times across rounds affects behavior and success in our repeated-game setting. Specifically, we often observe the same subject facing the same situation more than once, and we can measure whether thinking for longer or for less long than the subject would normally do in that situation affects the subject’s choices and her probability of winning the round.^{15,16} We find that thinking more than normal results in worse performance: when a subject thinks for longer than she would normally do in a particular situation, she wins less frequently and earns less. The behavioral mechanism that drives the reduction in performance is a tendency to move away from Nash equilibrium behavior: when the subject thinks for longer than normal she is more likely to increase her choice relative the previous round and she

¹²We are not aware of any other work that measures the relationship between personality and response times in strategic games. With respect to cognitive ability, in a public goods game Nielsen et al. (2014) find: (i) a positive relationship between scores in the three-question Cognitive Reflection Test and response time; and (ii) a marginally significant negative relationship between scores in a twenty-question Raven test and response time. However, these relationships were no longer significant when the same game was framed as ‘taking’ rather than ‘giving’.

¹³To give a flavor, we define three of the ten situations here. Situation 1: the subject won in the previous round with the lowest choice, and the two other subjects chose the same higher number. Situation 2: the subject won in the previous round with an ‘intermediate’ choice (that is, one of the other subjects lost with a lower number and the other lost with a higher number). Situation 3: the subject lost in the previous round with the highest choice, and the other two subjects chose lower numbers different from each other.

¹⁴Rubinstein (2016) further defines a subject’s ‘contemplative index’ to be her propensity to choose contemplative actions across different games.

¹⁵Recall from the previous paragraph that we categorize situations according to the characteristics of play in the previous round (that is, according to the particular subject’s earnings in the previous round and the rank-order of the choices of the three group members in that round).

¹⁶Some existing papers constrain thinking time (see the survey by Spiliopoulos and Ortmann, 2016). However, in the context of strategic games, constraining players’ thinking time does not give a clean measure of how a player’s strategic behavior varies in her own thinking time because a player’s behavior could also change in anticipation of the time constraint on the behavior of the other players. Furthermore, under a time constraint experimental subjects are not choosing how long to think and the imposed time constraint is often very short. Most closely related to our competitive strategic setting, with a time limit of just fifteen seconds Kocher and Sutter (2006) find slower convergence to equilibrium in a modified beauty contest, with a time limit of fifteen seconds to both read the instructions and then decide Lindner and Sutter (2013) find fewer sophisticated choices in the 11-20 game (but also choices closer to the mixed Nash equilibrium), and in a design where subjects have just twenty seconds to search payoff boxes using their mouse and then decide Spiliopoulos et al. (2015) find that time pressure reduces payoff search and induces simpler heuristics in normal-form games.

is less likely to choose the equilibrium number. We conclude that ‘overthinking’ is detrimental to performance in our strategic setting. Interestingly, these results based on within-subject variation show that overthinking is detrimental even though, as noted above, between-subject variation reveals that subjects who think for longer on average (across situations and rounds) perform better on average. However, this between-subject variation fails to capture the effect of overthinking at the individual level, and the between-subject correlation between average response time and average performance might be driven by the effect of unobserved individual characteristics on both response times and performance.

Together, our findings shed new light on how subjects allocate cognitive resources in games and the consequences of doing so. Our finding that, in repeated games, thinking time varies across situations provides evidence that subjects respond to the characteristics of the situation that they face when deciding how much cognitive effort to allocate to the situation. However, only one type of subject responds to strategic complexity, which highlights the importance of taking seriously across-subject heterogeneity. Our finding that overthinking can be detrimental to performance suggests that allocating too many cognitive resources in games can be counterproductive relative to quicker more instinctive thinking.

We conclude with hope that our findings will spur further empirical and theoretical research on the important topic of response times in games. A better understanding of how and when subjects allocate time and cognitive resources in games will help to refine existing models of boundedly-rational thinking in games, such as level- k thinking (Stahl and Wilson, 1995), and also help to build new ones. In the context of the level- k model, our results suggest that the level of strategic reasoning that people achieve might vary predictably with the particular characteristics of the strategic situation that they face. Furthermore, our findings suggest that new models that incorporate explicitly the choice of how hard to think in strategic interactions could yield better predictive power.

The paper proceeds as follows: Section 2 describes the experimental design; Section 3 provides descriptive statistics on response times; Section 4 uses response times to measure strategic complexity and explores between-subject heterogeneity in responsiveness to complexity; Section 5 studies how individual-level variation in thinking time across rounds affects behavior and performance; and Section 6 concludes.

2 Experimental design

As explained in the introduction, we use the experimental data collected by us in Gill and Prowse (2016).¹⁷ We ran thirty-seven experimental sessions at the University of Arizona’s Experimental Science Laboratory (ESL). Each session lasted approximately seventy-five minutes. In total, 780 student subjects participated in our experiment, with eighteen or twenty-four subjects per session. On average subjects earned twenty United States dollars, on top of a show-up fee of five dollars (subjects were paid privately in cash). The Supplementary Web Appendix provides the experimental instructions.¹⁸

Subjects played ten rounds of the p -beauty contest game in fixed groups of three without rematching. In every round each group member privately chose an integer $x \in \{0, 1, \dots, 100\}$. We implemented

¹⁷See footnote 6.

¹⁸We drew the participants from the ESL subject pool (which is managed using a bespoke online recruitment system) and we excluded graduate students in economics. We randomized seating positions. We provided the experimental instructions to each subject on their computer screen and we read the instructions aloud (questions were answered privately). The experiment was programmed in z-Tree (Fischbacher, 2007).

the beauty contest with $p = 0.7$: the group member whose chosen number was closest to seventy percent of the mean of the three numbers chosen by the group members (the ‘target’) was paid six dollars and the other group members received nothing. In the case of ties, the six dollars was split equally among the subjects who tied. The unique Nash equilibrium is for all players to choose zero.

The subjects had ninety seconds to make their choice in each round.¹⁹ While making their choice, the subjects could see a reminder of the rules. At the end of the ninety seconds, all groups advanced together to a feedback stage that lasted thirty seconds. We provided feedback about the group members’ choices in that round, seventy percent of the mean of the choices, and the earnings of the group members in that round.

Before the start of the first round, we measured the subjects’ cognitive ability using a thirty-minute computerized Raven test, and subjects were matched into groups of three to play the p -beauty contest according to their Raven test score.²⁰ We also measured the personality of 270 of our 780 subjects using an eight-minute questionnaire that was administered before the test of cognitive ability. For these 270 subjects, we measured the Big Five (openness, conscientiousness, extraversion, agreeableness and emotional stability), as well as grit and a measure of future orientation called Consideration of Future of Consequences (CFC).²¹ We find a high degree of correlation between our seven measures of personality, which justifies the construction of a smaller number of uncorrelated personality factors. Varimax rotation (Jolliffe, 1995) generates three factors: Factor 1 mainly captures conscientiousness, grit and CFC, Factor 2 mainly captures agreeableness and emotional stability, while Factor 3 mainly captures openness, extraversion and CFC. Gill and Prowse (2016) provide further details, including details of the subject matching by Raven test score and of the personality factor loadings.

3 Descriptive statistics on response times

Our sample consists of 780 subjects observed for ten rounds, which gives 7,800 subject-round observations. We aim to study behavior while the subjects are learning to play equilibrium. Once groups reach equilibrium, i.e., once all three group members simultaneously choose zero for the first time, behavior rarely changes and response times fall markedly.²² Thus, throughout this paper we study response times before subjects observe that their group has reached equilibrium, which make up the vast majority of our data. Specifically, we use subject-round observations for which the subject’s three-player group has never played equilibrium in a previous round, which constitute more than ninety-five percent of the subject-round observations from the experiment.

Figure 1(a) shows that, despite subjects gaining experience with the game, the average response time is around twenty-eight seconds in every round of the experiment.²³ Figure 1(b) shows the distribution of the subject-round observations of response time (this captures both within-subject

¹⁹If a subject failed to make a choice within ninety seconds, then a flashing request prompted an immediate choice.

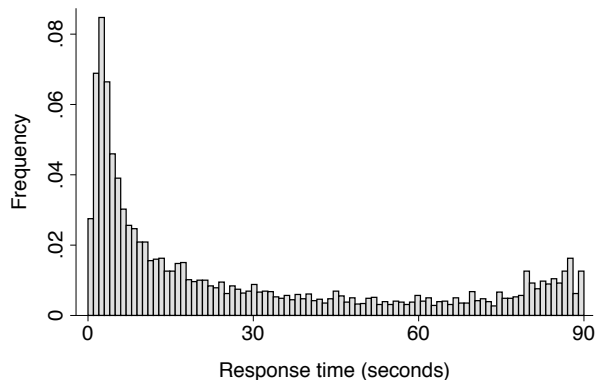
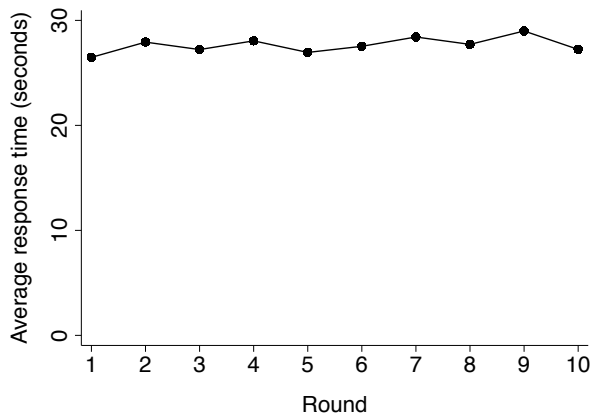
²⁰The Raven test is recognized as a leading measure of analytic or fluid intelligence (Carpenter et al., 1990; Gray and Thompson, 2004, Box 1, p.472). We used the Standard Progressive Matrices Plus version of the Raven test, which consists of sixty questions. We did not provide any monetary incentives for completing the Raven test.

²¹We measured the Big Five using the forty-four-item Big Five Inventory (John et al., 1991; John et al., 2008), grit using the twelve-item Grit Scale (Duckworth et al., 2007), and CFC using the twelve-item CFC Scale (Strathman et al., 1994). 0.3% of the responses are missing (57 of $270 \times 68 = 18,360$). For each question, we replaced any missing responses by the sample average of the non-missing responses to that particular question.

²²Around ninety-seven percent of subjects whose group played equilibrium in a particular round choose the equilibrium number in the next round. The average response time for subjects whose group has previously played equilibrium is 17.7 seconds. In contrast, the average response time for subjects whose group has never played equilibrium in a previous round is around twenty-eight seconds in every round (see Figure 1(a)).

²³We test the significance of the trend in the average response time over the experiment by regressing response time on a linear round variable. The two-sided p -value for the coefficient on the linear round variable is 0.388.

heterogeneity, which arises when a subject varies her response time from one round to the next, and across-subject heterogeneity, which originates from systematic differences in average response times across subjects). We see considerable heterogeneity in the subject-round observations of response time: while around one-quarter of responses occur with the first three seconds of the response window, around five percent of responses occur during the final five seconds of the ninety-second-long response window.



(a) Average response time in rounds 1–10

(b) Distribution of subject-round observations of response time

Figure 1

As we discussed in the introduction, a few papers find a positive between-subject relationship between response times and success in strategic games (see footnote 11). We replicate this finding for our beauty contest game. Specifically, we run a between-subject analysis in which we regress measures of success in the experiment on the subject-level average response time. We consider three measures of success: the fraction of rounds won; earnings per round; and log earnings per round. The first three columns of Table 1 shows that a subject’s average response time is a predictor of her success: subjects who think for longer on average are more likely to win and earn more. To understand the behavior that underlines these patterns, we regress p -beauty contest choices on the subject-level average response time. The last column of Table 1 shows that subjects who think for longer on average choose lower numbers in the beauty contest, i.e., they choose numbers closer to Nash equilibrium.

We also explore how response times are related to cognitive ability and personality by regressing the subject-level average response time on Raven test scores and the three factors that measure personality (Section 2 explained how we constructed the personality factors). Table 2 shows that there is no significant relationship between average response times and Raven test scores or the three personality factors. A subject’s average thinking time, therefore, is explained by characteristics or skills that are not captured by our measures of cognitive ability or personality.

	Fraction of rounds won	Earnings per round (cents)	Log earnings per round (cents)	Average choice
Average response time (minutes)	0.069*** (0.018)	36.220*** (9.328)	0.166*** (0.042)	-2.457*** (0.827)
Intercept	0.377*** (0.009)	183.253*** (4.367)	4.784*** (0.020)	20.705*** (0.604)
Subjects	780	780	780	780

Notes: All averages are taken at the subject level and, as explained in the first paragraph of Section 3, we exclude subject-round observations for which the subject's three-player group has played equilibrium in a previous round. When calculating the fraction of rounds won, a subject is considered to be a winner if she won all or part of the prize. Log earnings per round is calculated by taking the log of earnings at the round level and then averaging at the subject level over rounds. When taking the log of earnings, we add fifty cents to earnings in each round (the show-up fee of five dollars divided by the number of rounds) to avoid taking the log of zero. Heteroskedasticity-consistent standard errors with clustering at the group level are shown in parentheses. *, ** and *** denote significance at the 10%, 5% and 1% levels (two-sided tests).

Table 1: Average response time, success and choices

	Average response time (minutes)		
Raven test score (cognitive ability)	-0.003 (0.013)		0.009 (0.020)
Personality factor 1 (conscientiousness, grit and future orientation)		-0.033 (0.023)	-0.032 (0.023)
Personality factor 2 (agreeableness and emotional stability)		0.004 (0.019)	0.003 (0.019)
Personality factor 3 (openness, extraversion and future orientation)		-0.016 (0.021)	-0.016 (0.021)
Intercept	0.462*** (0.013)	0.454*** (0.020)	0.454*** (0.020)
Subjects	780	270	270

Notes: All averages are taken at the subject level and, as explained in the first paragraph of Section 3, we exclude subject-round observations for which the subject's three-player group has played equilibrium in a previous round. The Raven test score and personality factors 1–3 have been standardized to have means of zero and standard deviations of one (Section 2 explained how we constructed the personality factors). Personality was measured for 270 of our 780 subjects. Heteroskedasticity-consistent standard errors with clustering at the group level are shown in parentheses. *, ** and *** denote significance at the 10%, 5% and 1% levels (two-sided tests).

Table 2: Raven test score, personality and average response time

4 Strategic complexity

We now turn to our first substantive contribution: we use the dynamic nature of the repeated response times in our beauty contest game to measure the strategic complexity of the situations that our subjects face and to explore the heterogeneity in how subjects adjust their thinking time in response to changes in strategic complexity. We proceed in three stages: first, we categorize situations according to the characteristics of play in the previous round; second, we measure the strategic complexity of each situation by how long subjects think on average when they face that situation; and finally, we explore heterogeneity in how subjects respond to changes in strategic complexity.

4.1 Categorizing situations

In our repeated-game setting with fixed groups, subjects may perceive that the strategic complexity of the situation that they face in a given round varies with the characteristics of play in the previous round. Motivated by this observation, we categorize situations according to the subject's earnings in the previous round and the rank-order of the choices of three group members in the previous round. Let n denote the number chosen by the subject in the previous round, and let \underline{n}^o and \bar{n}^o denote the numbers chosen in the previous round by the subject's opponents. Without loss of generality, we order the opponents such that $\underline{n}^o \leq \bar{n}^o$. The subject's situation in a given round is determined by the subject's earnings in the previous round and the ordering of n , \underline{n}^o and \bar{n}^o . We focus on the ten situations described in the first column of Table 3.²⁴ The second column of Table 3 shows that we have between 123 and 1,739 subject-round observations for each situation, giving a total of 6,670 subject-round observations across the ten situations. We order the situations by strategic complexity: the third column of Table 3 reports our measure of strategic complexity, which we explain in Section 4.2.

4.2 Measuring strategic complexity

We measure the strategic complexity of a situation by how long subjects think on average when they face that particular situation. To identify the influence of the situation on average response times we cannot simply measure the average response time in each situation because subjects vary systematically in how long they think and subjects who tend to think for longer might be more likely to face certain situations. Thus, we leverage the repeated observations of response times to separate the component of response times that is attributable to the situation from the component of response times that is due to systematic differences in thinking times between subjects. We do this by running the following fixed-effects regression for subject i 's response time in seconds in round r :

$$\text{ResponseTime}_{i,r} = \sum_k \theta_k D_{i,r}^k + \alpha_i + \tau_r + \varepsilon_{i,r} \quad \text{for } r = 2, \dots, 10, \quad (1)$$

where $D_{i,r}^k$ is an indicator for subject i being in situation k in round r , α_i is a subject fixed effect that absorbs systematic between-subject difference in response times, τ_r is a round fixed effect that captures any trends in response times over the experiment (round five is the reference category), and $\varepsilon_{i,r}$ is an error term. As explained in the first paragraph of Section 3, we exclude subject-round observations for which the subject's three-player group has played equilibrium in a previous round.

²⁴These ten situations represent all feasible situations with two exceptions: the situation where $n < \underline{n}^o < \bar{n}^o$ and the subject earned three dollars (the subject tied with the opponent who chose \underline{n}^o); and the situation where $\underline{n}^o < n < \bar{n}^o$ and the subject earned three dollars (again the subject tied with the opponent who chose \underline{n}^o). We omit these two situations because we have only four subject-round observations from each of the situations.

When estimating (1), one situation is arbitrarily chosen to be the reference category ($k = 1; \theta_1 = 0$), and the parameter θ_k then measures the pure effect on response time of being faced with situation k instead of this reference situation. Using estimates from the fixed-effects regression described in (1), we calculate the strategic complexity of situation k as follows:

$$\text{StrategicComplexity}_k \equiv \hat{\theta}_k + \overline{\hat{\alpha}_i}, \quad (2)$$

where $\hat{\theta}_k$ denotes the estimate of θ_k and $\overline{\hat{\alpha}_i}$ is the average of the estimates of the subject fixed effects.

The third column of Table 3 reveals that strategic complexity, measured by average thinking time as detailed in (2), varies substantially across the ten situations. The most complex situation is where the subject won the entire prize of six dollars in the previous round with a choice that was between that of her opponents; in this situation, subjects think for an average of around thirty seconds. The two least complex situations are those where the subject lost the previous round with a choice that was higher than that of both of her opponents; in these situations subjects think for an average of around twenty-five seconds. The variation in strategic complexity across situations is statistically significant: the null hypothesis that strategic complexity is constant across situations is strongly rejected (p -value = 0.006).

Situation	Subject-round observations	Strategic Complexity (Average response time in seconds)
$\underline{n}^o < \bar{n}^o < n$ (\therefore subject earned \$0)	1,739	25.086
$\underline{n}^o = \bar{n}^o < n$ (\therefore subject earned \$0)	265	25.870
$n < \underline{n}^o < \bar{n}^o$ & subject earned \$0	633	26.038
$\underline{n}^o < n < \bar{n}^o$ & subject earned \$0	1,102	26.276
$n = \underline{n}^o = \bar{n}^o$ (\therefore subject earned \$2)	123	26.283
$\underline{n}^o < n = \bar{n}^o$ (\therefore subject earned \$0)	362	26.832
$n < \underline{n}^o = \bar{n}^o$ (\therefore subject earned \$6)	181	27.237
$n < \underline{n}^o < \bar{n}^o$ & subject earned \$6	1,102	27.985
$n = \underline{n}^o < \bar{n}^o$ (\therefore subject earned \$3)	530	28.321
$\underline{n}^o < n < \bar{n}^o$ & subject earned \$6	633	30.284

Notes: n denotes the number chosen by the subject in the previous round, and \underline{n}^o and \bar{n}^o denote the numbers chosen in the previous round by the subject's opponents (with $\underline{n}^o \leq \bar{n}^o$). Earnings refer to the subject's earning in the previous round. As explained in the first paragraph of Section 3, we exclude subject-round observations for which the subject's three-player group has played equilibrium in a previous round. We have a total of 6,670 subject-round observations across the ten situations. The third column reports strategic complexity as described in (2).

Table 3: Strategic complexity

4.3 Heterogeneity

We now study whether subjects vary in how they adjust thinking time in response to changes in strategic complexity. We postulate that some subjects may be sensitive to changes in strategic complexity and choose to devote more time to thinking in more complex situations. Other subjects may not tailor their thinking time to the complexity of their situation, either because they fail to recognize that situations vary in complexity or because they lack the self-control needed to think for longer.

We explore the empirical support for this reasoning by estimating the following two-type mixture regression model, which captures heterogeneity in how thinking time responds to changes in strategic complexity:

$$\text{ResponseTime}'_{i,r} = \beta_i \text{StrategicComplexity}'_{i,r} + v_r + \sigma_i \epsilon_{i,r} \quad \text{for } r = 2, \dots, 10, \quad (3)$$

where

$$\text{StrategicComplexity}_{i,r} \equiv \sum_k \text{StrategicComplexity}_k \times D_{i,r}^k \quad \text{for } r = 2, \dots, 10, \quad (4)$$

denotes the strategic complexity of the situation facing subject i in round r , $X'_{i,r}$ denotes $X_{i,r}$ expressed in deviation form (i.e., differenced relative to the average of $X_{i,r}$ for subject i), v_r is a round fixed effect, and $\epsilon_{i,r}$ is an independent error term with a standard normal distribution (see Section 4.2, and in particular (1) and (2) for the definitions of $\text{StrategicComplexity}_k$ and $D_{i,r}^k$). The parameter β_i describes how the subject's response time changes when strategic complexity deviates from the average strategic complexity faced by the subject, and σ_i measures the standard deviation of the component of the subject's response time that is unresponsive to changes in strategic complexity. We distinguish two types of subjects: type 1 subjects have $[\beta_i, \sigma_i] = [\beta_1, \sigma_1]$ and type 2 subjects have $[\beta_i, \sigma_i] = [\beta_2, \sigma_2]$. The probability of a subject being of type j is π_j for $j = 1, 2$.

	Type 1	Type 2
β	1.460*** (0.261)	0.068 (0.083)
σ	22.293*** (0.329)	3.726*** (0.361)
π (type probability)	0.693*** (0.024)	0.307*** (0.024)

Notes: Parameter estimates were obtained by applying Maximum Likelihood estimation to the sample of 6,670 subject-round observations described in Table 3. Heteroskedasticity-consistent standard errors with clustering at the group level are shown in parentheses. *, ** and *** denote significance at the 10%, 5% and 1% levels (two-sided tests for β and one-sided tests elsewhere).

Table 4: Heterogeneous effect of strategic complexity on response time

Table 4 reports the parameter estimates for this two-type mixture regression model. We find that the thinking times of the two types respond very differently to changes in strategic complexity. About seventy percent of subjects are type-1 subjects who substantially increase thinking time when strategic complexity increases (specifically, when the strategic complexity of the situation increases by one second, the response time of type-1 subjects increases by about 1.5 seconds, and this effect is significant at the one-percent level). In contrast, type-2 subjects, who make up about thirty percent of the subject population, are unresponsive to changes in the complexity of their situation. Interestingly, the estimate of σ is larger for type-1 subjects, which means that type-1 subjects also have larger variations in response times for reasons that are unrelated to changes in strategic complexity.²⁵

²⁵The p -value for the null hypothesis that there is only one type (against the alternative of two types) is 0.000. Table SWA.1 in Supplementary Web Appendix II reports the parameter estimates for a three-type mixture regression model: we continue to find that thinking time is responsive to changes in strategic complexity for the majority of subjects.

5 The value of thinking more

In Section 4 we categorized ten situations according to the characteristics of play in the previous round and we measured the strategic complexity of a situation by how long subjects think on average when they face that particular situation. In Section 3 we showed that subjects who think for longer on average are more successful. We now turn to a more subtle question: is a subject more successful when she thinks for longer than she would normally do in a particular situation?

We answer this question by exploiting within-subject variation in response times across rounds that is not attributable to variation in the strategic complexity of the situation facing the subject. Mirroring our analysis in Section 3, we consider three measures of subject i 's success in round r : being a winner of the round; earnings in the round; and the log of earnings in the round. Specifically, for each measure of success we run the following fixed-effects regression, which describes how a subject's success in a particular round depends on her thinking time in that round:

$$\text{Success}_{i,r} = \lambda \text{ResponseTime}_{i,r} + \eta_{i,s} + \gamma_r + e_{i,r}, \quad \text{for } r = 2, \dots, 10, \quad (5)$$

where $\eta_{i,s}$ is a subject-situation fixed effect, γ_r is a round fixed effect (round five is the reference category) and $e_{i,r}$ is an error term. The subject-situation fixed effects absorb subject-specific systematic differences in success across situations. In particular, they absorb the effect of systematic differences in thinking times across situations at the subject level that are correlated with the subject's success. The parameter λ is thus identified from within-subject variation in response times across rounds in which the subject faced the same situation. It follows that we can interpret our estimate of λ as the effect of a subject thinking for longer than she ordinarily does in a particular situation.

Table 5 reports parameter estimates from the fixed-effects regression described in (5). The table shows that thinking for longer than normal results in worse performance. In particular, when a subject thinks for one minute longer than she would normally do in a particular situation, the probability that the subject wins the round decreases by six percentage points. This translates into a reduction in earnings in the round of around twenty-nine cents or fourteen percent. We conclude that 'overthinking' is detrimental to success in our strategic setting.

	Winner of round	Earnings in round	Log earnings in round
Effect of thinking longer than normal (minutes)	-0.060** (0.029)	-28.693* (15.630)	-0.137** (0.069)
Intercept	0.429*** (0.020)	223.490*** (9.834)	4.938*** (0.044)
Subject-round observations	4,499	4,499	4,499

Notes: We start with the sample of 6,670 subject-round observations described in Table 3. From this sample, we then use the subject-round observations for which the subject is observed in the same situation in at least one other round (this ensures that subject-round observations are not fully absorbed by the fixed effects). This gives us 4,499 subject-round observations and 1,678 subject-situation fixed effects. “Intercept” is the mean of the estimates of the subject-situation fixed effects and thus can be interpreted as the mean predicted value of the dependent variable in round 5 given a response time of zero. A subject is considered to be a winner if she won all or part of the prize. When taking the log of earnings, we add fifty cents to earnings in each round (the show-up fee of five dollars divided by the number of rounds) to avoid taking the log of zero. Heteroskedasticity-consistent standard errors with clustering at the group level are shown in parentheses. *, ** and *** denote significance at the 10%, 5% and 1% levels (two-sided tests).

Table 5: Effect of thinking for longer than normal on success

We explore the behavioral mechanism that drives the detrimental effect of overthinking on performance by running fixed-effects regressions as described by (5), but with dependent variables based on the subject’s choice in the round. The behavioral mechanism that drives the reduction in performance is a tendency to move away from Nash equilibrium behavior: Table 6 shows that overthinking makes the subject more likely to increase her choice relative to her choice in the previous round and less likely to choose the equilibrium action of zero. In more detail, when a subject thinks for one minute longer than normal, the probability that her choice increases relative to the previous round goes up by six percentage points and the probability that she chooses the equilibrium action falls by three percentage points. Thinking more than normal also increases the subject’s choice in the round, but this effect is noisy and therefore insignificant.

	Choice in round	Choose zero in round	Increased choice in round
Effect of thinking longer than normal (minutes)	1.023 (1.067)	-0.026** (0.013)	0.058*** (0.019)
Intercept	19.730*** (1.022)	0.062*** (0.013)	0.234*** (0.021)
Subject-round observations	4,499	4,499	4,499

Notes: ‘Choose zero in round’ is an indicator for the subject having chosen the equilibrium action of zero. ‘Increased choice in round’ is an indicator for the subject’s choice in the round being greater than her choice in the previous round. Also see notes to Table 5. Heteroskedasticity-consistent standard errors with clustering at the group level are shown in parentheses. *, ** and *** denote significance at the 10%, 5% and 1% levels (two-sided tests).

Table 6: Effect of thinking for longer than normal on strategic behavior

6 Conclusion

In this paper we extend a recent line of research that uses response times to better understand economic behavior and preferences. To give two examples, Clithero and Rangel (2013) use response times to help predict out-of-sample behavior, while Hutcherson et al. (2015) relate response times to social preferences. We study response times in strategic interactions. Response times in games can allow researchers to gain insight into subjects' reasoning processes that choices alone cannot (Rubinstein, 2016), and subjects themselves can potentially use information about the response times of others to infer information that choices alone do not reveal (Frydman and Krajbich, 2016).

We use experimental data on response times from repeated games to develop a measure of the strategic complexity of a situation based on how long subjects think on average when they face that situation, where situations are defined according to the characteristics of play in the previous round. Our finding that strategic complexity varies significantly across situations provides evidence that subjects respond to the characteristics of the situation that they face when deciding how much cognitive effort to allocate to the situation. But not all subjects do this: one type of subject responds strongly to strategic complexity, while another type hardly responds at all.

We also leverage our response-time data from repeated strategic interactions to show that when subjects think for longer than they would normally do in a particular situation, they perform less well. In particular, when subjects 'overthink' they tend to move away from Nash equilibrium choices, they are less likely to win and they earn less. Our findings suggest that allocating too many cognitive resources in games can be counterproductive: instead, quicker more instinctive responses can be more successful.

We hope that our findings will provoke further research on response times in games that succeeds in improving the predictive power of existing models of boundedly-rational thinking in games, while also inspiring better models that incorporate explicitly the choice of how hard to think in strategic interactions.

References

- Achtziger, A.** and **Alós-Ferrer, C.** (2013). Fast or rational? A response-times study of Bayesian updating. *Management Science*, 60(4): 923–938
- Agranov, M., Caplin, A., and Tergiman, C.** (2015). Naive play and the process of choice in guessing games. *Journal of the Economic Science Association*, 1(2): 146–157
- Alaoui, L. and Penta, A.** (2016). Endogenous depth of reasoning. *Review of Economic Studies*, 83(4): 1297–1333
- Arad, A. and Rubinstein, A.** (2012). Multi-dimensional iterative reasoning in action: The case of the Colonel Blotto game. *Journal of Economic Behavior & Organization*, 84(2): 571–585
- Brañas-Garza, P., Meloso, D., and Miller, L.M.** (forthcoming). Strategic risk and response time across games. *International Journal of Game Theory*
- Busemeyer, J.R. and Townsend, J.T.** (1993). Decision field theory: A dynamic-cognitive approach to decision making in an uncertain environment. *Psychological Review*, 100(3): 432–459
- Cappelen, A.W., Nielsen, U.H., Tungodden, B., Tyran, J.R., and Wengström, E.** (2015). Fairness is intuitive. *Experimental Economics*, DOI 10.1007/s10683-015-9463-y
- Carpenter, P.A., Just, M.A., and Shell, P.** (1990). What one intelligence test measures: A theoretical account of the processing in the Raven Progressive Matrices test. *Psychological Review*, 97(3): 404–413
- Chabris, C.F., Laibson, D., Morris, C.L., Schuldt, J.P., and Taubinsky, D.** (2009). The allocation of time in decision-making. *Journal of the European Economic Association*, 7(2-3): 628–637
- Chen, F. and Fischbacher, U.** (2015). Response time and click position: Cheap indicators of preferences. *Thurgau Institute of Economics Research Paper 102, University of Konstanz*
- Clithero, J.A. and Rangel, A.** (2013). Combining response times and choice data using a neuroeconomic model of the decision process improves out-of-sample predictions. *Mimeo, California Institute of Technology*
- Di Guida, S. and Devetag, G.** (2013). Feature-based choice and similarity perception in normal-form games: An experimental study. *Games*, 4(4): 776–794
- Duckworth, A.L., Peterson, C., Matthews, M.D., and Kelly, D.R.** (2007). Grit: Perseverance and passion for long-term goals. *Journal of Personality and Social Psychology*, 92(6): 1087–1101
- Fischbacher, U.** (2007). z-Tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics*, 10(2): 171–178
- Frydman, C. and Krajbich, I.** (2016). Using response times to infer others’ beliefs: An application to information cascades. *Working Paper 2817026, SSRN*
- Gill, D. and Prowse, V.** (2016). Cognitive ability, character skills and learning to play equilibrium: A level- k analysis. *Journal of Political Economy*, 124(6): 1619–1676
- Glazer, J. and Rubinstein, A.** (2012). A model of persuasion with boundedly rational agents. *Journal of Political Economy*, 120(6): 1057–1082
- Gneezy, U., Rustichini, A., and Vostroknutov, A.** (2010). Experience and insight in the Race game. *Journal of Economic Behavior & Organization*, 75(2): 144–155
- Gray, J. and Thompson, P.** (2004). Neurobiology of intelligence: Science and ethics. *Nature Reviews Neuroscience*, 5(6): 471–482
- Ho, T.H., Camerer, C., and Weigelt, K.** (1998). Iterated dominance and iterated best response in experimental “ p -beauty contests”. *American Economic Review*, 88(4): 947–969
- Hutcherson, C.A., Bushong, B., and Rangel, A.** (2015). A neurocomputational model of altruistic choice and its implications. *Neuron*, 87(2): 451–462
- John, O.P., Naumann, L.P., and Soto, C.J.** (2008). Paradigm shift to the integrative Big Five trait taxonomy: History, measurement, and conceptual issues. In O.P. John, R.W. Robins, and L.A. Pervin, editors, *Handbook of personality: Theory and research*, 114–158. New York, NY: Guilford Press
- John, O.P., Donahue, E.M., and Kentle, R.L.** (1991). The Big Five Inventory: Versions 4a and 54. *Institute of Personality and Social Research, University of California, Berkeley*
- Jolliffe, I.T.** (1995). Rotation of principal components: Choice of normalization constraints. *Journal of Applied Statistics*, 22(1): 29–35
- Kahneman, D.** (2011). *Thinking, fast and slow*. Macmillan

- Kocher, M.G.** and **Sutter, M.** (2006). Time is money? Time pressure, incentives, and the quality of decision-making. *Journal of Economic Behavior & Organization*, 61(3): 375–392
- Kononov, A.** and **Krajbich, I.** (2016). Revealed indifference: Using response times to infer preferences. *Mimeo, Ohio State University*
- Krajbich, I., Bartling, B., Hare, T., and Fehr, E.** (2015). Rethinking fast and slow based on a critique of reaction time reverse inference. *Nature Communications*, 6(7455): 1–9
- Kuo, W.J., Sjöström, T., Chen, Y.P., Wang, Y.H., and Huang, C.Y.** (2009). Intuition and deliberation: Two systems for strategizing in the brain. *Science*, 324(5926): 519–522
- Lindner, F.** and **Sutter, M.** (2013). Level-k reasoning and time pressure in the 11–20 money request game. *Economics Letters*, 120(3): 542–545
- Lohse, J., Goeschl, T., and Diederich, J.** (2014). Giving is a question of time: Response times and contributions to a real world public good. *Department of Economics Discussion Paper 566, University of Heidelberg*
- Lotito, G., Migheli, M., and Ortona, G.** (2013). Is cooperation instinctive? Evidence from the response times in a public goods game. *Journal of Bioeconomics*, 15(2): 123–133
- Nagel, R.** (1995). Unraveling in guessing games: An experimental study. *American Economic Review*, 85(5): 1313–1326
- Nagel, R., Bühren, C., and Frank, B.** (forthcoming). Inspired and inspiring: Hervé Moulin and the discovery of the beauty contest game. *Mathematical Social Sciences*
- Nielsen, U.H., Tyran, J.R., and Wengström, E.** (2014). Second thoughts on free riding. *Economics Letters*, 122(2): 136–139
- Piovesan, M.** and **Wengström, E.** (2009). Fast or fair? A study of response times. *Economics Letters*, 105(2): 193–196
- Rand, D.G., Greene, J.D., and Nowak, M.A.** (2012). Spontaneous giving and calculated greed. *Nature*, 489(7416): 427–430
- Recalde, M.P., Riedl, A., and Vesterlund, L.** (2015). Error prone inference from response time: The case of intuitive generosity. *Department of Economics Working Paper 15/004, University of Pittsburgh*
- Roth, A.** and **Xing, X.** (1994). Jumping the gun: Imperfections and institutions related to the timing of market transactions. *American Economic Review*, 84(4): 992–1044
- Rubinstein, A.** (2007). Instinctive and cognitive reasoning: A study of response times. *Economic Journal*, 117(523): 1243–1259
- Rubinstein, A.** (2013). Response time and decision making: An experimental study. *Judgment and Decision Making*, 8(5): 540–551
- Rubinstein, A.** (2016). A typology of players: Between instinctive and contemplative. *Quarterly Journal of Economics*, 131(2): 859–890
- Schotter, A.** and **Trevino, I.** (2014). Is response time predictive of choice? An experimental study of threshold strategies. *Discussion Paper SP II 2014-305, WZB*
- Spiliopoulos, L.** (2016). The determinants of response time in a repeated constant-sum game: A robust Bayesian hierarchical model. *Working Paper 2740989, SSRN*
- Spiliopoulos, L.** and **Ortmann, A.** (2016). The BCD of response time analysis in experimental economics. *Working Paper 2401325, SSRN*
- Spiliopoulos, L., Ortmann, A., and Zhang, L.** (2015). Attention and choice in games under time constraints: A process analysis. *Working Paper 2620163, SSRN*
- Stahl, D.O.** and **Wilson, P.W.** (1995). On players’ models of other players: Theory and experimental evidence. *Games and Economic Behavior*, 10(1): 218–254
- Stone, M.** (1960). Models for choice-reaction time. *Psychometrika*, 25(3): 251–260
- Strathman, A., Gleicher, F., Boninger, D.S., and Scott Edwards, C.** (1994). The consideration of future consequences: Weighing immediate and distant outcomes of behavior. *Journal of Personality and Social Psychology*, 66(4): 742–752
- Turocy, T.L.** and **Cason, T.N.** (2015). Bidding in first-price and second-price interdependent-values auctions: A laboratory experiment. *Mimeo, University of East Anglia*
- Wilcox, N.T.** (1993). Lottery choice: Incentives, complexity and decision time. *Economic Journal*, 103(421): 1397–1417

Supplementary Web Appendix

(Intended for Online Publication)

Supplementary Web Appendix I Experimental Instructions

Please look at your screen now. I am reading from the instructions displayed on your screen. Please now turn off cell phones and any other electronic devices. These must remain turned off for the duration of this session. Please do not use or place on your desk any personal items, including pens, paper, phones etc. Please do not look into anyone else's booth at any time. Thank you for participating in this experimental session on economic decision-making. You were randomly selected from the Economic Science Laboratory's pool of subjects to be invited to participate in this session. There will be a number of pauses for you to ask questions. During such a pause, please raise your hand if you want to ask a question. Apart from asking questions in this way, you must not communicate with anybody in this room or make any noise.

You will be paid a show-up fee of \$5 together with any money you accumulate during this session. The amount of money you accumulate will depend partly on your actions and partly on the actions of other participants. You will be paid privately in cash at the end of the session.

{Further instructions in the 15 sessions that included questionnaires to measure personality: Please raise your hand if you have any questions. Before we start the experimental session, I would like you to complete a pre-experimental survey. The survey is made up of 68 questions. There are 17 pages with 4 questions on each page. For each question, please enter your answer in the column to the right of the question. There are no right or wrong answers to the questions. You will have 8 minutes to complete the pre-experimental survey. During the 8 minutes, you can move back and forth between the 17 pages and you can change your previous answers. The top right-hand corner of the screen will display the time remaining (in seconds). Before we start the pre-experimental survey, please raise your hand if you have any questions. During the pre-experimental survey, please raise your hand if you have a problem with your computer. [Subjects complete questionnaire] Thank you for completing the pre-experimental survey. I will now describe the experimental session.}

The session is made up of 2 parts. In the first part you will complete a test. Right at the end of the session you will find out your own test score, but you will not be paid for completing the test. I will describe the second part of the session after you have completed the test. Please raise your hand if you have any questions.

I will now describe the test which makes up the first part of the session. The test is made up of 60 questions, divided into parts A, B, C, D and E. Each of these parts is made up of 12 questions. For every question, there is a pattern with a piece missing and a number of pieces below the pattern. You have to choose which of the pieces below is the right one to complete the pattern. For parts A and B of the test, you will see 6 pieces that might complete the pattern. For parts C, D and E you will see 8 pieces that might complete the pattern. In every case, one and only one of these pieces is the right one to complete the pattern.²⁶ For each question, please enter your answer in the column to the right of the pattern. You will score 1 point for every right answer. You will not be penalized for wrong answers. You will have 3 minutes to complete each of parts A and B, and you will have 8 minutes to complete each of parts C, D, and E. During each part, you can move back and forth between the 12 questions in that part and you can change your previous answers. The top right-hand corner of the screen will display the time remaining (in seconds). Before we start the test, please raise your hand if you have any questions. During the test, please raise your hand if you have a problem with your computer. [Subjects complete test]

Your screen is now displaying whether your test score was in the top half of the test scores of

²⁶The wording of this description follows the standard Raven test convention.

all participants in the room or was in the bottom half of the test scores of all participants. [30 second pause] [Example (not read aloud): Your test score was in the top half of the test scores of all participants in the room.] At the end of the session you will find out your own test score.

I will now describe the second and final part of the session. This second part is made up of 10 rounds. You will be anonymously matched into groups of 3 participants. You will stay in the same group for all 10 rounds. In each round, you and your other 2 group members will separately choose a whole number between 0 and 100 (0, 100 or any whole number in between is allowed). The group member whose chosen number is closest to 70% of the average of all 3 chosen numbers will be paid \$6 for that round and the other 2 group members will be paid nothing. If more than one group member chooses a number which is closest to 70% of the average of all 3 chosen numbers, the \$6 will be split equally among the group members who chose the closest number or numbers. Your total payment will be the sum of your payments in each round together with your show-up fee of \$5. In each round you will have 90 seconds to choose your number. If you choose your number early you will still have to wait until the end of the 90 seconds. The top right-hand corner of the screen will display the time remaining (in seconds). The screen will also include a reminder of the rules.

At the end of each round you will discover: (i) the numbers chosen by all your group members; (ii) the average of all 3 chosen numbers; (iii) what 70% of the average of all 3 chosen numbers was; and (iv) how much each group member will be paid for the round. Please raise your hand if you have any questions.

You will stay in the same group of 3 for all 10 rounds. Each group member has been randomly allocated a label, X, Y or Z. Your screen is now displaying your label and whether the test scores of the members of your group were in the top half or the bottom half of the test scores of all participants in the room. [60 second pause] [Example (not read aloud): You are group member Y. Your test score was in the top half of the test scores of all participants in the room. You have been matched with 2 participants (group member X and group member Z). Group member X was randomly selected from those whose test scores were also in the top half. Group member Z was randomly selected from those whose test scores were in the bottom half.] Please raise your hand if you have any questions. There will be no further opportunities for questions.

[10 rounds of beauty contest with feedback as described in Section 2.3]

[Screen asks subjects to report their gender]

[Screen reports the subject's score in the Raven test]

The session has now finished. Your total cash payment, including the show-up fee, is displayed on your screen. Please remain in your seat until you have been paid. Thank-you for participating.

Supplementary Web Appendix II Further results

	Type 1	Type 2	Type 3
β	1.616*** (0.312)	0.371 (0.278)	0.077 0.083
σ	23.983*** (0.546)	8.594*** (1.634)	2.201*** (0.423)
π (type probability)	0.570*** (0.046)	0.270*** (0.035)	0.161*** (0.038)

Notes: See notes to Table 4 for descriptions of the estimation method and sample. Heteroskedasticity-consistent standard errors with clustering at the group level are shown in parentheses. *, ** and *** denote significance at the 10%, 5% and 1% levels (two-sided tests for β and one-sided tests elsewhere).

Table SWA.1: Further results on the heterogeneous effect of strategic complexity on response time