Comprehension and deliberation: decomposing response time in economic experiments

Eszter Timár* Hubert János Kiss§ Dániel Horn¶

Abstract

Response time is increasingly used as a window to look into decision-making process in economic experiments. Yet its interpretation is ambiguous: a fast response may reflect quick comprehension or an impulsive choice. We show that the source of this ambiguity is conceptual and can be resolved by a simple design feature: the separation of instruction and decision screens. Using data from approximately 1,100 students in a lab-in-the-field experiment linked to standardized test scores, we decompose total response time into comprehension time and deliberation time. We show that each component is informative in its own right. Comprehension time is correlated with cognitive ability, providing a cost-free ability proxy in any experiment with split screens. Deliberation time is independent of ability but responds to the structure of the choice: heuristic choices are faster and decisions near the indifference point take longer. The decomposition recovers behavioral patterns that total response time obscures.

JEL classifications: C90; C91; D83

Keywords: response time; comprehension; deliberation; cognitive ability; experimental economics

Acknowledgments

The authors would like to thank seminar participants at Corvinus University of Budapest and at ELTE KRTK KTI, and Ariel Rubinstein for helpful comments provided on the first draft of this paper. This research was supported by the National Research, Development and Innovation Office of Hungary (project no. K-143415).

*Corvinus University of Budapest and ELTE KRTK, Budapest, Hungary. ORCID: https://orcid.org/0000-0002-9811-8298. **Corresponding author.** E-mail: eszter.timar@stud.uni-corvinus.hu.

§ELTE KRTK and Corvinus University of Budapest, Budapest, Hungary. ORCID: https://orcid.org/0000-0003-3666-9331. E-mail: kiss.hubert@krtk.elte.hu.

¶Corvinus University of Budapest and ELTE KRTK, Budapest, Hungary. ORCID: https://orcid.org/0000-0002-2888-6240. E-mail: daniel.horn@uni-corvinus.hu.

1 Introduction

Information about how time is allocated during a task offers a glimpse into the underlying cognitive processes. Driven by this insight, experimental economists have increasingly leveraged process data on the time it takes to reach a decision. Response time, defined as the time elapsed between the presentation of a decision problem and the participant's response, has become the most widely used process measure (Spiliopoulos and Ortmann, 2018). Collecting it is costless, universally available in computerized experiments, and does not influence participants' behavior.

The use of response time in economics is relatively novel (as noted by Rubinstein (2016), and Spiliopoulos and Ortmann (2018)), but its applications are already numerous. Economists have so far leveraged response time data to interpret the meaning of choices in strategic games (Rubinstein, 2006; Lotito et al., 2013; Cappelen et al., 2016), infer task complexity (Agranov et al., 2025) and effort (Wilcox, 1993; Ofek et al., 2007), predict players' out-of-sample behavior (Schotter and Trevino, 2021; Clithero, 2018a), model traits and abilities (Pedersen et al., 2023; Pezeshki and Embretson, 2025), or classify participants into behavioral types (Rubinstein, 2013, 2016).

Yet the interpretation of response time remains ambiguous (Clithero, 2018a). A long response time, for example, may reflect careful deliberation. Alternatively, it may indicate difficulty understanding the task. A short response time, similarly, may reflect either heuristic decision-making or efficient cognitive processing. Between-subject differences in response times are often striking, but their meaning is rarely pinned down (Spiliopoulos and Ortmann, 2018). Given this ambiguity, the multitude of interpretations found across studies is unsurprising.

In this paper, we argue that the source of this ambiguity is conceptual. Observed response time in the economics literature conflates two distinct processes: task comprehension and choice deliberation. Task comprehension is the time required to read and understand the decision problem. Choice deliberation is the time spent making a decision, conditional on understanding the task. To our knowledge, this distinction has not been made in the economics literature. In addition, standard computerized experimental designs do not allow researchers to isolate the components of total response time.

The contribution of this paper is to empirically disentangle comprehension time from deliberation time. We show that both components are analytically valuable but play fundamentally different roles. To do so, we rely on data from a controlled laboratory experiment with high-school students in which the instruction screen and the decision screen are separated. With this simple design tweak, we can observe the time spent understanding the task separately from the time spent making a choice. We link the experimental data to administrative data from centralized academic achievement tests. The resulting dataset contains process data, choice data, and independently measured proxies for cognitive ability.

Three sets of findings emerge. First, our analysis confirms that comprehension time and deliberation time are empirically distinct constructs with substantial variation across subjects. Second, we show that comprehension time can serve as a cost-free proxy for cognitive ability in experiments that lack external measures. Third, we find that deliberation time is independent of cognitive ability, but responds to the structure of the choice. Heuristic choices are faster, and decisions near the indifference point take longer, exactly as sequential sampling models predict.

These findings have direct implications for the use of response time in experimental economics. When instructions and decisions appear on the same screen, a fast response is inherently ambiguous: it may reflect rapid comprehension or a quick, potentially impulsive, choice. The two cannot be disentangled. A simple separation of the instruction and decision screens eliminates this ambiguity at no cost, allowing researchers to recover distinct measures of comprehension and deliberation.

The rest of the paper is structured as follows. In Section 2, we discuss the relevant literature and how our paper connects to it. Section 3 introduces the conceptual framework and research questions guiding the analysis. Section 4 presents the data and methodology. Section 5 reports the results, and Section 6 concludes.

2 Related literature

2.1 Experimental economics

The adoption of process data has enabled economists to study the cognitive processes underlying economic choices. The analysis of response time (RT), the most commonly used process measure, provides insights beyond what can be learned from choice data alone (Clithero, 2018b; Spiliopoulos and Ortmann, 2018; Konovalov and Krajbich, 2019a).

Albeit relatively new, the use of RT data in economic experiments has proven useful in several ways. First, it has been used to interpret the meaning of participants' decisions (Rubinstein, 2016). For example, RT has been used to infer the strength of preferences underlying a choice, based on the assumption that longer deliberation reflects weaker preferences (Chabris et al., 2008; Campbell et al., 2018). Longer response times have been interpreted as signals of cognitive dissonance or decision conflict (Matthey and Regner, 2011; Jiang, 2013; Evans et al., 2015), whereas shorter response times are seen as intuitive choices (Lotito et al., 2013; Cappelen et al., 2016; Rubinstein, 2016). RT has also been used to infer subjective task complexity (Agranov et al., 2025) and the level of effort participants exert to complete tasks (Wilcox, 1993; Ofek et al., 2007).

A closely related strand of the literature uses response time to test theories of heuristic decision-making. The Social Heuristics Hypothesis (Rand et al., 2014) posits that prosocial behavior reflects an automatic default response that can be overridden by deliberation. In this framework, fast choices are interpreted as heuristic and slow choices as deliberate. Cappelen et al. (2016) find that fair choices in a dictator game are significantly faster than selfish ones, interpreting this as evidence that fairness is intuitive. Rubinstein (2016) classifies choices based on response time to distinguish between instinctive and contemplative decision types. Alós-Ferrer and Buckenmaier (2021) show that RT is related to cognitive sophistication, suggesting that differences in response speed across individuals partly reflect differences in cognitive processing rather than their preferences alone. A common challenge in this literature is that fast responses may reflect either heuristic decision-making or rapid comprehension. Without separating these components, the heuristic interpretation of short response times remains ambiguous.

A second strand of the literature uses non-choice measures to predict out-of-sample choices. It appears that RT-based predictions may even outperform choice-based predictions. In a global game experiment, Schotter and Trevino (2021) use RT alone to identify a participant's threshold strategy. They find that its predictive power is superior to predictions based on normative solutions. RT predictions may also be as informative as classical measures of participant preferences, such as risk aversion (Rubinstein, 2013). According to Chabris et al. (2008), the time discount parameters inferred from RT predict behavior almost as well as choice-based measures of time preferences.

Third, and more recently, RT has been used to model traits and abilities. Junghaenel et al. (2022) examine whether the item response latency in surveys can be used to infer respondents' cognitive ability. Linking survey data with cognitive test results, they find that approximately one quarter of the variation in cognitive scores can be explained by measures of average RT and intra-individual RT variation. To our knowledge, Junghaenel et al. (2022) is the only study to use RT data to infer cognitive ability from tools not specifically designed to measure cognition (i.e., not intelligence or achievement tests). In a related vein, Pedersen et al. (2023) validate a game-based cognitive assessment platform that predicts eight cognitive abilities from in-game behavior, including response times, and Pezeshki and Embretson (2025) show that jointly modeling response accuracy and response time improves ability estimation in automated testing. These studies share the premise that RT carries information about cognitive traits, but none separates comprehension from deliberation within the task.

We systematically review this literature in Table 1 in the Appendix, which summarizes 35 studies that use response time in incentivized decision tasks without time constraints. The inventory covers the sample, task, RT measurement, and main findings of each study. A striking pattern emerges: every study records a single, aggregate response time. None separates the time spent understanding the task from the time spent making the choice.

2.2 Cognitive science

The joint analysis of choices and response time has a long tradition in cognitive science and experimental psychology. This literature has typically focused on modeling intelligence and cognition. It relies on simple choice tasks, such as perceptual decisions and memory tests, and works with short response times (on the order of split seconds). Numerous studies show that response time is moderately correlated with measures of cognitive ability, including intelligence, working memory, and academic achievement (Schubert et al., 2015; Doebler and Scheffler, 2016; Kyllonen and Zu, 2016; Schmitz and Wilhelm, 2016; Schubert et al., 2017; Kumar et al., 2020).

A meta-analysis finds that both mean RT and RT variability are negatively correlated with intelligence, although the correlations are typically small to moderate (Doebler and Scheffler, 2016). RT is often considered an indicator of how quickly an individual processes information, a core component of cognitive ability (Kyllonen and Zu, 2016; Kumar et al., 2020). Cohort studies have found stronger correlations between response time and intelligence scores in more complex choice tasks (Deary et al., 2001; Der and Deary, 2017).

We consider a small number of studies at the crossroads between cognitive science and behavioral economics. These studies share with cognitive experimental psychology the feature that the tasks have objectively correct solutions (e.g., cognition tasks, reasoning and achievement tests) and analyze response time jointly with response accuracy. Unlike most economic experiments, they directly elicit abilities rather than preferences. At the same time, they resemble economic experiments in involving longer response times and more complex tasks. Jimenez et al. (2018) administer the Cognitive Reflection Test (CRT) (Frederick, 2005), a widely used empirical tool in research on dual-process theories pioneered by Kahneman (2011), and finds that fast answers on the CRT tend to be incorrect. Chen et al. (2018) find a curvilinear relationship between response time and accuracy on a series of academic achievement and reasoning tests.

Cognitive science also offers theoretical approaches to response time. Clithero (2018b) argues that economic analyses lack a theoretical foundation, and that sequential sampling models from cognitive psychology could help fill that gap. In particular, the Drift-Diffusion Model (DDM) (Ratcliff, 1978, 2013) was developed to jointly model the speed-accuracy trade-off and the efficiency of information processing. In the model, decisions emerge from a noisy process of evidence accumulation that continues until a threshold is reached.

Importantly, the DDM operationalizes total response time as the sum of decision time and non-decision time. Clithero (2018b, p. 63) explicitly notes a limitation: non-decision time is typically neither observed nor analyzed in the economic literature and is instead treated as a nuisance parameter or assumed to be constant or uniformly distributed (Ratcliff and Tuerlinckx, 2002). This distinction between the two components of total response time will be the focus of our empirical analysis.

2.3 Positioning this study in the literature

This paper contributes to the literature using response time as a behavioral measure in incentivized decision tasks. Our contribution is to empirically separate two components of response time and to show that each is informative in its own right. Our approach is related to Junghaenel et al. (2022), who infer cognitive ability from survey response latencies. Our approach differs in two important ways: we use incentivized economic experiments rather than survey items, and we isolate comprehension from deliberation through experimental design rather than statistical modeling.

Our work also connects to Alós-Ferrer and Buckenmaier (2021), who show that response times are systematically related to cognitive sophistication, not only the strength of preferences. Similarly, Clithero (2018a) notes that non-decision time (the sequential-sampling analogue of what we call comprehension time) is rarely observed or analyzed in the economic literature. Cappelen et al. (2016) observe that response time carries information about the meaning of choices, but that this relationship is blurred by cognitive ability. As our literature inventory confirms (Table 1 in the Appendix), no existing study separates the time spent reading instructions from the time spent choosing.

The decomposition also allows us to revisit two well-established empirical literatures using a sharper measure. A large body of work has tested the Social Heuristics Hypothesis (Rand et al., 2014), which predicts that prosocial choices are fast and selfish choices are slow, using total response time as the main outcome measure (Rand et al., 2012; Cappelen et al., 2016; Rubinstein, 2016). Because total response time includes comprehension, these tests conflate heuristic speed with reading speed. Our decomposition isolates the deliberation component and provides a cleaner test of whether heuristic choices are genuinely faster to *decide*, not merely faster to *read and process*. Similarly, a growing literature documents that response time increases as decision difficulty rises. For instance, as options approach the indifference point in intertemporal choice, it should take longer to decide (Chabris et al., 2008; Krajbich et al., 2015; Konovalov and Krajbich, 2019b). These findings are consistent with sequential sampling models but have so far been tested using total response time. We test whether the difficulty–response time relationship survives when comprehension is separated from deliberation time.

We demonstrate the practical value of the decomposition by revisiting the relationship between response time and fairness in a dictator game, studied by Cappelen et al. (2016). The original paper finds that participants who choose the fair allocation are significantly faster. The authors note that total response time conflates different processes and call for a decomposition. Our split-screen design makes such a decomposition possible.

By decomposing response time, we aim to consolidate the different and sometimes conflicting interpretations interpretations of response time found in the literature. Total response time has been interpreted as reflecting deliberation effort, preference strength, cognitive ability, decision conflict, and heuristic processing. We show that these interpretations are not contradictory. Rather, they are driven by different components of the same aggregate measure.

3 Conceptual framework

We conceptualize observed response time in economic experiments as the sum of two distinct cognitive processes: task comprehension and choice deliberation. Let total response time for individual i be given by

$$RT_i = CT_i + DT_i, \tag{1}$$

where CT_i denotes comprehension time (the time required to read and understand the decision problem) and DT_i denotes deliberation time (the time spent forming a choice conditional on understanding the task).

Comprehension time captures the cognitive work required to read, parse, and understand the decision problem: mapping textual instructions into an internal representation of the task, identifying feasible actions, and understanding payoff consequences. Deliberation time captures the cognitive process that follows comprehension: evaluating alternatives, resolving trade-offs, and forming a choice.

This decomposition raises the following three research questions:

RQ1: Are comprehension time and deliberation time empirically distinct constructs? Our starting point is a simple observation: when researchers interpret total response time as a measure of deliberation, they implicitly assume that the time required for task comprehension is homogeneous across participants. In other words, it is assumed that inter-individual differences in response time are driven entirely by differences in decision-making. To our knowledge, this assumption has never been tested. It is, however, unlikely to hold. People differ in reading speed, cognitive ability, and familiarity with experimental interfaces. All of these factors affect how quickly they process instructions. If comprehension time varies across individuals, total response time conflates two distinct sources of heterogeneity, and its interpretation becomes ambiguous. If CT_i and DT_i capture distinct cognitive processes, they should not be reducible to a single latent factor.

RQ2: What does comprehension time capture? If comprehension time captures the cognitive work of processing instructions, it should reflect individual differences in cognitive ability. People differ in reading speed, working memory, and information processing speed. These skills are central components of cognitive ability. Individuals with higher cognitive ability should therefore understand experimental tasks more quickly. If comprehension time reflects these skills, it should be negatively associated with independently measured cognitive ability.

RQ3: What does deliberation time capture? Existing theory provides further predictions about what deliberation time should reflect. Dual-process theories and the Social Heuristics Hypothesis (Rand et al., 2014) predict that heuristic choices (e.g., focal allocations such as sending nothing, exactly half, or everything) are fast, while intermediate choices require more deliberation. Because total response time conflates decision speed with reading speed, deliberation time should reveal these patterns more sharply: focal choices should be associated with shorter deliberation times than intermediate choices.

Sequential sampling models predict that deliberation time increases as the difference in subjective value between alternatives decreases (Ratcliff, 1978; Krajbich et al., 2015). Near a participant's indifference point, the evidence accumulation process has a lower drift rate and therefore takes longer to reach a threshold. Deliberation time should therefore increase as the offered amount approaches the participant's indifference point. We test this using the adaptive staircase structure of the time preference tasks, which generates within-person variation in the distance between the offered amount and the participant's indifference point.

4 Data and methodology

4.1 Experimental setting and data

To address these questions, we use data from a large-scale lab-in-the-field experiment carried out in Hungarian upper secondary schools between March 2019 and March 2020. The study included 1,108 students from 53 classes in nine schools. Sessions took place during regular school hours and lasted 45 minutes, corresponding to the length of a standard school lesson in Hungary. It is important to note that the experiment was originally designed to study preferences (Horn et al., 2022). The present analysis of response time is post-hoc, exploiting the split-screen feature retrospectively.

All tasks were computerized and implemented in z-Tree (Fischbacher, 2007). Participation was voluntary and anonymous. Because the participants were minors, informed consent was obtained from both students and their parents. Payments were made in the form of cafeteria vouchers. At the end of each session, one incentivized task and one decision within that task were randomly selected for payment. This payment rule preserved incentive compatibility while keeping expected earnings broadly comparable across participants. Average earnings were approximately 1,000 HUF, roughly enough to cover the price of a typical school lunch at the time.

Participants completed eight incentivized experimental tasks spanning time preferences, risk attitudes, social preferences, and competitiveness. We briefly summarize the tasks below; full instructions, payoff rules, and implementation details are reported in Horn et al. (2022). Appendix D contains all instruction and decision screens seen by participands.

Time preferences. Two tasks measure intertemporal choice following the approach of Falk et al. (2018). In each task, participants make five sequential binary choices using a staircase procedure (Cornsweet, 1962). In Task 1, they choose between an immediate payment of 1,000 HUF and a larger payment received two weeks later. Later in the experiment, in Task 6, they choose between payments received after four and six weeks. The monetary amounts used at the two horizons are identical, so the tasks differ only in timing. This design allows us to separately recover a longer-run discount factor, δ, from the four-versus-six-week task and to combine it with the immediate-versus-two-week task to construct a proxy for present bias, β. After each choice, the later amount is adjusted by bisection: choosing the earlier payment increases the next delayed offer, whereas choosing the later payment reduces it. The procedure is designed to converge to the individual's indifference point. The indifference point is defined as the delayed amount at which the participant is approximately indifferent between the two options.[1]

Risk preferences. Risk preferences were measured in Task 4 using the Bomb Risk Elicitation Task (BRET) (Crosetto and Filippin, 2013). Participants were shown 100 boxes, one of which contained a randomly placed "bomb." They decided how many boxes to remove. Each box removed earned a fixed payment. If the bomb was among the selected boxes, the entire payoff was lost. Risk tolerance is proxied by the number of boxes removed.

Social preferences. Social preferences, covering altruism, cooperation, trust, and trustworthiness, were elicited in five tasks. In Task 2, altruism was measured using a dictator game in which participants allocated part of an endowment to a randomly selected classmate from the same session. In Task 3, participants completed a second dictator game with a randomly selected schoolmate, that is, another student from the same school but from a different class. This was the only non-incentivized task. In both dictator games, altruism is measured by the share of the endowment transferred to the recipient.

Cooperation was measured in Task 5 using a two-player public goods game. Each participant received an endowment and decided how much to contribute to a common pool. The marginal per capita return was 75%, and cooperativeness is measured by the share of the endowment contributed.

Trust and trustworthiness were measured in Tasks 7 and 8 using the standard trust game of Berg et al. (1995). In the first stage, each participant received an endowment and decided how much to send to a randomly selected peer in the room. The transferred amount was tripled before being received by the other participant. Trust is measured by the share of the endowment sent. In the second stage, trustworthiness is measured by the share of the tripled amount returned to the sender. This stage was implemented using the strategy method: participants specified return proportions for all possible amounts

[1]The indifference point is estimated as the midpoint between the last accepted and the last rejected offer after five bisection steps. For example, if the initial offer is 1,500 HUF and the participant always chooses the later option, the offers converge downward and the indifference point is estimated near the lower bound of the interval.

they might receive.

The experiment also included a real-effort competitiveness task based on the design of Niederle and Vesterlund (2007), but it is excluded from this analysis since it does not measure preferences.

Table 1 reports descriptive statistics on response time by game. Participants spent more time on the instruction screen than on the decision screen: the median comprehension time across the six split-screen games was 255 seconds, whereas the median deliberation time was 157 seconds.

Table 1: Descriptive statistics by game

	Total RT		Comprehension		Deliberation	
	Mean	Median	Mean	Median	Mean	Median
Total (6 games)	444	417	269	255	175	157
Time pref 1	38	35	20	20	18	15
Time pref 2	39	35	17	15	23	20
Risk	71	68	57	55	13	9
Trust (amount sent)	60	53	50	45	10	6
Trustworthiness (amount returned)	154	137	50	45	104	90
Cooperation (public goods)	83	76	76	70	7	5
Altruism (dictator†)	38	35				

Notes: All times in seconds. "Total (6 games)" aggregates the six split-screen tasks: two time preferences, risk, trust, trustworthiness, and cooperation. †The dictator game uses a single screen; comprehension and deliberation times cannot be separated. Only total response time is reported.

The experimental data are matched at the individual level to administrative records from the National Assessment of Basic Competencies (NABC) (Balázsi and Ostorics, 2020), an assessment of full cohorts of secondary school students conducted since 2008. These records provide standardized mathematics and reading test scores, demographic characteristics (age and gender), and detailed measures of family background and socioeconomic status. The test scores were collected before the experiment and are standardized to have mean zero and unit variance.[2] Because they predate experimental participation, they serve as independent proxies for cognitive ability and cannot be affected by experimental behavior or response times.

4.2 Separating comprehension and deliberation time

A key feature of the experimental design is the strict separation between the instruction screen and the decision screen in six tasks: the two time preference tasks, the risk task, the public goods game, and the two stages of the trust game. This separation was not implemented for the dictator games. The participants first viewed the instructions describing the task and proceeded to the decision screen only after actively indicating that they were ready to continue.[3] There were no time limits on either screen. This design allows us to observe two distinct components of response time for each decision:

- Comprehension time: the time spent on the instruction screen.
- Deliberation time: the time spent on the decision screen before making a choice.

Total response time is the sum of these two components. Because the instructions are identical for all participants within a given task, variation in comprehension time should primarily reflect individual differences in processing and understanding rather than differences in task content.

[2] We use test scores from the 6th grade NABC, that is, 3 to 7 years before the experiment depending on subjects' cohorts.
[3] See each screen as it appeared to participants in Appendix D.

4.3 Identification strategy

Factor structure of comprehension and deliberation time. To test whether comprehension time and deliberation time reflect distinct latent constructs (RQ1), we perform a factor analysis of the item-level CT and DT measures across games. If comprehension time and deliberation time capture the same underlying process, all items should load on a single factor. If the two are distinct, the CT items should load on one factor and the DT items on another.

Using maximum likelihood, we estimate two models: a one-factor model in which all items load on a common latent factor, and a two-factor model with separate Comprehension and Deliberation factors. The indicators are five CT items and six DT items drawn from the six split-screen tasks: the two time preference tasks, the bomb risk task, the public goods game, and the send and return stages of the trust game.[4] We compare the models using standard fit indices (RMSEA, CFI, and SRMR) and information criteria (AIC and BIC).

Cognitive ability and response times. We denote observed comprehension and deliberation times as CT_i and DT_i. To test how these two components are associated with cognitive ability (relevant for RQ2 and RQ3), we aggregate CT_i and DT_i across the six split-screen games (the two time preference tasks, the bomb risk task, the send and return stages of the trust game, and the public goods game) and estimate the following specification:

$$\text{Ability}_i = \alpha + \beta\,\text{CT}_i + \delta\,\text{DT}_i + \mathbf{X}_i'\gamma + \mu_c + \varepsilon_i \tag{2}$$

where Ability_i is the standardized mathematics or reading test score of student i; CT_i and DT_i are total comprehension and deliberation times aggregated across the six split-screen tasks; $\mathbf{X}_i$ is a vector of individual-level controls (parental education, socioeconomic status, whether the household receives social assistance, age, and gender); and μ_c denotes class fixed effects.[5] Standard errors are clustered at the class level. If comprehension time captures the cognitive work involved in understanding a task, we expect $\beta < 0$: students who process instructions more slowly should score lower on ability tests. If deliberation time captures a different process, namely forming a choice rather than understanding the problem, we expect $\delta \approx 0$. We also estimate a restricted specification in which CT_i and DT_i are replaced by total response time, to test whether the decomposition adds explanatory power beyond the standard approach.

Deliberation time and heuristic decision-making. To examine whether the decomposition reveals sharper behavioral patterns, we present descriptive comparisons of mean deliberation time and mean comprehension time across choice categories in the trust game (send and return stages) and the public goods game. We focus on these games because they feature discrete choice categories that map naturally onto theoretical predictions from dual-process models. Focal choices (e.g., sending nothing, half, or everything) are predicted to be fast, heuristic responses, whereas intermediate choices are more likely to reflect deliberation over trade-offs (Rubinstein, 2016; Rand et al., 2014). Comparing DT and CT side by side allows us to assess whether these behavioral patterns are driven primarily by the deliberation component rather than by differences in comprehension speed.

Deliberation time and decision difficulty. Sequential sampling models predict that deliberation time should increase with decision difficulty. We test this prediction for DT by exploiting the adaptive

[4] The trust game uses a single instruction screen for both the send and return stages, so CT is observed only once. We therefore use a single CT indicator for the trust game and two separate DT indicators, one for sending and one for returning, yielding five CT items and six DT items in total.

[5] Class fixed effects control for differences in experimental conditions, such as time of day or room size, as well as differences in student composition across classes. Accordingly, all results are identified from within-class variation.

staircase structure of the two time preference tasks. Each task consists of five sequential binary choices in which the offered amounts converge toward the student's indifference point. By design, later questions tend to be more difficult because the offered amount is closer to the threshold at which the student is indifferent. We estimate:

$$\mathrm{DT}_{ijq} = \alpha + \varphi\,\mathrm{Distance}_{ijq} + \mu_{ij} + \lambda_q + \varepsilon_{ijq} \tag{3}$$

where DT_{ijq} is the deliberation time of student i in task j on question q. $\mathrm{Distance}_{ijq} = |\text{Offered amount}_{ijq} - \text{Indifference point}_{ij}|$ measures how far the offer is from the student's indifference point. μ_{ij} are individual×task fixed effects that absorb all time-invariant individual- and task-level heterogeneity, and λ_q are question-order fixed effects that absorb any common learning pattern across the five sequential decisions. The prediction is $\varphi < 0$: holding question position constant, offers closer to the indifference point should require more deliberation because the drift rate in the evidence accumulation process is lower.

Response time decomposition and social preferences. Finally, we illustrate the practical value of the decomposition by revisiting a prominent finding in the literature. Cappelen et al. (2016) study response time in a dictator game and find that participants who choose the fair allocation respond significantly faster than those who keep everything for themselves. Recognizing that response time conflates comprehension and deliberation, they control for cognitive ability and for a proxy for processing speed. Our split-screen design allows us to implement the decomposition they call for more directly. The dictator game in our experiment presents the instructions and the decision on a single screen.[6] As a result, we cannot decompose response time in this task into comprehension and deliberation components. At the same time, this feature allows us to replicate the design of Cappelen et al. (2016) closely. We therefore follow their approach, but replace their external proxies for cognitive ability and processing speed with aggregate comprehension time measured in the six split-screen games.

We estimate:

$$RT_i^{\mathrm{dict}} = \alpha + \beta\,\mathrm{Choice}_i + \theta\,\overline{CT}_i + \mathbf{X}_i'\gamma + \mu_c + \varepsilon_i, \tag{4}$$

where RT_i^{dict} is response time on the dictator game screen (in seconds); Choice_i is either a binary indicator for choosing the fair allocation (an equal split) or the continuous share sent (0–100%); $\overline{CT}_i$ is aggregate comprehension time from the six split-screen games (in minutes); $\mathbf{X}_i$ includes demographic and family-background controls (gender, age, parental education, socioeconomic status); and μ_c denotes class fixed effects. In the cognitive-control specification, $\mathbf{X}_i$ additionally includes standardized mathematics and reading scores, GPA, and school grades, following Cappelen et al. (2016). We estimate several specifications, first restricting the sample to fair (50%) and selfish (0%) participants, and then extending the analysis to the full sample using the continuous measure of dictator giving.

Table 2 provides a summary of the research questions, estimation methods, and corresponding predictions.

[6]This was not a deliberate design choice, but rather reflected the brevity of the task and the simplicity of the required response, which did not warrant a separate instruction screen. Ex post, however, it proves advantageous because it replicates the single-screen design used by Cappelen et al. (2016).

Table 2: Summary of research questions, methods, and predictions

Research question	Method / specification	Prediction
RQ1: Are CT and DT distinct constructs?	1. Factor analysis of item-level CT and DT	1. Two factors: one for CT, another for DT
RQ2: What does CT capture?	1.$\text{Ability}_i = \alpha + \beta CT_i + \delta DT_i + X_i'\gamma + \mu_c + \varepsilon_i$ 2. Cappelen-style within-game replication	1. $\beta < 0$ 2. CT attenuates the fairness–RT link
RQ3: What does DT capture?	1. Same ability equation 2. Bar charts of DT, CT, RT by choice category 3. $DT_{ijq} = \alpha + \varphi\,\text{Distance}_{ijq} + \lambda_q + \varepsilon_{ijq}$	1. $\delta \approx 0$ 2. Focal DT lower 3. $\varphi < 0$

5 Results

We begin with the first research question: are comprehension time and deliberation time empirically distinct? In the economic literature, response time is typically interpreted as a proxy for deliberation. This implicitly assumes that the comprehension component is homogeneous across participants. Figure 1 shows that this assumption is misleading: comprehension time varies substantially across individuals and is strongly right-skewed. Figure 2 plots CT against DT. The two components are positively but only moderately correlated (Spearman's $\rho = 0.57$). For any level of DT, there is substantial heterogeneity in CT. Appendix Figure B1 shows that stratifying students into CT quintiles, the within-quintile spread in DT is comparable in magnitude to the between-quintile spread.

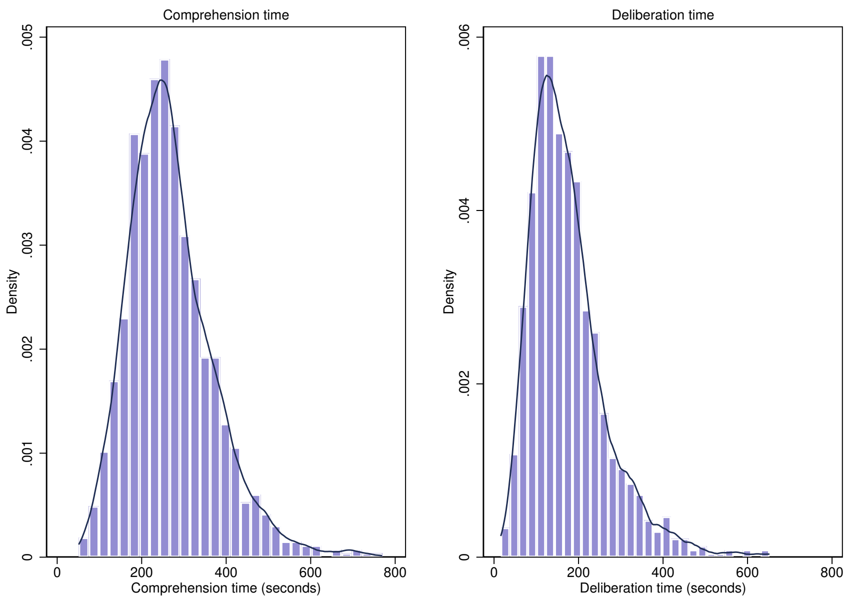

Figure 1: Distribution of CT_i and DT_i across individuals

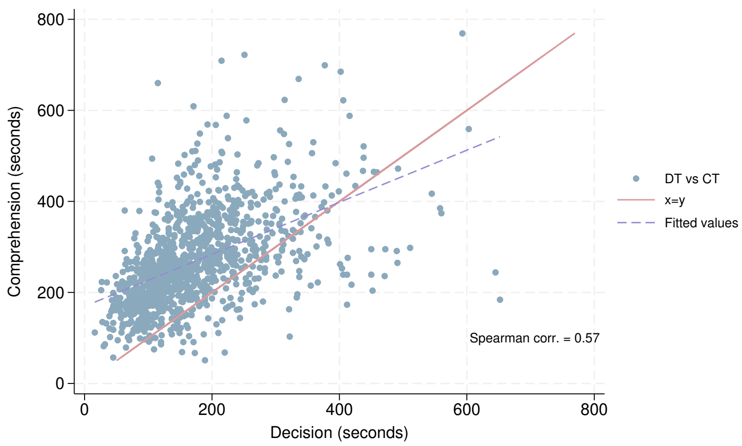

Figure 2: Scatterplot of CT_i and DT_i

We test this more formally using a factor analysis implemented as a structural equation model (SEM). We estimate two models: a one-factor model in which all 11 items (5 CT, 6 DT) load on a single "Response Time" factor, and a two-factor model in which CT items load on a Comprehension factor and DT items load on a Deliberation factor, with the two factors allowed to correlate freely. The two-factor model fits the data substantially better on every criterion: RMSEA declines from 0.092 to 0.072, CFI increases from 0.852 to 0.913, SRMR declines from 0.059 to 0.049, and AIC falls by 169 points (Table 3).

In the two-factor model, all CT items load strongly on the Comprehension factor (standardized loadings 0.54–0.75). The DT loadings on the Deliberation factor are more heterogeneous: the two time preference tasks load strongly (0.71 and 0.74), the return stage of the trust game loads moderately (0.49), but the bomb risk task (0.23) and the public goods game (0.22) load weakly, suggesting that deliberation time in these relatively short tasks contains more noise. The estimated correlation between the two latent factors is 0.74. This is higher than the raw Spearman ρ of 0.57, as the SEM corrects for attenuation due to measurement error. A correlation of 0.74 implies that the two constructs share roughly half of their variance: they are related, but not reducible to a single factor. The shared variance is plausible due to a latent "speed" trait.

Table 3: Factor analysis: standardized loadings and model fit

Item	One-factor (RT)	Two-factor (CT + DT)
Comprehension items		
CT: Time pref 1	0.503	0.537
CT: Time pref 2	0.580	0.599
CT: Bomb risk	0.635	0.678
CT: Trust	0.729	0.749
CT: Public goods	0.664	0.663
Deliberation items		
DT: Time pref 1	0.605	0.711
DT: Time pref 2	0.620	0.740
DT: Bomb risk	0.198	0.228
DT: Trust send	0.390	0.384
DT: Trust return	0.456	0.495
DT: Public goods	0.198	0.224
Factor correlation		
Corr(C, D)	—	0.745
Fit indices		
RMSEA	0.092	0.072
CFI	0.852	0.913
SRMR	0.059	0.049
AIC	32,255	32,086
BIC	32,420	32,257

Notes: $N = 1{,}108$. Standardized coefficients from maximum likelihood SEM estimation. All loadings significant at $p < 0.001$. The one-factor model restricts all items to load on a single latent factor; the two-factor model assigns CT items to a Comprehension factor and DT items to a Deliberation factor.

5.1 Comprehension time

Comprehension time and test scores. Having established that *CT* and *DT* are empirically distinct, we next ask what each component captures. Table 4 reports estimates of Equation 2. Columns (1) and (4) use total response time as a single predictor, whereas columns (2) and (5) decompose it into comprehension time and deliberation time. Columns (3) and (6) add individual controls. Consistent with a large literature on mental chronometry (Jensen, 2006), total response time is negatively and significantly associated with standardized test scores. Students who take longer across all experimental tasks score lower on both the mathematics test ($\hat{\beta} = -0.00033$, $p < 0.05$) and the reading test ($\hat{\beta} = -0.00040$, $p < 0.01$).

However, this aggregate relationship masks an important asymmetry. When we replace total response time with its two components, comprehension time is negative and statistically significant in both equations ($p < 0.05$ for math, $p < 0.01$ for reading), while deliberation time is indistinguishable from zero in all specifications.

Adding individual controls leaves the estimates essentially unchanged (columns 3 and 6), and the pattern is robust to a log specification (Appendix Table A1).[7] This directly answers RQ2 and RQ3: the relationship between cognitive ability and response time appears to be driven entirely by comprehension time.

The effect is notably stronger for reading than for math. In the specifications with individual controls, the CT coefficient for reading ($\hat{\beta} = -0.058$, $p < 0.01$, column 6) is nearly twice the size of the CT

[7] Allowing *CT* and *DT* to interact yields a non-significant interaction term for reading and a marginally significant one for math ($\beta = 0.013$, $p = 0.04$); see Appendix Table A2. Appendix Figure B2 reports binscatters of *CT* and *DT* against the two test scores, absorbing class fixed effects.

coefficient for math ($\hat{\beta} = -0.034$, $p < 0.05$, column 3). This asymmetry is intuitive: comprehension time in our experiment fundamentally reflects a reading task. Students read and parse textual instructions describing the decision problem. The cognitive skills that determine how quickly a student reads and understands a text overlap substantially with those measured by a standardized reading test.

Columns (4) and (8) provide the most demanding test of this interpretation. Column (4) adds reading scores as a control to the math equation. The CT coefficient shrinks and loses significance, indicating that the relationship between CT and mathematics scores is largely mediated by reading ability. Column (8) performs the reverse test by adding mathematics scores as a control to the reading equation. Here, CT remains significant ($\hat{\beta} = -0.037$, $p < 0.05$). Even after accounting for mathematical ability, students who take longer to process experimental instructions score lower on the reading test. This is perhaps the most telling result in the table: comprehension time appears to capture a reading-specific component of cognitive ability that is not reducible to general intelligence or mathematical skills.

Table 4: Response time and cognitive ability

VARIABLES	(1) Math score 6th grade	(2) Math score 6th grade	(3) Math score 6th grade	(4) Math score 6th grade	(5) Reading score 6th grade	(6) Reading score 6th grade	(7) Reading score 6th grade	(8) Reading score 6th grade
Total CT		-0.0384*	-0.0340*	-0.00470		-0.0579***	-0.0582***	-0.0371**
		(0.0196)	(0.0186)	(0.0155)		(0.0185)	(0.0177)	(0.0148)
Total DT		-0.00268	0.00711	0.00210		0.0165	0.00998	0.00556
		(0.0221)	(0.0214)	(0.0164)		(0.0204)	(0.0204)	(0.0153)
Total RT	-0.0206*				-0.0208**			
	(0.0110)				(0.0101)			
Reading score				0.503***				
				(0.0256)				
Math score								0.621***
								(0.0322)
Constant	0.153*	0.181**	0.186	-0.127	0.155**	0.213***	0.622	0.507
	(0.0818)	(0.0819)	(0.688)	(0.542)	(0.0752)	(0.0758)	(0.914)	(0.732)
Observations	1,083	1,083	1,081	1,081	1,083	1,083	1,081	1,081
R^2	0.571	0.571	0.592	0.719	0.471	0.474	0.495	0.653
Class FE	YES	YES	YES	YES	YES	YES	YES	YES
Indiv. controls	NO	NO	YES	YES	NO	NO	YES	YES

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Indiv. controls: family background, age, gender

These results are consistent with the cognitive science literature, which has long documented that faster information processing is associated with higher cognitive ability (Jensen, 2006; Junghaenel et al., 2022). Comprehension time captures the speed at which students encode novel task instructions. It is conceptually similar to the non-decision time parameter in drift-diffusion models (Ratcliff, 1978).

Comprehension time as a cost-free ability control. A number of authors have called for caution when interpreting RT differences without an explicit ability adjustment (Clithero, 2018a; Spiliopoulos and Ortmann, 2018). Cappelen et al. (2016) confront this concern directly in a dictator-game setting and address it by including a "swiftness" proxy and Raven's progressive matrices as ability controls. In this section, we show that the split-screen decomposition offers a within-experiment alternative. Comprehension time captures the ability-driven component of total RT and can serve as a cost-free ability proxy in any experiment with split screens.

We illustrate this by revisiting Cappelen et al. (2016), who find that participants who choose the fair allocation (an equal split) in the dictator game respond approximately 0.45 standard deviations faster than those who keep everything. They note that total response time conflates reading the problem, making a decision, and implementing the choice. Their single-screen design does not allow them to separate these components.

Our split-screen design provides exactly the decomposition that Cappelen et al. (2016) call for. We replicate their analysis using the dictator game in our data, in which students choose what share of an endowment (0–100%) to send to a classmate. Because the dictator game in our experiment presents the instructions and the decision on a single screen, we cannot decompose response time within this game into comprehension and deliberation components. We can, however, replace their external ability proxies with aggregate comprehension time from the six split-screen games.

Following Cappelen et al. (2016), we construct a binary indicator equal to one for students who chose the fair allocation (50%) and zero for those who kept everything (0%). We exclude intermediate choices. In our sample, 614 students (88%) chose the fair allocation and 87 (12%) chose the selfish allocation (total $N = 701$).

Table 5 reports the results under three ability-control specifications. Column (1) includes only class fixed effects, with no ability proxy. Fair choices are 5.5 seconds faster than selfish ones, but the difference is not statistically significant ($p = 0.106$). Column (2) adds the cognitive battery along with demographic and family-background controls.[8] The fair coefficient strengthens to -6.5 seconds and becomes marginally significant ($p = 0.079$), closely resonating with the original Cappelen findings. Column (3) replaces the cognitive battery with our aggregate comprehension time from the six split-screen games.[9] The fair coefficient is essentially unchanged in magnitude (-6.7) but is now significant at the 5% level ($p = 0.023$). Precision sharpens: the standard error falls by approximately 20% relative to column (2).

Two patterns stand out. First, ability is a substantive confound in this setting: without controlling for it, the heuristic-speed finding is statistically invisible in our data. This is analogous to what Cappelen et al. (2016) find before introducing their ability proxies. Second, comprehension time is at least as effective an ability control as a full cognitive test battery. This is consistent with the interpretation that CT captures cognitive processing as it operates during the experiment itself, whereas the cognitive battery (collected three to seven years before the experiment) is a stale proxy that introduces noise.

The qualitative pattern is preserved when we generalize from the binary comparison to the full sample using the continuous dictator choice (0–100%), as shown in Appendix Table A3. Appendix Table A3 also reports a horse-race specification that includes the cognitive battery and CT together. CT remains highly significant ($t \approx 9.5$). Appendix Figure B3 visualizes the aggregate relationships between dictator

[8] Standardized mathematics and reading scores, GPA, and school grades.

[9] In other words, all games except the dictator game.

giving and the two RT components.

Table 5: Dictator game response times

	(1) No controls	(2) Cognitive batt.	(3) Comprehension time
Fair choice (0/1)	−5.49	−6.48*	−6.71**
	(3.33)	(3.62)	(2.87)
Comprehension time (6 games, min)			6.42***
			(0.67)
Reading score		−3.11**	
		(1.21)	
Hungarian grade		2.93**	
		(1.37)	
Cognitive test battery	NO	YES	NO
Family / demographic controls	NO	YES	YES
Class fixed effects	YES	YES	YES
Observations	701	684	686
R^2	0.14	0.18	0.37

Notes: Dependent variable is response time on the dictator game screen (seconds). Sample restricted to subjects choosing the fair allocation (50%) or the selfish allocation (0%). Robust standard errors clustered at the class level in parentheses. Cognitive test battery includes standardized mathematics and reading test scores, GPA, and school grades. Family / demographic controls include gender, age, parental education, socioeconomic status, and family-support indicators. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

5.2 Deliberation time

Deliberation time showed no significant relationship with test scores (Table 4). This null result suggests that the time students spend making a choice is largely independent of cognitive ability. In other words, comprehension time carries information about ability, while deliberation time does not. What does deliberation time capture?

Heuristic decision-making. A central use of response time in behavioral economics is to test dual-process theories. A prominent example is the Social Heuristics Hypothesis by Rand et al. (2012, 2014), which argues that cooperation is the intuitive default and should therefore be faster than non-cooperative choices. The empirical evidence cited in support of this hypothesis is almost entirely based on total response time, and subsequent replications and reanalyses have raised concerns that the "fast cooperator" finding may reflect sorting on individual processing speed rather than a structural property of cooperation itself (Tinghög et al., 2013; Bouwmeester et al., 2017; Recalde et al., 2018a). Cappelen et al. (2016) formalized the same concern in a dictator-game setting: their single-screen design could not separate how much of a "fast fair-choice" effect was attributable to the choice and how much to the person, and they could only address this with external ability proxies. Alós-Ferrer and Buckenmaier (2021) argue that RT differences across choice categories can reflect heterogeneity in depth of reasoning.

The decomposition introduced in this paper allows us to empirically distentangle the decision-stage time DT from the instruction-stage time CT. We can thus ask separately whether choice categories differ in deliberation and whether they differ in comprehension. Sorting on cognitive ability would manifest in the latter; structural heuristic processing in the former. Total RT collapses the two.

Figure 3 shows mean DT (left column) and mean CT (right column), with 95% confidence intervals, for each choice category in the trust and public goods games.

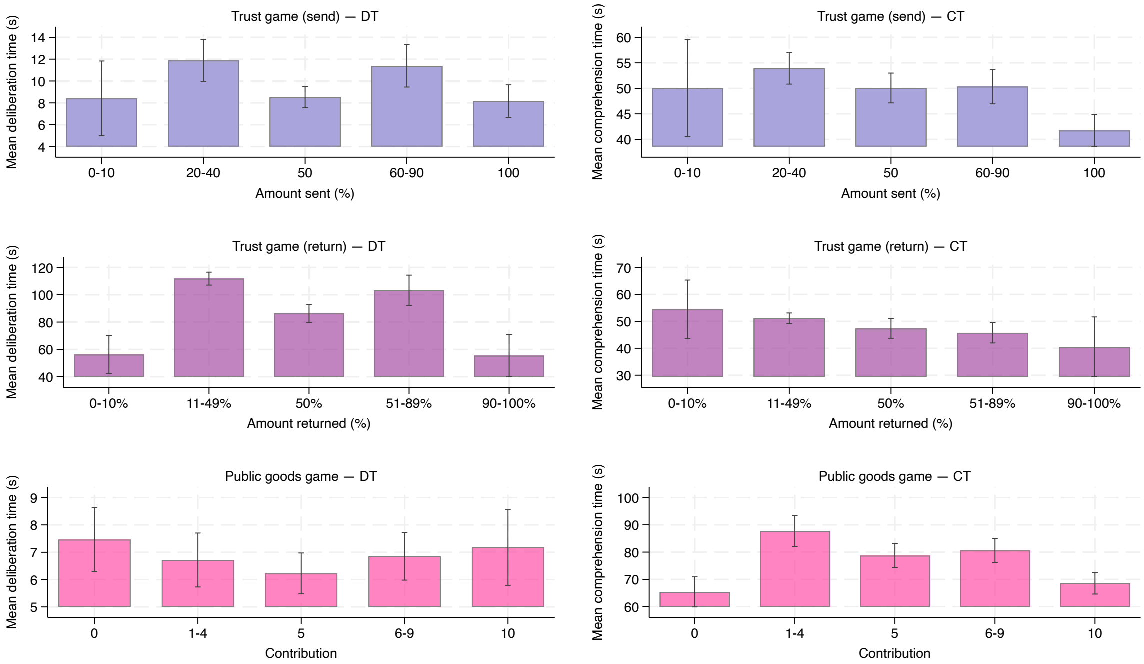

Figure 3: Mean deliberation time (left column) and mean comprehension time (right column) by choice category, with 95% confidence intervals. Detailed descriptive statistics are reported in Appendix Tables B4-6.

Trust game. In both stages of the trust game, the DT panel exhibits the M-shape predicted by dual-process theory and the Social Heuristic Hypothesis.[10] Focal choices are processed faster, consistent with the view that they are selected as defaults while intermediate amounts require weighing strategic uncertainty (Rubinstein, 2016; Caplin and Martin, 2016; Krajbich et al., 2015).

The CT panels are largely flat, but subjects who send the full endowment complete comprehension noticeably faster in both stages. This is precisely the sorting confound that reanalyses of Rand et al. (2012) have flagged: total RT for full senders pools a heuristic effect in DT with a comprehension-speed effect in CT (Recalde et al., 2018b; Bouwmeester et al., 2017). The decomposition makes the two effects separately visible. A researcher with only total RT would attribute all of the "extreme-trust = fast" finding to heuristic processing. In this case, much of the speed gap is attributable to faster comprehension.

Public goods game. This game illustrates both the heuristic prediction and the sorting confound side by side. In DT, the pattern is consistent with the Social Heuristics Hypothesis (Rand et al., 2014): contributing half the endowment (mean 6.2 seconds) and the full endowment (7.2 seconds) are faster than free-riding (7.5 seconds), consistent with the view that cooperation operates as a default heuristic while free-riding requires a more deliberate deviation from the prosocial norm (Rand et al., 2012).

The CT panel reveals a strikingly different pattern. Free-riders and full contributors have the shortest comprehension times (65 and 69 seconds), whereas intermediate contributors are substantially slower (80–88 seconds, with non-overlapping confidence intervals). Unlike the DT pattern, this variation does not match any dual-process prediction. The most plausible interpretation is sorting by cognitive ability:

[10] The return stage of the trust game uses the strategy method: students simultaneously specify return percentages for each of 11 possible received amounts on a single screen (see the experimental instructions in the appendix). Deliberation times are therefore substantially longer than in single-decision games (median 90 seconds versus 5–10 seconds). The within-game comparisons remain valid because all choices are made under the same task structure.

RQ2 established that faster comprehenders have higher cognitive ability, and the literature documents that cognitive ability predicts strategic behavior (Proto et al., 2019; Burks et al., 2009). If higher-ability students both read faster and gravitate toward clear-cut choices (full cooperation or full free-riding), CT varies across choice categories without reflecting deliberation. The decomposition cleanly separates the two stories: the heuristic prediction is confirmed in DT, and the ability-sorting confound manifests in CT.

Taken together, the three games make a single methodological point. Total response time conflates ability-driven variation in comprehension with process-driven variation in deliberation, and the two can move in the same or different directions across choice categories. This is what the reanalysis literature (Tinghög et al., 2013; Bouwmeester et al., 2017; Recalde et al., 2018a) suspected was driving the original "cooperators are fast" result, and what Cappelen et al. (2016) could only address with external ability proxies. Our split-screen design provides the within-experiment decomposition these concerns called for.

The next subsection turns to the related caution of Alós-Ferrer and Buckenmaier (2021): that RT differences across choice categories may reflect the depth of reasoning involved. We show that deliberation time indeed responds to within-person variation in decision difficulty.

Deliberation time and decision difficulty. The bar charts above have shown that deliberation time varies with the type of choice. But does this variation merely reflect a stable individual trait, with some people simply making decisions faster than others? To answer this question, we exploit the staircase design of the time preference tasks and test whether deliberation time responds to decision difficulty within individuals. Recall that each time preference task presents five sequential binary choices whose offered amounts converge toward the student's indifference point (see Section 4 for details of the bisection procedure). Questions that come later in a task tend to be harder because the offered amount is closer to the threshold at which the student is indifferent between the two options.

According to the literature on intertemporal choice, RT is longest when options are close in discounted value and shortest when one option is clearly preferred (Chabris et al., 2008, 2009; Rodriguez et al., 2014; Dai and Busemeyer, 2014; Krajbich et al., 2015; Konovalov and Krajbich, 2019b). These patterns are consistent with sequential-sampling accounts: near the indifference point, evidence accumulation has a lower drift rate and therefore takes longer (Krajbich et al., 2015; Clithero, 2018b). If deliberation time reflects genuine deliberation rather than merely a stable trait, it should increase as the distance between the amount offered and the participant's indifference point decreases.

Table 6: Deliberation time and decision difficulty

	(1) Person FE	(2) Person x Task FE	(3) FE + Question order
Distance from indifference (100 HUF)	0.422***	0.622***	-0.451***
	(0.0351)	(0.0345)	(0.0680)
Observations	10830	10830	10830
Adj. R-sq	0.246	0.293	0.351

Distance = |offered amount - indifference point| in 100 HUF.
Q = question-number FE (1-5).
SE clustered at the individual level.

Table 6 reports estimates of Equation 3 on a panel of 10,830 individual × task × question observations. Column (1) includes only individual fixed effects and yields a positive coefficient on distance ($\hat{\varphi} = 0.422$, $p < 0.01$). Taken at face value, this estimate would suggest that students deliberate *more* on easier (that is, earlier) questions. Column (2) tightens the specification by adding individual × task fixed effects, yielding an even larger and still positive coefficient ($\hat{\varphi} = 0.622$, $p < 0.01$). This pattern is confirmed by the median decision time by question depicted in Appendix Figure B4.

However, column (3) shows that both estimates are confounded. Adding question-order controls flips the sign of the distance coefficient ($\hat{\varphi} = -0.451$, $p < 0.01$) and improves model fit. The positive coefficients in columns (1) and (2) arise because students must first become familiar with the task. In other words, there are two opposing forces: a learning effect and the effect of decision difficulty. Once the learning effect is absorbed in column (3), the underlying difficulty effect emerges. Holding question number constant, offers closer to the indifference point require significantly more deliberation. Appendix Figure B5 visualizes the column (3) estimate in partial-regression form: conditional on question position, decisions closer to the indifference point take longer, as the sequential sampling prediction implies.

The sign reversal across columns is itself informative. It demonstrates the importance of controlling for question order in adaptive designs and confirms that the negative relationship in column (3) is not an artifact of the specification. It is the result of within-person, within-task variation in decision difficulty. Deliberation time is therefore not simply a measure of whether some participants are slower or faster decision-makers. It responds to how demanding the choice is.

6 Discussion and conclusions

This paper proposed a simple design-based decomposition of response time in economic experiments. By presenting task instructions and decision screens separately, total response time can be split into comprehension time and deliberation time. Using data from approximately 1,100 Hungarian secondary school students who completed six incentivized tasks (two time preference tasks, a risk elicitation task, a public goods game, and the send and return stages of a trust game), we show that these two components are empirically distinct and carry different information.

A factor analysis shows that comprehension and deliberation items load on separate latent factors, confirming that the decomposition captures two distinct constructs rather than noisy measures of a single "speed" trait. Yet, as our inventory of 35 studies shows (Appendix C), no economic studies have yet decomposed these two constructs.

Comprehension time is strongly associated with cognitive ability. Students who take longer to read the instructions score significantly lower on standardized mathematics and reading tests, even after controlling for family background, age, and gender. The effect is stronger for reading than for mathematics, which is intuitive: comprehension time is fundamentally a reading task. Even after controlling for math scores, comprehension time remains a significant predictor of reading ability. This result is consistent with the well-documented correlation between response time and cognitive ability in the psychometric literature (Doebler and Scheffler, 2016; Kumar et al., 2020; Jensen, 2006). While this correlation has been established in the context of simple perceptual tasks and IQ tests (Deary et al., 2001; Der and Deary, 2017), our contribution is to show that this relationship extends to the comprehension component of complex economic decision tasks. The finding also resonates with Junghaenel et al. (2022), who show that survey response latencies predict cognitive ability. Our approach goes further by isolating the comprehension channel through experimental design rather than statistical modeling.

The practical importance of controlling for cognitive ability when interpreting response time extends beyond the specific applications in this paper. Faster decisions have been read, depending on the literature, as evidence of higher processing speed (Cappelen et al., 2016; Junghaenel et al., 2022), lower cognitive effort (Alós-Ferrer and Buckenmaier, 2021), stronger preferences and clearer evaluations (Krajbich et al., 2015; Alós-Ferrer et al., 2021), more error-prone behavior (Recalde et al., 2018a), or heuristic processing (Rand et al., 2014; Rubinstein, 2016). Without an independent measure of cognitive ability, different forces of the decision-making process have been shown Cappelen et al. (2016); Alós-Ferrer and Buckenmaier (2021) or speculated Recalde et al. (2018b) to be confounded. The split-screen decomposition enables experimentalists to separate the comprehension and deliberation channels.

These results also speak to the interpretation of the non-decision time parameter in drift-diffusion models. In standard drift-diffusion model applications to economic choice (Clithero, 2018a; Smith et al., 2014), non-decision time is typically treated as a nuisance parameter capturing motor response and stimulus encoding. Clithero (2018b, p. 63) notes that this component is rarely analyzed. Our evidence suggests that it is not mere noise, it contains valuable information about cognitive ability.

Deliberation time captures something different. It shows no relationship with cognitive ability. Instead, it reflects the decision process. This null result is consistent with the argument of Alós-Ferrer and Buckenmaier (2021) that deliberation times capture cognitive effort in the decision itself, distinct from subjects' general processing speed. It also supports the conceptual distinction between decision time and non-decision time in the drift-diffusion model (Ratcliff, 1978; Clithero, 2018b).

Across the trust and public goods games, heuristic choices (focal allocations) are associated are fast while intermediate choices are associated with longer deliberation. This pattern is consistent with the Social Heuristics Hypothesis (Rand et al., 2014) and with Rubinstein (2016)'s distinction between instinctive and contemplative decision types. Importantly, the heuristic signature is visible in deliberation time but not in comprehension time. Comprehension time also varies across choice categories, but the pattern may reflect ability-driven sorting (Proto et al., 2019; Burks et al., 2009). This distinction matters: studies that test the Social Heuristics Hypothesis using total response time (Cappelen et al., 2016; Rand et al., 2012; Evans et al., 2015) may find compressed or distorted heuristic patterns because total *RT* confounds deliberation speed with comprehension speed.

In a time preference task with increasing decision difficulty, subjects have spent significantly more time on questions near their indifference point. This is exactly what sequential-sampling models predict: as the difference in subjective value between alternatives shrinks, the drift rate decreases and decisions take longer (Ratcliff, 1978; Krajbich et al., 2015; Clithero, 2018b). This finding extends the empirical evidence on the difficulty–response time relationship documented by Chabris et al. (2008), Konovalov and Krajbich (2019b).

The implication for experimental design is simple. Separating instruction screens from decision screens recovers two behavioral signals that are conflated on a single screen. This modification is costless and allows researchers to distinguish cognitive processing speed from deliberative decision-making. Our review of 35 studies using response time in economic experiments finds that none implements this separation. Hence, the existing literature cannot disentangle whether fast responders are quick thinkers, impulsive deciders, or both.

There are notable limitations. First, our sample consists of adolescents tested in a lab-in-the-field setting during school hours. There may be population effects, and results may not generalize to adult participants in laboratory experiments. Second, the original aim of the experiment was to study the preferences of Hungarian adolescents rather than to analyze response time. A purpose-built design would have implemented split screens in all eight tasks (including the dictator games), standardized task lengths so that deliberation-time factor loadings are uniform across tasks, and randomized task order to absorb fatigue effects.

Future research could extend this approach in several directions. Applying the decomposition to adult samples and different experimental paradigms would test the generality of our findings. Embedding the decomposition within a structural drift-diffusion framework could further separate non-decision time into comprehension and residual motor processing. Testing whether comprehension time predicts real-world outcomes beyond test scores (such as educational attainment or labor market performance) would establish its value as a cost-free cognitive ability proxy in field experiments where administering IQ tests is impractical.

Taken together, the message is simple: response time in economic experiments conflates two processes, and these carry different information about different aspects of the decision-maker. Separating them requires nothing more than an extra screen.

Statements and Declarations

Funding

This research was supported by the National Research, Development and Innovation Office of Hungary (project no. K-143415).

Competing interests

The authors have no competing interests to declare.

Consent to participate and ethics approval

Informed consent was obtained from all participants in the experiment. Because participants were secondary-school students, written informed consent was also obtained from their parents or legal guardians. The original experiment from which our data are drawn was approved by [ETHICS COMMITTEE NAME], decision [NUMBER / DATE].

Data and code availability

The experimental data used in this paper come from Horn et al. (2022). The replication code and processed data underlying all results in this paper are available at [REPOSITORY URL OR DOI]. The raw experimental data are available from the authors on reasonable request, subject to the conditions of the original data collection.

Use of AI tools

The authors declare that Anthropic's Opus 4.7. was used to assist with the compilation of source files, the compilation of the online appendix, proofreading and editing.

References

Agranov, M., Schotter, A., and Trevino, I. (2025). Complex for whom? An experimental approach to subjective complexity. Manuscript (February 27, 2025).

Alós-Ferrer, C. and Buckenmaier, J. (2021). Cognitive sophistication and deliberation times. *Experimental Economics*, 24:558–592.

Alós-Ferrer, C., Fehr, E., and Netzer, N. (2021). Time will tell: Recovering preferences when choices are noisy. *Journal of Political Economy*, 129(6):1828–1877.

Balázsi, I. and Ostorics, L. (2020). The hungarian educational assessment system. In *Monitoring student achievement in the 21st century: European policy perspectives and assessment strategies*, pages 157–169. Springer.

Berg, J., Dickhaut, J., and McCabe, K. (1995). Trust, reciprocity, and social history. *Games and economic behavior*, 10(1):122–142.

Bouwmeester, S., Verkoeijen, P. P. J. L., Aczél, B., Barbosa, F., Bégue, L., Brañas-Garza, P., Chmura, T. G. H., Cornelissen, G., Døssing, F. S., Espín, A. M., et al. (2017). Registered replication report: Rand, Greene, and Nowak (2012). *Perspectives on Psychological Science*, 12(3):527–542.

Burks, S. V., Carpenter, J. P., Goette, L., and Rustichini, A. (2009). Cognitive skills affect economic preferences, strategic behavior, and job attachment. *Proceedings of the National Academy of Sciences*, 106(19):7745–7750.

Campbell, D., Mørkbak, M. R., and Olsen, S. B. (2018). The link between response time and preference, variance and processing heterogeneity in stated choice experiments. *Journal of Environmental Economics and Management*, 88:18–34.

Caplin, A. and Martin, D. (2016). The dual-process drift diffusion model: Evidence from response times. *Economic Inquiry*, 54(2):1274–1282.

Cappelen, A. W., Nielsen, U. H., Tungodden, B., Tyran, J.-R., and Wengström, E. (2016). Fairness is intuitive. *Experimental Economics*, 19(4):727–740.

Chabris, C. F., Laibson, D., Morris, C. L., Schuldt, J. P., and Taubinsky, D. (2008). Measuring intertemporal preferences using response times. Working Paper 14353, National Bureau of Economic Research.

Chabris, C. F., Morris, C. L., Taubinsky, D., Laibson, D., and Schuldt, J. P. (2009). The allocation of time in decision-making. *Journal of the European Economic Association*, 7(2-3):628–637.

Chen, H., De Boeck, P., Grady, M., Yang, C.-L., and Waldschmidt, D. (2018). Curvilinear dependency of response accuracy on response time in cognitive tests. *Intelligence*, 69:16–23.

Clithero, J. A. (2018a). Improving out-of-sample predictions using response times and a model of the decision process. *Journal of Economic Behavior & Organization*, 148:344–375.

Clithero, J. A. (2018b). Response times in economics: Looking through the lens of sequential sampling models. *Journal of Economic Psychology*, 69:61–86.

Cornsweet, T. N. (1962). The staircase-method in psychophysics. *The American journal of psychology*, 75(3):485–491.

Crosetto, P. and Filippin, A. (2013). The "bomb" risk elicitation task. *Journal of Risk and Uncertainty*, 47(1):31–65.

Dai, J. and Busemeyer, J. R. (2014). A probabilistic, dynamic, and attribute-wise model of intertemporal choice. *Journal of Experimental Psychology: General*, 143(4):1489.

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

Der, G. and Deary, I. J. (2017). The relationship between intelligence and reaction time varies with age: Results from three representative narrow-age cohorts at 30, 50 and 69 years. *Intelligence*, 64:89–97.

Doebler, P. and Scheffler, B. (2016). The relationship of choice reaction time variability and intelligence: A meta-analysis. *Learning and Individual Differences*, 52:157–166.

Evans, A. M., Dillon, K. D., and Rand, D. G. (2015). Fast but not intuitive, slow but not reflective: Decision conflict drives reaction times in social dilemmas. *Journal of Experimental Psychology: General*, 144(5):951–966.

Falk, A., Becker, A., Dohmen, T., Enke, B., Huffman, D., and Sunde, U. (2018). Global evidence on economic preferences. *The Quarterly Journal of Economics*, 133(4):1645–1692.

Fischbacher, U. (2007). z-tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics*, 10(2):171–178.

Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic Perspectives*, 19(4):25–42.

Horn, D., Kiss, H. J., and Lénárd, T. (2022). Gender differences in preferences of adolescents: evidence from a large-scale classroom experiment. *Journal of Economic Behavior & Organization*, 194:478–522.

Jensen, A. R. (2006). *Clocking the mind: Mental chronometry and individual differences*. Elsevier.

Jiang, T. (2013). Cheating in mind games: The subtlety of rules matters. *Journal of Economic Behavior & Organization*, 93:328–336.

Jimenez, N., Rodriguez-Lara, I., Tyran, J.-R., and Wengström, E. (2018). Thinking fast, thinking badly. *Economics Letters*, 162:41–44.

Junghaenel, D. U., Schneider, S., Orriens, B., Jin, H., Lee, P.-J., Kapteyn, A., Meijer, E., Zelinski, E., Hernandez, R., and Stone, A. A. (2022). Inferring cognitive abilities from response times to web-administered survey items in a population-representative sample. *Journal of Intelligence*, 11(1):3.

Kahneman, D. (2011). *Thinking, Fast and Slow*. Farrar, Straus and Giroux, New York.

Konovalov, A. and Krajbich, I. (2019a). Revealed strength of preference: Inference from response times. *Judgment and Decision Making*, 14(4):381–394.

Konovalov, A. and Krajbich, I. (2019b). Revealed strength of preference: Inference from response times. *Judgment and Decision making*, 14(4):381–394.

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(1):7455.

Kumar, A. P., Omprakash, A., Kuppusamy, M., Maruthy, K. N., Sathiyasekaran, B. W. C., Vijayaraghavan, P. V., and Ramaswamy, P. (2020). How does cognitive function measured by the reaction time and critical flicker fusion frequency correlate with the academic performance of students? *BMC Medical Education*, 20:507.

Kyllonen, P. and Zu, J. (2016). Use of response time for measuring cognitive ability. *Journal of Intelligence*, 4(4):14.

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.

Matthey, A. and Regner, T. (2011). Do I really want to know? A cognitive dissonance-based explanation of other-regarding behavior. *Games*, 2(1):114–135.

Niederle, M. and Vesterlund, L. (2007). Do women shy away from competition? do men compete too much? *The quarterly journal of economics*, 122(3):1067–1101.

Ofek, E., Yildiz, M., and Haruvy, E. (2007). The impact of prior decisions on subsequent valuations in a costly contemplation model. *Management Science*, 53(8):1217–1233.

Pedersen, M. K., Díaz, C. M. C., Wang, Q. J., Alba-Marrugo, M. A., Amidi, A., Basaiawmoit, R. V., Bergenholtz, C., Christiansen, M. H., Gajdacz, M., Hertwig, R., Ishkhanyan, B., Klyver, K., Ladegaard, N., Mathiasen, K., Parsons, C., Rafner, J., Villadsen, A. R., Wallentin, M., Zana, B., and Sherson, J. F. (2023). Measuring cognitive abilities in the wild: Validating a population-scale game-based cognitive assessment. *Cognitive Science*, 47(6):e13308.

Pezeshki, M. and Embretson, S. (2025). Impact of parameter predictability and joint modeling of response accuracy and response time on ability estimates. *Applied Psychological Measurement*, 49(6):247–265.

Proto, E., Rustichini, A., and Sofianos, A. (2019). Intelligence, personality, and gains from cooperation in repeated interactions. *Journal of Political Economy*, 127(3):1351–1390.

Rand, D. G., Greene, J. D., and Nowak, M. A. (2012). Spontaneous giving and calculated greed. *Nature*, 489(7416):427–430.

Rand, D. G., Peysakhovich, A., Kraft-Todd, G. T., Newman, G. E., Wurzbacher, O., Nowak, M. A., and Greene, J. D. (2014). Social heuristics shape intuitive cooperation. *Nature communications*, 5(1):3677.

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review*, 85(2):59–108.

Ratcliff, R. (2013). Parameter variability and distributional assumptions in the diffusion model. *Psychological Review*, 120(1):281–292.

Ratcliff, R. and Tuerlinckx, F. (2002). Estimating parameters of the diffusion model: Approaches to dealing with contaminant reaction times and parameter variability. *Psychonomic Bulletin & Review*, 9(3):438–481.

Recalde, M. P., Riedl, A., and Vesterlund, L. (2018a). Error-prone inference from response time: The case of intuitive generosity in public-good games. *Journal of Public Economics*, 160:132–147.

Recalde, M. P., Riedl, A., and Vesterlund, L. (2018b). Error-prone inference from response time: The case of intuitive generosity in public-good games. *Journal of Public Economics*, 160:132–147.

Rodriguez, C. A., Turner, B. M., and McClure, S. M. (2014). Intertemporal choice as discounted value accumulation. *PloS one*, 9(2):e90138.

Rubinstein, A. (2006). Dilemmas of an economic theorist. *Econometrica*, 74(4):865–883.

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. *The Quarterly Journal of Economics*, 131(2):859–890.

Schmitz, F. and Wilhelm, O. (2016). Modeling mental speed: Decomposing response time distributions in elementary cognitive tasks and correlations with working memory capacity and fluid intelligence. *Journal of Intelligence*, 4(4):13.

Schotter, A. and Trevino, I. (2021). Is response time predictive of choice? an experimental study of threshold strategies. *Experimental Economics*, 24(1):87–117.

Schubert, A.-L., Hagemann, D., and Frischkorn, G. T. (2017). Is general intelligence little more than the speed of higher-order processing? *Journal of Experimental Psychology: General*, 146(10):1498–1512.

Schubert, A.-L., Hagemann, D., Voss, A., Schankin, A., and Bergmann, K. (2015). Decomposing the relationship between mental speed and mental abilities. *Intelligence*, 51:28–46.

Smith, A., Bernheim, B. D., Camerer, C. F., and Rangel, A. (2014). Neural activity reveals preferences without choices. *American Economic Journal: Microeconomics*, 6(2):1–36.

Spiliopoulos, L. and Ortmann, A. (2018). The BCD of response time analysis in experimental economics. *Experimental Economics*, 21(2):383–433.

Tinghög, G., Andersson, D., Bonn, C., Böttiger, H., Josephson, C., Lundgren, G., Västfjäll, D., Kirchler, M., and Johannesson, M. (2013). Intuition and cooperation reconsidered. *Nature*, 498(7452):E1–E2.

Wilcox, N. T. (1993). Lottery choice: Incentives, complexity and decision time. *The Economic Journal*, 103(421):1397–1417.