NBER WORKING PAPER SERIES

COMPLEX DISCLOSURE

Ginger Zhe Jin
Michael Luca
Daniel J. Martin

Working Paper 24675
http://www.nber.org/papers/w24675

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
June 2018

Part of the research was conducted when Jin took leave at the Federal Trade Commission. The views expressed are those of the authors and do not necessarily represent those of the U.S. Federal Trade Commission, any individual Commissioner, or the National Bureau of Economic Research. Martin would like to thank both the Paris School of Economics and the Camargo Foundation for their hospitality during the writing of this paper. Early stages of this project were supported by the French National Research Agency, through the program Investissements d'Avenir, ANR-10- LABX_93-01. We would like to thank Patrick Rooney, Byron Perpetua, Philip Marx for excellent assistance, and David Laibson, Matthew Rabin, and many others for constructive comments. All rights reserved. All errors are ours.

Complex Disclosure
Ginger Zhe Jin, Michael Luca, and Daniel J. Martin
NBER Working Paper No. 24675
June 2018
JEL No. D8,D91,K2,L15

ABSTRACT

Disclosure policies have the potential to help consumers and make markets more efficient. Yet, the effectiveness of disclosure policies can be undermined if firms strategically make unfavorable information unnecessarily complicated to understand. To explore the incentives for using complexity in disclosure, we implement a game of mandatory disclosure where senders are required to report their private information truthfully, but can choose how complex to make their reports. We find that senders use complex disclosure over half the time, and most of this obfuscation is profitable because receivers make systematic mistakes in assessing complex reports. Stated beliefs suggest that receivers correctly infer the strategic implications of complexity, but are overconfident about their ability to assess complex reports.

Ginger Zhe Jin
University of Maryland
Department of Economics
3115F Tydings Hall
College Park, MD 20742-7211
and NBER
jin@econ.umd.edu

Michael Luca
Harvard Business School
Soldiers Field Road
Boston, MA 02163
mluca@hbs.edu

Daniel J. Martin
Northwestern University
Kellogg School of Management
2211 Campus Drive
Evanston, IL 60208
d-martin@kellogg.northwestern.edu

1. Introduction

Firms are often required to disclose contract terms and other relevant information to consumers. For example, credit card companies are often required to disclose interest rates. Tech companies are often required to disclose privacy policies. And public firms are often required to disclose financial performance.

Yet in many cases, firms can choose exactly *how* to present the mandated information. One important component of this decision is whether to disclose information in manner that is simple to understand or complicated. For example, credit card companies can present payment schedules, penalties, and fees clearly or bury potentially important details in the fine print.[2] Privacy policies can be written in easy to understand language, or shrouded in pages of complex legalese. When public firms make financial disclosures, they can summarize them into several paragraphs or run as long as 257 pages.[3]

Sometimes, terms are complex by necessity – simply because there is a lot of information to provide. For example, there may be many contingencies and states of the world. However, the ability to choose how complex to make disclosures also raises the possibility that companies may strategically manipulate information to make it unnecessarily complicated in order to exploit consumer biases. In other words, there are situations where firms might use complicated terms when simple ones would suffice; legalistic privacy policies when plain English would be more informative; and pages of details about irrelevant firm activities when high-level summaries of firm performance would be more useful.

[2] The Truth in Lending Act of 1968 (TILA) requires lenders to disclose consumer credit terms and cost in a standardized way. The Real Estate Settlement Procedures Act of 1974 (RESPA) requires lenders and others involved in mortgage lending to provide borrowers with pertinent and timely disclosures regarding the nature and costs of a real estate settlement process. In 2015, the US Consumer Finance Protection Bureau consolidated the disclosure requirements under TILA and RESPA, resulting in the Loan Estimate Form and the Closing Statement Form, which standardize the content and format of disclosure in mortgage transactions.

[3] Since the SEC does not impose a limit on the length of a financial filing, an average 10-K has grown from roughly 30,000 words in 2000 to 42,000 words in 2013, with GE's 2014 10-K stretching to 103,484 words and 257 pages. Source: https://www.wsj.com/articles/the-109-894-word-annual-report-1433203762 accessed on September 26, 2017.

In practice, there are times when this seems to have been the case. For example, some accounting scandals occurred, not because the firm did not disclose their creative accounting in their SEC filings, but because the information was buried in thousands of pages and few readers could understand their real contents (e.g., Enron).[4] These examples are beyond anecdotes. Ben-Shahar and Schneider (2014) and Lowenstein, Sunstein, and Golman (2014) have criticized a long list of hard-to-understand disclosures and have argued that mandated disclosure has failed as public policy as a consequence. Financial experts even blame the complexity of financial products for the 2008 financial crisis, although the risks embedded in these products were supposedly disclosed to a ratings agency.[5]

The ability to strategically manipulate complexity to exploit consumers depends centrally on the inferences consumers make when they observe complex information. If consumers are skeptical of firms using complex disclosures, then firms that offer better terms or higher-quality products will want to present this information clearly and simply. For example, if the worst firms use complex disclosures, then firms that offer the second-best terms or medium quality products will want to use simple disclosures to separate themselves from those firms. As a result, we would expect all but the worst firms to offer the simplest possible terms (constrained by the true complexity of the transaction), similar to the "unraveling" results in voluntary disclosure (Viscusi 1978; Grossman and Hart 1980; Grossman 1981; Milgrom 1981).

In a world with unraveling, complex contracts would exist only out of necessity, with little scope for strategic complexity. However, systematic mistakes by consumers trying to extract the truth from complex reports can also give rise to strategic complexity, motivating companies to choose complexity over simplicity in their disclosures in order to mislead consumers. The welfare implications of complex disclosure depend heavily on the reasons why complexity is used. If complexity arises mainly out of necessity, it can be good for consumers – helping to provide as much relevant information as possible. However, strategic complexity is likely to be welfare reducing, misleading consumers and leading to worse decisions.

[4] http://www.investopedia.com/updates/enron-scandal-summary accessed on September 26, 2017.
[5] https://www.ft.com/content/24f73610-c91e-11dc-9807-000077b07658 accessed on September 26, 2017.

Because multiple factors can lead to complex disclosure, it is difficult to identify strategic uses of complexity in the field. To study the strategic use of complexity directly, we design a laboratory experiment that shuts down all the external factors including product complexity, pricing, and competition. In doing so, we focus cleanly on the strategic incentives to complexify information.

There are two roles in our experiment: an information sender (e.g., the firm) and an information receiver (e.g., the consumer). Subjects are randomly paired in each round, one randomly assigned as the sender and the other as the receiver.[6] In each round, the sender observes a new state (which is an integer drawn uniformly from 1 to 10), and chooses how complex to make their report of this number. When the report is simple, the number is presented as a single integer. When the report is complex, the state is presented as several computer-generated numbers (up to 20) that add up to the true number. We build in a clear conflict of interest: senders would like receivers to guess that the true state is as high as possible and the receiver would like to guess as accurately as possible.

In the main sessions, we debrief both players at the end of each round about the true state, the sender's choice of complexity, and the receiver's guess in that round. This way, subjects have many opportunities to learn the strategic forces in the game and the consequences of their actions.[7] We embed these learning opportunities in the main treatment because Jin, Luca, and Martin (2015) have shown that immediate and repeated feedback can steer voluntary disclosure towards the predictions of unraveling. Because the choice of simplicity over complexity can be viewed as a form of voluntary disclosure (though senders are mandated to disclose their true state), this design also tests how robust the convergence to unraveling is when some element of mandatory disclosure is subject to voluntary choice.

Our experiment generates two main findings, one related to sender behavior and the other related to receiver behavior. First, the choices of complexity we observe from senders depart substantially from the predictions of unraveling. Defining "low

[6] Roles were randomly assigned so that subjects could experience both roles, which allowed subjects to be well informed about the actions and payoffs available in both roles. This design feature increases the opportunities for learning and was used in two of the primary treatments of Jin, Luca, and Martin (2015).
[7] As a robustness check, we also run sessions without feedback. In addition, we run a robustness check where we limit the number of complexity levels available to senders. See Section 4 for more details.

complexity" as messages with <=5 numbers to sum, unraveling theory predicts that senders should use low complexity 90% to 100% of the time. In contrast, senders use low complexity less than 50% of the time, even if we look just at the second half of rounds. When not using low complexity, senders opt mostly to use high complexity (defined as messages with >=15 numbers to sum), and they do so a systematic way.[8] Over rounds, senders gravitate towards two extremes: using low complexity for high states (where secret numbers >=8) and using high complexity for low states (where secret numbers <=3). When the state is neither high nor low, senders use high complexity approximately 33% of the time, even in the second half of rounds. These results contrast with the convergence towards full disclosure found in simple voluntary disclosure with feedback (Jin, Luca, and Martin 2015), implying that the unraveling prediction may fail when we change the method of disclosure.

Though it is not consistent with the unraveling prediction, using high complexity to hide both low and middle states is optimal in our experiment because receivers guess higher than the actual secret number in both low and middle states, and this persists into the second half of rounds.[9] Because sender behavior is largely consistent with these strategic incentives, the average losses of senders are small and decrease over rounds.[10]

Our second main finding relates to the reasons behind receiver mistakes, which justify the use of high complexity at low and middle states. We estimate a structural model of receiver decision-making to investigate whether math errors alone can explain the guesses we observe.[11] This model fails to predict over-guessing at middle states or the extent of under-guessing we observe at high states, even when we add social preferences or risk aversion to the model.

We add behavioral biases to this model in order to better explain receiver guesses. In the theoretical literature on complex disclosure, the leading assumption that justifies the strategic use of complexity is that receivers are naive about sender strategies (for

[8] Senders use middle complexity at a similar rate across states, so the only variation across states is the fraction of rounds in which low and high complexity are used.
[9] We define the "optimal" action for senders as the one that has the highest expected payoff, and we measure losses in terms of expected payoff.
[10] The primary sources of losses are choosing high complexity at high states and choosing low complexity at low states. These mistakes, along with their possible sources, are examined in Section 4.3.
[11] Most model parameters are estimated out-of-sample using a math test and guesses with low complexity.

example, see Gabaix and Laibson 2006; Spiegler 2006; Carlin 2009; and Armstrong and Vickers 2012). In addition, naiveté about sender strategies has been found in experiments that study other forms of disclosure (Cai and Wang 2006; Jin, Luca, and Martin 2015).

To examine the potential impact of this behavioral bias in our experiment, we add naiveté about sender strategies to our model of receiver decision-making. The combined model is able to accurately predict receiver guesses. As a result, it reproduces the strategic incentives for complex disclosure that we observe in our experiment. However, naiveté is not the only behavioral bias that can explain our data. Adding overconfidence about math ability (instead of naiveté) also allows the model to accurately predict receiver guesses. Overconfidence about ability has been found in a number of domains and can take a wide variety of forms (Moore and Healy 2008; Grubb 2015).

To distinguish between these possible explanations for receiver over-guessing, we elicit beliefs from subjects both about the strategies of other senders and their own math ability. These beliefs suggest that receivers are not naive about sender strategies, but they are overconfident about their math ability. To our knowledge, overconfidence about the ability to internalize complex information has not been previously proposed as an explanation for complex disclosure.[12]

In addition, overconfidence could help to explain why feedback is not effective at reducing receiver mistakes in our experiment. There is mounting evidence that ego-utility is an important driver of overconfidence and can lead to asymmetric updating in beliefs about ability after receiving feedback about performance (Eil and Rao 2011; Mobius, Niederle, Niehaus, and Rosenblat 2011).

The rest of the paper is organized as follows. Section 2 reviews the related literatures and articulates our contribution. Section 3 presents the complexity game we study and our experimental design. Section 4 discusses our experimental results. Section 5 concludes with policy implications.

[12] Grubb (2015) presents evidence of how other forms of overconfidence interact with complex disclosures, such as overconfidence about the precision of estimates, overconfidence about self-control, and overconfidence about attention to fulfilling contract terms.

2. Literature Review

Our paper draws on and contributes to three literatures: the literature on voluntary and mandatory disclosure, the literature on obfuscation and behavioral biases, and the literature on communication experiments.

2.1 Voluntary and Mandatory Disclosure

In virtually every transaction imaginable, companies must decide what information to disclosure and whether to make disclosure simple or complicated. In practice, voluntary disclosure is observed in many industries, but is far from complete.[13] As summarized in Dranove and Jin (2010), this incompleteness can be explained by external factors such as disclosure cost and consumer knowledge before disclosure or by a seller's strategic incentives.[14] Depending on the perceived driver of incomplete disclosure, moving to mandatory disclosure can be beneficial or hurtful to the society, and could redistribute welfare between sellers and buyers. Also, mandatory disclosures can be subject to a range of behavioral biases, limiting their effectiveness (Lowenstein, Sunstein, and Golman 2014).

While theorists often contrast voluntary disclosure and mandatory disclosure as two distinct regimes, there is often a mixture of both voluntary and mandated elements in reality. For instance, policies that mandate disclosure on a limited number of dimensions may encourage firms to redirect resources to the mandated dimensions, but shirk on other dimensions (Lu 2012). Even on the mandated dimensions, firms may game the definition of the mandated statistics (Dranove et al. 2003; Jacob and Levitt 2003) or shroud it in a way that obfuscates important details (Brown, Hossain, and Morgan 2010). This effectively allows firms to voluntarily choose the content or format, even if the disclosure

[13] See Mathios (2000), Jin (2005), Bollinger et al (2011), Bederson et al. (2018), Anderson et al. (2015), Fung et al. (2007), and Luca and Smith (2015) for specific examples.

[14] For instance, see Jovanovic (1981) for the impact of disclosure cost on disclosure decisions, see Matthews and Postlewaite (1985) on the incentive to not knowing true quality, see Board (2009) on the incentive to use disclosure for differentiation, see Feltovich, Harbaugh, and To (2002) on relating disclosure to counter-signaling, see Grubb (2011) on the incentive to hide due to dynamic concerns, and see Marinovic and Varas (2016) on disclosure decisions in light of litigation risk.

itself is mandatory.

Because we focus on the voluntary choice of simplicity or complexity, we exclude (by design) other external factors that could complicate a firm's choice of simplicity in a mandatory disclosure setting. Senders do not lack information on the true state, have no legal concerns, face no disclosure cost, and have just a single attribute. In doing so, we simplify the strategic interaction between sender and receiver, which helps us to isolate how a subject's action is driven by their information set and their beliefs about their opponent. We believe these elements are fundamental for the general understanding of disclosure decisions.

2.2 Obfuscation and Behavioral Biases

A growing literature models why firms may choose obfuscation in light of consumer naiveté. For example, Ellison (2005) shows that add-on pricing can be rationalized if one adds a subpopulation of irrational consumers. Gabaix and Laibson (2006) develop a model in which firms can shroud dimensions of product information when some consumers are myopic or unaware. Heidhues, Koszegi, and Murooka (2016) further give out the conditions under which a shrouding equilibrium arises when naive consumers ignore add-on prices until at least one firm unshrouds (reveals) the additional price. Spiegler (2006) assumes consumers are only capable of evaluating one of many dimensions of the product, which motivates firms to obfuscate by making the product more attractive on some dimensions but less attractive on others. He also shows that competition could increase obfuscation in equilibrium. Similarly, Armstrong and Vickers (2012) model bank overdraft fees in a market where some consumers are sophisticated and some consumers are naive. They show that competition may end up subsidizing the sophisticated at the expense of the naive. In the finance literature, Carlin (2009) has modeled why firms might use complex pricing when some consumers are myopic. In the accounting literature, Hirshleifer and Teoh (2003) consider the impact of naiveté on financial disclosures, where receivers can be naive about non-disclosed information and inattentive to disclosed information.

Theoretically, receiver naiveté is not a necessary condition for senders to choose obfuscation. Firms may still engage in obfuscation even if all information receivers are rational. In a model where consumers must spend time to search for price, Ellison and Wolitzky (2012) show that firms have incentive to increase consumer's search cost through obfuscation. In doing so, obfuscation increases the search cost of consumers, raises equilibrium price, and benefits all firms even if some firms do not use obfuscation themselves. In a different setting, Perez-Richet and Prady (2012) consider obfuscation in front of a third party certifier (say bond rating agencies), whose job is to digest and certify the disclosed information. They find that even good types may add complexity in the disclosed information, a result that defies unraveling. This occurs because complexity reduces the certifier's ability to understand the report, which could motivate the certifier to lower its validation threshold.

The empirical literature has documented many examples of obfuscation. Brown, Hossain, and Morgan (2010) show that shipping and handling cost is often shrouded on e-commerce platforms. Sullivan (2017) shows that some hotels keep mandatory resort fees separate from room rate, and some online travel platforms conduct the price search by room rate only and do not disclose resort fees until consumers reach the hotel-specific page before payment. Obfuscation can also appear in a more sophisticated way. Ellison and Ellison (2009) document a loss-leader strategy by Internet retailers. In that strategy, the retailer sets a low price for a low-quality product on a price comparison site, and then persuades consumers to buy higher-quality products at a greater markup after consumers visit the retailer's website. Célérier and Vallée (2017) find that banks offer retail investment products in ways that are consistent with strategic obfuscation. For instance, more complex products are more expensive and are more harmful for consumers.

While these studies tend to focus on seller's choice of obfuscation, other empirical studies turn to document the behavior of information receivers. Chetty et al. (2009) study two price regimes that include or exclude tax in the list price (tax rate is well known). They find that people are much less responsive to tax in the second regime because taxes are more complicated to compute. Blake et al. (2017) study an online ticket platform that switched from transparent pricing to hiding transaction fees until payment. They find that consumers are more likely to buy more tickets and pay higher price if transaction fees are "back-end." Pope (2009) and Luca and Smith (2013) show that the salience of quality disclosure determines the extent to which customers respond. In a variety of settings, people are found inattentive to relevant details even after disclosure occurs (Armstrong and Chen 2010; DellaVigna and Pollet 2005; DellaVigna and Pollet 2009; Lacetera et al. 2012). In a similar spirit, Hanna, Mullainathan, and Schwartzstein (2014) show that consumers often only attend to certain once-overlooked information when information is presented in a summary form.

Our lab experiment complements this field work by jointly studying the decisions of senders and receivers in an environment where we can control the incentives and information of subjects and remove non-behavioral reasons for complex disclosure. By measuring sender and receiver mistakes at the same time, we can accurately determine departures from equilibrium and shed light on the extent to which our subjects behave optimally in response to opponent actions.

In addition, we are able to measure subjects' information processing difficulties and explore their potential behavioral bias in the form of over-confidence or failure to form correct beliefs about opponent strategies. In this sense, our work is related to the theory of Cursed Equilibrium (Eyster and Rabin 2005), Level-k reasoning (Crawford and Iriberri 2007), coarse thinking (Mullainathan, Schwartzstein, and Shleifer 2008), and rational inattention (Sims 2003).

2.3 Communication Experiments

Our experimental design is related to the cheap talk experiments of Cai and Wang (2006) and the voluntary disclosure experiments of Jin, Luca, and Martin (2015). For instance, we also frame states as "secret numbers" and use a similar payoff structure. The key differences in our experimental design are that the sender must truthfully reveal their type and can choose to make their reports complex. Hence, our experiment tests models of complex disclosure, rather than cheap talk or voluntary disclosure.

Few lab experiments have studied complexity explicitly. Kalayci and Potters (2011) implement an experiment where sellers have control over the complexity of product quality, but in their experiment buyers face time pressure and are given no information about the objectives and incentives of sellers, so it is difficult to know what buyers believe about why sellers present products in a complex way. Carlin, Kogan, and Lowery (2013) have used lab experiments to study how subjects trade assets after viewing information of different complexity levels. But their subjects are all information receivers and therefore they cannot draw a close link between sender and receiver behavior. In Martin (2015), buyers are given information about the seller's incentives, but the complexity of product quality is determined exogenously. In comparison, our experiment studies the endogenous choice of complexity.

Our measure of complexity is similar to what other experiments have used to generate cognitive costs for subjects. For instance, Caplin, Dean, and Martin (2011) find evidence of sacrificing behavior by having subjects choose among strings of numbers, where the value of an option is determined by the sum of the string. Caplin and Martin (2016) ask subjects to choose among sums of strings and find evidence consistent with a dual-process model of choice.

Though unraveling has been confirmed by multiple disclosure experiments, Jin, Luca, and Martin (2015) show that immediate and repeated feedback is crucial for subjects to converge to the predictions of unraveling. In comparison, our experiment focuses on mandatory disclosure rather than voluntary disclosure, but the choice of simplicity is voluntary and subject to the same unraveling logic. Our results suggest that in a setting different from the classical game of voluntary disclosure even immediate and repeated feedback (about the real meaning of complex report) is not enough to salvage unraveling.

Our work is also related to the experiments that study vagueness and ambiguity as a way to shroud information. For instance, Serra-Garcia, van Damme, and Potters (2011) allow non-disclosure to take the form of vague messages. They find that intermediate senders sometimes use vague messages, which receivers do not make correct inferences about. Agranov and Schotter (2012) study the use of both vague (natural language) and ambiguous (interval) messages, and find that an announcer in coordination games might want to use such messages. Relative to this literature, we consider complexity as another way to shroud information, under the constraint that the reported information must convey the truth state no matter whether it is simple or complex. Since receivers may process complex and vague/ambiguous reports differently, our work speaks directly to the real examples of complex disclosure when firms are subject to disclosure mandates.

3. Experimental Design

In this section, we first present a simple game of complex disclosure and then describe how we implement it in the lab.

3.1 A Simple Game of Complex Disclosure

The one-shot disclosure game we study involves two agents: an information sender and an information receiver. At the beginning of the game, nature determines the state b (which can be interpreted as the sender's type) by taking a draw from a distribution F that has full support over a finite subset of the real numbers B. The sender knows the realized state b, but ex ante, the receiver knows only the distribution of possible states. In our experiment, b is determined by a computer and is framed as the "secret" number, F is the uniform distribution, and B is the set of integers $\{1,2,3,4,5,6,7,8,9,10\}$.

The sender must report the realized state truthfully, but can choose the complexity of the report. As detailed below, we define complexity as the length of the report, which in our experiment consists of c computer-generated random integers that add up to the true state b. The sender chooses c between 1 and 20, where 1 corresponds to the shortest report and 20 is the longest.[15] We assume that if the sender chooses to make the report complex, the receiver observes a noisy signal x of the state b, where the noise $e=x-b$ reflects the error receivers make in reading complex reports. In our structural estimation, we assume further the noise e is drawn from a symmetric distribution G with full support over the integers $\{-9,-8,\dots,8,9\}$. Clearly, the distribution G should depend on the complexity of the report: the simplest report contains one number, and this number must be the true state b by definition, so it seems reasonable to assume that in this case G puts all probability on $e=0$ (no noise). As the report gets more complex, it also seems

[15] In practice, senders do not often choose intermediate levels of complexity.

reasonable to assume that the noise distribution *G* gets more dispersed.

After observing *x*, the noisy signal of the state, the receiver takes an action a from a subset of real numbers A. In our experiment, A will also be the set of integers *{1,2,3,4,5,6,7,8,9,10}* and is framed as the guess of the secret number *b*.

The true state and the receiver's action determine the payoffs for the two parties. The sender's utility is given by a function $U_S(a)$, which is concave, monotonically increasing in the receiver's action, and independent of the state. The receiver's utility is given by a function $U_R(a, b)$, which is concave in the receiver's action a and reaches its maximum when a is equal to *b*. In other words, the receiver benefits more from selecting an action that is closer to the true state, while the sender benefits the most when the receiver's action is as high as possible. These utility functions produce a strong conflict of interest when the state is low.

When there is uncertainty about actions and states, we assume that senders and receiver maximize expected utility. For receivers, the posterior probability of states (after observing the noisy signal *x*) is based on that signal, the prior distribution *F*, and the distribution of noise parameters *G* for a given complexity level. If *G* does not generate uncertainty about the true state, then regardless of the sender's action, the receiver will always choose an action equal to the realized state. However, with sufficient uncertainty about the true state, receivers may sometime choose a state that is not equal to the true state.

If there is sufficient uncertainty about the true state for complex reports, the techniques found in Milgrom (1981) can easily be adapted to show that in every sequential equilibrium of this disclosure game, the sender chooses to make a simple report if the state is not the minimum element in *B*, and if the report is complex, the receiver takes the action that is the minimum element in *B*. In other words, the sender always reports the state in the simplest form (unless it is the worst possible type), and the receiver always guesses the worst possible state if the sender chooses a complex report. When the realized state is the minimum element in *B*, the sender is indifferent between using simple or complex reports, so any mixture over these actions is consistent with equilibrium.

The force behind this equilibrium is that the best states using complex reports will want to separate from worse states using complex reports. Suppose states b_1, b_2, *and* b_3 all use the same level of complex report and $b_1<b_2<b_3$. Knowing this strategy from the sender, a rational receiver observing such a complex report will never guess the true state above b_3. When the signal is noisy enough, the receiver will sometimes even guess a worse state. In light of this, the sender with true state b_3 would prefer to reveal b_3 using a simple report. This leaves b_1 and b_2 to be the only two potential states behind the complex report. The same logic will motivate b_2 to switch to a simple report. Then observing the complex report automatically implies that the true state is b_1, which makes b_1 indifferent between simple and complex reports. Applied iteratively, this leads to unraveling in the use of complex disclosure at any level.

3.2 Implementing this Game Experimentally

In each round, subjects were paired together, and in each pairing, one subject was randomly assigned to be the sender and the other to be the receiver (with equal likelihood). To reduce framing effects, the sender was referred to as the "S Player", and the receiver was referred to as the "R Player".

In each round and for each pair, the computer drew a whole number from 1 to 10, called the "secret" number. Each of these numbers was equally likely to be drawn, and both senders and receivers were made aware of this probability distribution.

Each sender was shown the secret number for their pairing and then made their decision about report complexity while the receivers waited. In our main sessions, the sender chose a "report length", which was a whole number c between 1 and 20. The computer program randomly selected c integers between -10 and 10 until those numbers add up to the true state b. Both senders and receivers were told this is how the c numbers were generated.

After all senders made their decisions, the receivers' screens became active. If a sender decided to report their secret number with length c, the receiver they were paired with was shown this message: "The number I received is", followed by a table of c integers that range from -10 to 10 that add up to the secret number. The instructions were clear that the sender only chooses the report length c and the specific random numbers shown in the report are generated by the computer. In the Appendix, we present the full instructions and an example of a report with maximum length (20).

Below the area for the sender's message, receivers were asked to guess the secret number, and these guesses could be any integer between 1 and 10. The receiver had 60 seconds to view the sender's report and guess b. If nothing was guessed after that time, a random guess is entered for the receiver. In our main sessions, less than 4% of receivers hit this time limit.

Receiver payoffs, denominated in "Experimental Currency Units" (ECU), were $ECU_R = 110 - 20|(b-a)/2|^{1.4}$, where b is the secret number and a is the receiver's guess.[16] These payoffs decrease monotonically as the guess moves further from the secret number. The sender payoffs in each round were $ECU_s = 110 - 20|(10-a)/2|^{1.4}$. These payoffs are independent of the secret number and increase monotonically with receiver guesses because guesses cannot be higher than 10. These payoffs are similar to the quadratic specification found in Crawford and Sobel (1982) when there is a large bias towards higher actions. Because we use just a small number of states and actions, the payoffs could be shown in a table, so that subjects did not need to know or interpret these functional forms.

With these payoff functions, there was a clear misalignment of interests between senders and receivers. Receiver payoffs were higher when their guesses were closer to the secret number, and sender payoffs were higher when the receiver made higher guesses. Subjects were told in the instructions about these two features of sender and receiver payoffs.

At the end of each session, subjects were privately paid in cash a show up fee of \$5 plus all additional earnings they accumulate over the course of the session. ECU were converted to U.S. dollars at a rate of 150 to 1 (rounded up to the nearest dollar). While it is possible for subjects to end up with a negative balance of ECU, because subjects are paid for every round, this outcome is extremely unlikely and never came close to

[16] We allowed subjects accrue ECU in all rounds because payoffs could vary substantially between roles and realizations of the state, and we wanted performance to play a larger role than luck in final payments. Cai and Wang (2006) use similar payoff functions and also paid subjects every round. However, this approach introduces the possibility of wealth and portfolio effects. To ameliorate such effects, subjects were not told the cumulative payoffs they had earned so far in the experiment.

occurring in the sessions we ran. However, because subjects are paid for every round, there is the potential for intentional variation in play (a "portfolio" strategy), but we find little evidence of such behavior.

3.3 Experimental Sessions

Our sessions were conducted at the Computer Lab for Experimental Research (CLER) facility at the Harvard Business School (HBS). In this laboratory, subjects are separated with dividers, and each subject was provided with a personal computer terminal. Subjects do not have to be Harvard University students, but we restricted subject to be no older than 25 years old. The software used to run the experiments was the z-Tree software package (Fischbacher 2007).

Each session consisted of 30 rounds of the disclosure game. In each round, subjects were randomly matched into pairs. To reduce reputational effects, subjects were matched anonymously and were told that it was very unlikely they would be paired with the same subject in consecutive rounds. For a session size of 14, the actual likelihood of being paired with the same subject in consecutive rounds is 7.7%. The purpose of switching roles is to insure that both sides have a good sense for the incentives and actions available to the other side. In a related experiment, Kalayci and Potters (2011) implement a laboratory experiment where sellers have control over the complexity of product quality, but in their experiment buyers are given no information about the objectives and incentives of sellers, so it is very difficult to know what buyers believe about why sellers who make quality complex.

3.4 Feedback, Beliefs, and Math Test

Our main sessions provide round-by-round feedback. Subjects were told four pieces of information after each round: 1) the actual secret number; 2) the report length chosen by the sender; 3) the receiver's guess of the secret number, and; 4) their own payoff. After all subjects pressed the "OK" button on the screen containing this feedback, the next round began. To reduce social considerations, subjects in the feedback treatment were not told the payoffs for the other player in their pairing, though it could be deduced using the payoff table. In addition, between rounds subjects only received feedback about their pairing, not all pairings in the session.

Once all rounds are completed, subjects were asked questions about their beliefs of how other subjects played in their session. First, subjects were asked to guess the average report length that senders chose for each secret number. Second, subjects were asked to guess the average secret number was when the sender chose complexity levels between 1 and 5, between 6 and 10, between 11 and 15, and between 16 and 20. The purpose of these questions was to assess whether subject beliefs about sender strategies influenced their decisions as receivers. These guesses were not incentivized, but in a recent paper, Trautmann and Kuilen (2015) show that such "introspective" elicitation can yield accurate beliefs.

In some sessions, subjects were asked to complete a four-question math test after answering the two belief questions. For each question in this test, subjects were asked to add up 20 numbers, and were paid $4 if a randomly selected question was correctly answered. Subjects were told that the numbers would sum up to an integer between 1 and 10, that all integers were equally likely, and that the 20 numbers would be generated in the same fashion as in the disclosure game. After completing the math test, subjects answered two additional belief questions. First, they were asked to guess the number of questions on the math test (from 0 to 4) that they thought they answered correctly. Second, they were asked to guess the average number of questions they thought others answered correctly.[17] These belief questions were also not incentivized.

3.5 Robustness Sessions and Demographics

For robustness, we adopted two alternative treatments. The first alternative replaces round-by-round feedback with "no feedback," where subjects were given no information after completing each round. After all receivers had made their decisions,

[17] The exact wording of the questions was "For yourself, what do you think was the number of rounds (between 0 and 4) answered correctly?" and "For all participants, what do you think was the average number of rounds (between 0 and 4) answered correctly?"

subjects proceeded to a screen that required them to click "OK" to start the next round. The no-feedback treatment is designed to contrast with the feedback treatment, so that we can determine whether round-by-round feedback is crucial in driving convergence towards unraveling as in Jin, Luca, and Martin (2015). The second alternative treatment also limits sender choice of report length to the two extremes (c is only 1 or 20) rather than the full range from 1 to 20. The reason for this alternative treatment is to determine whether play is substantially different if the "strategic complexity" of the game is reduced for both senders and receivers.

In short, our experiment includes three treatments: feedback (our main sessions), no feedback, and two report lengths. Subjects completed just one of the three treatments. In all three treatments, subjects were asked at the end of the experiment to complete a questionnaire that includes questions about demographic details. Specifically, subjects are asked for their gender, if they are a native English speaker, their year in school, and if they have a friend participating in that session.

4. Experimental results

In this section, we first report the results from our main sessions and then compare them to the results from our robustness sessions. We then explore the possible reasons behind sender and receiver mistakes. For receivers, we estimate a structural model to predict choices both with and without behavioral biases.

4.1 Results from the Main Sessions

Table 1 summarizes the characteristics for the subjects in our 29 main sessions. In total, we have 294 subjects, all of whom experience both roles (sender and receiver) and receive round-by-round feedback for 30 rounds. Roughly 41% of the subjects are male, 72% are undergraduate students, 85% are native English speakers, and 14% report that they have a friend in the same session.[18] These demographic distributions are similar to the ones reported by Jin, Luca, and Martin (2015), who also conducted experiments in the CLER lab.

4.1.1 Summary of Behavior and Mistakes

Table 2A summarizes sender choice of complexity by secret number. In contrast to the unraveling prediction, the average choice of complexity is 9.728 and increases almost monotonically as the secret number gets smaller.[19] For the two smallest secret numbers (1 and 2), a majority of senders choose the maximum complexity (report length $c=20$) and over 72% choose high complexity ($c>=15$). For the two highest secret numbers (9 and 10), a majority of senders choose the simplest report ($c=1$) and over 72% choose low complexity (c<=5). For secret numbers in the middle, the median choice of complexity goes down from 13 for secret numbers of 4 to 4 for secret numbers of 7.

Figure 1A depicts the distribution of complexity choices for each secret number, where the size of the bubble represents the number of senders choosing a specific complexity level conditional on a specific secret number. Most senders concentrate on high complexity when the secret number is below 5, and switch to low complexity when the secret number is above 5. If the secret number is exactly 5, sender choices are dispersed across all levels of complexity.

Turning to receivers, Table 2B shows that the median receiver guess is correct for every secret number, but the standard deviation of receiver guesses is non-trivial (ranging from 1.167 to 2.326). As a consequence, guesses are significantly different from secret numbers for every secret number except for 6 and 7 (using a two-sided t-test and a significance level of 5%). On average, the bias in receiver mistakes reveals much greater over-guessing for low secret numbers (1.183 for secret numbers of 1 and 0.936 for 2)

[18] One subject did not report any demographics, and three subjects skipped the question about whether they were native English speakers. Despite these missing values, we include all subjects in analysis because our regressions will include subject fixed effects and therefore absorb all demographic variables.
[19] In a regression of complexity choice onto secret number with individual fixed effects and robust standard errors, the coefficient is negative (-1.496) and statistically significant ($p<0.001$).

than under-guessing for high secret numbers (-0.367 for 9 and -0.403 for 10), which suggests that receiver mistakes are not driven entirely by mechanical boundary effects.[20]

To further explore receiver behavior, we define the size of receiver mistakes as the absolute distance between the receiver guess and the secret number. As shown in Table 2B, the average receiver mistake size is the highest for the lowest secret number ($c=1$) and decreases almost monotonically with secret number.[21] This is consistent with the fact that senders present simpler reports for higher secret numbers, which reduces the potential for math errors, shortens the response time for receivers, and lowers the probability of receivers not making a decision within the 60 second time limit. For the less than 4% of receivers that are over the time limit, the computer generates a random guess, which can lead to large mistakes. Excluding these observations, receiver mistake sizes remain large for the smallest secret numbers (0.946 for secret numbers of 1 and 0.777 for 2) as compared to the mistake sizes for large secret numbers (between 0.370 and 0.379 for secret numbers between 6 and 10). In fact, mistake sizes are significantly different between secret numbers of 1 and 10 using a two-sided Wilcoxon rank-sum test (p-value<0.001).

Because receivers observe the complexity of sender reports, Table 3 tabulates how receiver guesses and mistakes vary by the complexity level of sender reports, as well as the secret numbers behind these reports. On average, we observe a small amount of under-guessing for complexity up to a length of 4. For complexity between 5 and 12, the number of observations is smaller, and the average guess fluctuates between over-guess and under-guess. Once complexity is over 12, we observe consistent over-guessing that peaks at the highest level of complexity (0.655 for length 20). Interestingly, the size of receiver mistakes is less monotonic, but is clearly much higher for high complexity than for low and medium complexity.[22] These results are robust to excluding rounds where receivers did not make their decision within the time limit. Without those rounds, the

[20] Additional evidence in support of this conclusion is provided in Section 4.4.
[21] In a regression of mistake size onto secret number with individual fixed effects and robust standard errors, the coefficient is negative (-0.085) and statistically significant (p<0.001).
[22] In a regression of mistake size onto complexity with individual fixed effects and robust standard errors, the coefficient is positive (0.054) and statistically significant (p<0.001).

magnitude of receiver mistakes is slightly lower for the two highest complexity levels (0.783 versus 0.748 for length 19, and 1.284 versus 1.008 for length 20).

Absent behavioral factors, one would imagine that receiver mistakes should be zero for the simplest reports because such report reveals the secret number exactly. In contrast, the data shows an average mistake of 0.243 for length 1 and 0.257 for length 2, which are significantly different from 0 using a two-sided t-test. There could be multiple explanations for this phenomenon: some receivers may not understand the game, while others may understand the game but want to reward the senders that reveal low secret numbers with a simple report. To the extent that subjects learn about the game over time, errors due to the first explanation should decline over time, but the social preferences in the second explanation are likely to persist. These possibilities are discussed and analyzed further in Section 4.4.

To show the joint impact of secret numbers and complexity, Table 4 cross-tabulates secret numbers by low (<=5), medium (6-14), and high (>=15) levels of complexity. Consistent with social preferences, when the secret number is presented simply, receivers tend to have larger mistakes (0.6 on average) for the lowest secret number (1). This is sensible if receivers possess some social preferences, as a simple report of low states are helpful for receivers and therefore if the receiver wants to reciprocate, she could be willing to sacrifice her own utility to reward this "honest" behavior. In contrast, under-guessing a complex report of high states would hurt both the receiver and the sender. These results are also consistent with some strategic confusion, as there is a similar average mistake size for other secret numbers.

4.1.2 Optimal Responses

So far, we have documented that sender choice of complexity deviates from the unraveling prediction and receiver guesses deviate from the true states behind sender reports. Could these "mistakes" be a rational reaction to opponents who do not follow the equilibrium strategy as predicted by unraveling theory?

To address this possibility, we measure how far a subject is from taking the payoff-maximizing action in each decision problem, which provides a rough sense for the size and consequences of the "mistakes" they are making. To do this, we construct the average opponent strategy from our data, determine the expected payoffs from taking each possible action, and then calculate how far the expected payoff for the taken action is from the highest expected payoff.[23] For senders, the possible actions are grouped as low (1-5), medium (6-14), and high (15-20) complexity.[24] For receivers, the possible actions are limited to the guesses available to them, which are integers between 1 and 10.

All of our calculations take an ex-ante perspective, so when determining the highest expected payoff for receivers, we assume that all states are equally likely to happen and determine the average sender behavior separately for each state. In addition, we pool all rounds when determining average sender and receiver behavior, which is equivalent to assuming that a subject is equally likely to face an opponent from any round.[25]

Table 5 reports the monetary losses that result from actions taken in our main sessions. On average, senders are 15.3% away from the highest expected payoff if they take the empirical distribution of receiver guesses (conditional on every possible complexity group) as given. This deviation differs substantially across secret numbers: for the highest secret number (10), sender choice (mostly low complexity) is close to optimal (3.8% loss); but for the lowest secret number, sender choice (mostly high complexity) is still 51.6% away from the highest payoff. This is driven mostly by the failure to always use high complexity when facing a secret number of 1.

We also calculate expected payoffs relative to the payoff that senders would get in the unraveling equilibrium. Because the unraveling equilibrium predicts different receiver behavior that we observe, sender payoffs in equilibrium could be higher or lower than the sender payoff observed in our data. It turns out that sender choice of complexity results in a 71.4% expected gain for secret numbers of 2 and a 3.9% expected loss for secret

[23] Because the minimum possible payoff can be negative, we normalize payoffs by subtracting the minimum possible payoff (for the realized state) to the payoffs from taking any action in that state..
[24] We grouped these actions because some complexity levels are rarely chosen by senders for some secret numbers, thus we could have a non-reliable density in the empirical distribution of sender choice of complexity conditional on these secret numbers.
[25] These assumptions may not hold in a dynamic environment that features learning. We will present evidence of learning in Section 4.1.4 and control for these dynamic effects in the regression analyses presented in Section 4.1.5.

numbers of 10, relative to the unraveling equilibrium. We cannot do the same exercise for a secret number 1 because the normalized equilibrium payoff is 0.

Table 5 also reports the deviation from the highest expected payoff that receivers could achieve if they take the empirical distribution of sender choices as given. This deviation is 13.8% for low complexity, 16% for medium complexity, and 16.7% for high complexity. The observed receiver payoff is 30 to 33% worse than the payoff that receivers would get in an unraveling equilibrium, because receivers would know every state perfectly in this equilibrium. Note that the departure from highest expected payoffs is not readily comparable between senders and receivers, because their payoffs differ in both scale and range.

In short, neither sender choice nor receiver guesses appear to be an entirely rational reaction to the empirical distribution of opponent behavior pooled across rounds. There are non-trivial sender mistakes and receiver mistakes, even if we measure them in the payoff space. We will test the robustness of these results to dynamic effects in Section 4.1.5 and explore reasons behind these mistakes in Sections 4.3 and 4.4.

4.1.3 Beliefs and Math Ability

As shown before, receivers tend to make their biggest mistakes when the sender report is highly complex and when the secret number is low. One potential explanation is that receivers make more math mistakes when facing more complex reports. In addition, receivers may not fully understand the degree to which senders choose higher complexity in lower states, thus not appreciate the extent to which complexity is "bad news". We measure both directly in our experiment using additional tasks.

After all 30 rounds of the game were completed, we asked subjects to predict the average sender choice of complexity for each secret number. The distribution of average stated beliefs and the distribution of average sender choices in our data are strikingly similar (see Figure 2A), suggesting that subjects are not naive in their beliefs about sender strategies.

Assuming receivers use these stated beliefs as their prior beliefs at the beginning of each round and only use the observed complexity level (not the content of each report) to determine the value of the secret number, we can infer what they should have guessed via Bayes' Rule. As shown in first panel of Table 6, this value (referred to as the "inferred guess") is on average 2.546 for high complexity (length>=15), which is *lower* than the average actual guess (4.222) in the game. Since the average secret number is 3.712 for high complexity, we conclude that the stated sender beliefs cannot explain why receivers systematically *over-guess* the true state when they face a complex report.

We also asked subjects to report what they would guess for the secret number on average if the reported complexity is 1-5, 6-10, 11-15 or 16-20.[26] The average answers are presented, along with actual averages, in Figure 2B. Once again, stated beliefs are close to the empirical frequencies.

Using these stated beliefs, we find that subjects on average believe the secret number is 2.510 if the report has a complexity between 16 and 20. This average, referred as "complex guess" in the bottom panel of Table 6, is lower than both the average secret number (3.626) and the average guess in the game (4.191). Again, this provides evidence that receiver beliefs cannot explain why they *over-guess* when faced with a complex report.

To what extent are guesses impact by math errors? After beliefs are elicited, 160 subjects were asked to complete a math tests that consisted of four questions. Each question required them to sum 20 numbers in a table similar to the most complex table in our game. These questions are highly incentivized for correct answers, and these answers do not impact the payoffs of other subjects, which should minimize social considerations. As shown in Figure 3, only 54 subjects get all four questions correct (33.75%), 48 get one wrong (30%), 27 get two wrong (16.88%), and the remaining get either three wrong (10.62%) or all wrong (8.75%).

Interestingly, when we asked each subject how many math test questions they think they answered correctly, 41.88% believe they got all correct and 72.5% believe they got three or four correct. Both of these rates are higher than the actual fraction of subjects

[26] This belief question uses a different grouping of complexity levels due to a lack of perfect foresight about the clustering of sender actions. Throughout the rest of the paper, we group complexity into low (1-5), medium (6-14) and high (15-20) levels because the empirical distribution of sender choice has much higher density at the two ends (1 and 20) and there is clear bunching at 1, 5, 10, 15 and 20. This difference does not affect our analysis, as we report the summary statistics of stated beliefs separately from other variables.

who got this many correct (33.75% and 63.75% respectively), which is shown in Figure 3. When asked to predict the average number of questions that other subjects answered correctly, the mode prediction was 3 and the average prediction was 2.694, which is very close to actual average (also 2.694).

In short, our subjects on average appear to be sophisticated enough to infer the extent to which complexity is bad news and many of them appear to be overconfident about their own math ability. In Section 4.4, we will further explore the role of naiveté and overconfidence in receiver mistakes.

4.1.4 Evidence of Learning

To provide detail on sender complexity use over rounds, the first panel of Table 7 also presents how sender payoffs depart from the highest expected payoff over rounds (taking the empirical distribution of receiver behavior as given and fixed over rounds). Overall, we see a gradual improvement from the beginning 10 rounds (15.9% departure) to the last 10 rounds of the game (14.2%). Breaking this down by secret number, the biggest improvement comes from the lowest secret number (1), where the departure from the highest payoff drops from 55.1% in the first 10 rounds to 51.5% and 48.4% in the second and third blocks of 10 rounds. Strikingly, this improvement is accompanied by senders *increasing* their choice of complexity for this secret number. In comparison, at the highest secret number (10) senders get closer to the highest payoff (from 5.3% departure in the first 10 rounds to 3.1% in the last 10 rounds) while *decreasing* their use of complexity (from 5.829 to 2.512).

For other secret numbers above 5, we also see senders decrease their use of complexity over the experiment. However, for secret numbers at or below 5, senders continue to use substantial amounts of complex disclosure throughout the experiment, as reflected in an average complexity choice above 10 in the last block of rounds. This is in contrast to the convergence towards full disclosure found by Jin, Luca, and Martin (2015) in a game of simple voluntary disclosure with round-by-round feedback.

There is also evidence of learning on the receiver side. Figure 4 plots the average size of receiver mistakes for low (1-5), medium (6-14) and high (15-20) levels of complexity. While receivers may not get better in math within the short time of the experiment (and may even become fatigued), they could become more aware of the strategic meaning of report complexity and realize that high complexity implies low states. This explanation is also consistent with the dynamics of senders, as they chose high complexity in low states and low complexity in high states more often in the second half of rounds. This pattern can be seen by comparing Figures 1A and 1B.

More details about receiver mistakes are given in the last two panels of Table 7. Throughout the game, the average receiver mistake drops for all three groups of complexity, but the biggest drop occurs for high complexity. Departure from the highest payoff improves as well, while the magnitude of improvement tends to be much larger for medium and high complexity (from ~18% to ~13%) than for simple complexity (from 15.3% to 13.7%). This is consistent with the conjecture that many mistakes upon simple reports may reflect social preferences but mistakes in medium and high complexity could be driven by other factors, such as math errors, overconfidence, and naiveté.

4.1.5 Regression Results

Table 8A presents the results of our regressions based on sender behavior, and Table 8B presents those based on receiver behavior. The motivation for these regressions is to replicate our results while controlling for round-by-round changes in sender and receiver behavior, which we reported in the previous section.

For senders, the dependent variables are sender choice of complexity and the payoff departure from the highest expected payoff. In the first and third columns of Table 8A, we include subject demographics, math test performance, the degree to which they overestimate their performance on the math test, and how much they believe their performance was better than the average performance of others. Taking a secret number of 1 as the default, Table 8A shows that senders choose significantly less complexity and depart less from the highest payoff when their secret number increases. This is consistent with our results without subject or round controls. Math ability and overconfidence explain little in sender choice, but when a sender believes other subjects are better at math, they choose a higher complexity.

To capture sender learning, we include the round number (1-30) and the interaction with whether the secret number is in the medium (4-6) or high range (7-10). These coefficients suggest that senders learn to *increase* complexity for low states (1-3), but *decrease* complexity for medium and high states. We also include a dummy for the first five rounds, in case the initial learning about the game creates a level effect in choice of complexity. There is little evidence for a difference when controlling for other factors.

Columns (2) and (4) include sender fixed effects, which absorb individual demographics, math test performance, and beliefs about math performance. Results for most coefficients are similar to what we have without individual fixed effects, suggesting that sender choice and learning are not driven by unobserved individual characteristics.

Turning to receivers, Table 8B attempts to understand the absolute size of receiver mistakes and receiver's payoff losses when controlling for time trends. Because we want to study the mistakes that receivers actively made, we focus our analysis on the 96% of receiver guesses that are made before the time limit.

Since receivers observe the sender's choice of complexity, we include a separate dummy for each complexity level. In addition to controlling for the same subject variables as in the sender regression, we also include the receiver's inferred guess for a given complexity level based on their stated beliefs. If receivers state their true belief and are risk neutral, this variable captures what receivers would guess to be the secret number if they observe the sender's choice of complexity but do not have the complex report (table of c numbers) in front of them.

Compared with the default complexity (1), Table 8B shows that receiver mistakes drop significantly for some low complexity levels (3-5) but increase significantly for almost all levels of high complexity. This pattern is similar with and without subject fixed effects. Results on payoff losses are less consistent, but once we control for subject fixed effects, payoff losses are significantly different between the default and most complexity levels 13 and above.

Receivers do appear to lower their guess for high complexity over time (after we control for subject fixed effects). As a result, they depart less from the highest expected payoff with high complexity. The magnitude of learning is non-trivial: throughout the 30 rounds, receivers will lower their over-guess of high complexity by 0.453, about half of the average over-guess in the pooled data (0.913).

Table 8B also suggests that math ability matters. But since the coefficient is negative, the regression suggests that lower ability subjects tend to make smaller mistakes on average (after controlling for other factors). Conversely, time spent is positively correlated with receiver mistakes. These counterintuitive results can have two explanations: 1) receiver mistakes are not driven by large math errors, or; 2) math errors affect receiver mistakes in a more nuanced way than is specified in this reduced-form regression.[27] In addition, overconfidence does not appear to have a significant relationship with receiver mistakes or payoff losses. The impact of both factors will be examined further in a structural estimation in Section 4.4.

4.2 Results from the Robustness Treatments

Table 9A compares sender choice of complexity in the main sessions and robustness sessions. In particular, we have two types of robustness sessions: Robust 1 refers to sessions that maintain the random assignment of roles and the same set of complexity options (1, 2, …20) but do not provide round-by-round feedback to the subjects. Even so, subjects can still learn about the game by playing both roles, by observing the random realizations of secret numbers as a sender, by observing simple reports as receivers, and by reading complex reports as receivers. Robust 2 refers to sessions that also restrict sender choice of complexity to the two extremes (1 or 20).

All three types of sessions demonstrate similar monotonicity between secret number and choice of complexity: most senders choose high complexity for low states and low complexity for high states. This tendency is strongest in Robust 2, which makes sense because Robust 2 restricts sender choice to the extremes. A comparison between main sessions and Robust 1 suggests that feedback drives senders even more to the two extremes.

[27] Our regression specifications do not include interactions between math performance (our measure of math ability) and different levels of complexity.

Table 9B and Figure 5 provide the comparison across treatments for receivers. The size of their mistakes is similar across the main and robustness sessions, and if anything, receiver mistakes (for high complexity) are slightly higher in Robust 2. Once again, this is not surprising given that Robust 2 pushes all complex reports to the extreme. However, the size of receiver mistakes at high complexity is not significantly different between the main sessions and either of the robustness sessions.

In short, we conclude that the patterns we observe in the main sessions are robust to changes in feedback design and the number of complexity options. For the remainder of the paper, we focus on main sessions only.

4.3 Reasons Behind Sender Mistakes

From a policy perspective, sender mistakes often capture less interest than receiver mistakes, partly because in the real world senders tend to be firms, which have more resources to overcome their mistakes. However, because subjects play both roles in our experiment, we can hope to learn something about the sources of receiver mistakes by looking at the sources of sender mistakes.

The largest sender losses come from two types of mistakes: using high complexity when the state is high and using low complexity when the state is low. In our main sessions, the former decreases from 11.4% of high-state decisions in the first half of rounds to 7.8% in the second, and the latter occurs in 16.0% of low-state decisions in both the first and second half of rounds.

Both types of sender mistakes could be driven by incorrect beliefs about receiver actions, random errors, or confusion about game form.[28] These factors could be ameliorated with experience and feedback, so we might expect their impact to lessen over rounds. However, only the incidence rate of the first mistake – choosing a high complexity level in a high state – decreases over rounds. Evidence that these mistakes might be driven by errors or confusion can be found by comparing the choices a subject makes when she is a sender and a receiver: there is a positive correlation (0.1344)

[28] Martin and Munoz Rodriguez (2018) find evidence of inattention to game form in experiments that use the BDM mechanism.

between the likelihood of a subject choosing a high complexity level in high states as a sender and incorrectly guessing by more than one integer with a simple report when she is a receiver.

Both types of sender mistakes could also be driven by social preferences. Spite could drive senders to use high complexity when it is not justified in their own payoff, and social norms could drive senders to use low complexity when it is not justified in payoff. We find some evidence that choosing low complexity in a low state is driven by social preferences by once again comparing the choices a subject makes when they are a sender versus a receiver. If a subject thinks that the socially correct action is to disclose simply for even low states, then he or she might act in this way and reward senders who do the same. In fact, there is a positive correlation (0.2666) between the likelihood of a subject choosing a low complexity level in low states as a sender and over-guessing the state by one integer with a simple report as a receiver.

Because subjects play in both roles, we will include these two possible reasons for sender mistakes – confusion and social preferences – into our baseline model of receiver guesses. However, we will find that neither appears to be a major driver of receiver mistakes.

4.4 Reasons Behind Receiver Mistakes

In this section, we will study the reasons for the mistakes that receivers make when the secret number is presented in a complex way. Along the way, we will also explore the reasons behind the mistakes made with simple reports, but our primary focus is on complex reports because the vast majority of receiver mistakes occur when the secret number is disclosed with high complexity, and it is these mistakes that justify the complexity that is observed in our experiment. As a consequence, in the subsequent analyses we only use receiver guesses from rounds where senders chose high complexity and where receivers made a guess before the time limit.

We start by modeling receiver mistakes using the Quantal Response Equilibrium (QRE) approach of McKelvey and Palfrey (1995), which assumes that receivers have Logit demand for each action based on the expected payoffs to taking each action given the empirical distribution of opponent actions. This approach has a free parameter often interpreted as the sensitivity of errors to expected payoffs, which we estimate using maximum likelihood. As can be seen in Table 10, the predictions based on this estimated parameter produce an average likelihood of -1.733 and do a reasonably good job at predicting the rates of over-guessing and under-guessing in the experiment. In particular, it is able to capture over-guessing for middle secret numbers.

However, while QRE is successful at explaining receiver mistakes, it does not indicate why receivers are making these particular mistakes. Instead, it tells us that receiver mistakes have a systematic structure that is well approximated by the QRE approach. To understand why receiver mistakes have this particular structure, we explicitly model the decision problem faced by receivers.

Because receivers face an involved decision problem, we investigate the sources of receiver mistakes using a structural model that is based on the theoretical framework presented in Section 3. To simplify this analysis, we hold the distribution of sender behavior fixed so that we can treat the receiver's choice as an individual decision problem.

In this structural model, we assume that a receiver facing a complex message (N>=15) has prior beliefs about the likelihood of each secret number b given by F, so that

$$b \sim F, where\ F \in \Delta(B).$$

The receiver then observes a noisy signal of the secret number, which can be interpreted as either an error in summing the numbers or partial attention to the grid of numbers. We assume that this noise signal is generated by adding an error term e drawn from the distribution G to the secret number, so that

$$x = b + e, where\ e \sim G.$$

Based on the signal x and their prior beliefs F, the receiver forms posterior beliefs γ and takes an action a (makes the guess) that maximizes their expected utility subject to some probability of making strategic errors. This decision rule is given by the following optimization problem:

$$max_{a \in A} \sum_{b \in B} \gamma(b|x) U_R(a, b)$$

$$where\ \gamma(b|x) = \frac{F(b)\Pr G(x-b)}{\sum_{b' \in B} F(b')\Pr G(x-b')}$$

Throughout this analysis we will estimate parameters by pooling the choices of all receivers. This is necessary because we have insufficient power to study each individual in isolation. As a consequence, we treat the parameter estimates as coming from a representative agent.

4.4.1 Estimating Math Errors

We assume that math error determines the precision of the signal x, and therefore affects the receiver's posterior beliefs about the secret number. We could impose strong assumptions on the distribution of math errors and try to identify it using receiver decisions in the game, but we choose instead to estimate it out-of-sample for clean identification. In particular, we estimate the distribution of math errors non-parametrically, using the math errors found in the math test completed after playing the game. The questions in this test have a similar level of complexity (N=20) as a report with maximum complexity, but there should be minimal strategic or social considerations when answering these questions, and the payoff function is such that the receiver should report their modal belief of the secret number, regardless of their risk preferences.

We assume that the distribution of additive errors G is symmetric and has support over the integers $\{-9,-8,\ldots,0,\ldots,8,9\}$. This generates enough error for a secret number of 1 to get a signal of 10, and a secret number of 10 to get a signal of 1. Thus, this model has 10 parameters, which are estimated non-parametrically. By assuming that receivers guess their signal, we can identify from guesses and secret numbers the frequency with which each signal is realized.

To estimate G in this way, we used the math test answers for the 160 subjects who completed the math test.[29] The resulting estimate places a large mass (72.5%) on no noise ($e=0$), and the average parameter is 4.8 percentage points from the corresponding

[29] Because we do not observe guesses when subjects hit the time limit, we exclude these decisions from the estimation.

parameter in a distribution that places all weight on no noise. Our estimate of *G* is presented visually in Figure 6A.

4.4.2 Estimating Strategic Confusion and Social Preferences

Two factors that we will add to our model are strategic confusion and social preferences. By strategic confusion, we mean confusion about how to play the game, and by social preferences, we mean concerns about the payoffs of others.

We estimate the degree of strategic confusion and the social preferences of the subjects jointly, using the guesses of receivers when the message has been reported in a simple manner (N<=5). Again, we deliberately use out-of-sample estimation, in order to shy away from confounding factors such as math error. In doing so, we assume that there are minimal interactions between complexity and strategic confusion or social preferences. In practice, it is likely that social considerations when messages are complex are different from when messages are simple, as receivers may feel some positive reciprocity when simple reports are made. Because of this, our out-of-sample estimate of social preferences will only be added to our model later as a robustness check.

Here we make two functional form assumptions. First, we assume that strategic confusion results in a receiver sometimes guessing in a uniform random way. In the Level-k model of choice, this is often designated as the "Level-0" behavior. Because we are using a representative agent model, this is as if some fraction of agents are Level-0 agents. Second, we assume that a receiver sometimes uses social preferences that take the form proposed by Fehr and Schmidt (1999). Note that only one parameter of this model (advantageous inequality) will have bite. Together this gives us three parameters to estimate: the probability of uniform random guessing, the probability of using social preference, and a parameter of the Fehr-Schmidt model of social preferences.

The parameters of this model were estimated using the Nelder–Mead method, and the standard errors were computed using 1,000 bootstrapping samples. The estimates were a 7.4% probability of uniform random choice (with a standard error of 0.007), a 2.3% probability of using social preferences (with a standard error of 0.005), and a 0.658 advantageous inequality parameter (with a standard error of 0.194).

4.4.3 Baseline Predictions

In our baseline model: receivers hold correct prior beliefs over the distribution of states given a complex report (equal to the empirical frequency in the main sessions); make math errors in accordance with the estimated distribution *G*; understand that their errors come from this distribution, update their beliefs according to Bayes' rule; and then maximize risk-neutral expected utility given their posterior beliefs, but with the estimated probability of strategic confusion (random guessing). Importantly, all of the parameters in this model are estimated out-of-sample.

Even with correct beliefs, this model predicts over-guessing and under-guessing of the extremes because the boundary pushes math errors and strategic errors into the middle of distribution, which then pushes guesses into the middle of the distribution. However, it does not do so symmetrically. Because senders are much more likely to have low secret numbers when they use complexity, receivers should take this into account when they guess, given their uncertainty about the state.

This asymmetry is reflected in the predictions from the model, which are provided in Table 10 along with the predictions given by several variants of this model. For 100,000 simulated draws from the distribution of noise parameters, it predicts over-guessing of 0.712 for low states, -0.181 for middle states, and -1.662 for high states (with an overall average log-likelihood of -1.553). The actual rates of over-guessing were 0.772, 0.096, and -0.891. Because of the strong impact of prior beliefs, the baseline model failed to capture over-guessing for middle states and over-estimated the degree of under-guessing at high states.

A natural robustness check is adding social preferences to the model. Specifically, we add the rate and degree of social preferences estimated out-of-sample, though this is likely to be an overestimate of the actual social preferences for senders who use complex disclosures. For 100,000 simulations, the amended model predicts over-guessing of 0.792 for low states, -0.130 for middle states, and -1.640 for high states. While the model comes closer to predicting the actual rates of over-guessing, the improvements in predicting these rates are small, and the model still fails to capture over-guessing for middle states. In addition, the overall average log-likelihood of -1.547 is only a bit better for the amended model.

Because receivers face uncertainty about secret number, it could be that the model needs to account for the possibility of risk aversion. For this robustness check, we assume that utility takes the CRRA form, which means that we allow a free parameter. To estimate this parameter, we conduct a search over a grid of 100 values between 0 and 1 using again 100,000 simulations, and the standard errors were computed using 100 bootstrapping samples. The parameter that maximizes log-likelihood is set-identified, and the lower bound is 0.010 and the upper bound is 0.129.[30] As Table 10 shows, adding risk aversion to the baseline model does not noticeability improve the overall average log-likelihood or the predictions of over-guessing.

4.4.4 Naiveté

As discussed previously, the leading assumption in theories of complex disclosure that produces the incentives to use complexity is naiveté about the strategic use of complexity. As a starting point, we will add full naiveté to our model by assuming that receivers think that all states are equally likely. In the Level-k approach, this often constitutes Level-1 beliefs: that opponents are guessing randomly.

This change to the model improves fit tremendously over the baseline model. Again, based on simulations of 100,000 decisions, the overall average log-likelihood falls from -1.552 to -1.294. The amended model (still with no free parameters) now predicts over-guessing of 0.873 for low states, 0.089 for middle states, and -0.821 for high states, where the actual rates of over-guessing were 0.772, 0.096, and -0.891.

While naiveté of this degree is relatively good at explaining choices, it is not consistent with the stated beliefs of subjects. However, it is possible that the reported beliefs of subjects are not the beliefs used by subjects to play the game. It could be that subjects change their thinking when responding to belief elicitation questions (a possibility raised by Costa-Gomes and Weizsacker 2008). Also, because we elicit beliefs

---

[30] The risk aversion parameter is set identified because changes in the parameter value lead to discontinuous changes in the choice probabilities.

at the end of the experiment, it could be that subjects are overweighting their experience in the final rounds. However, we observe similar rates of over-guessing in the final rounds, which suggests that the reasons behind receiver mistakes should persist into the final rounds as well.

4.4.5 Overconfidence

If math errors drive the precision of the signal, what matters for belief updating is the *perceived* precision of the signal, which depends on how confident the subject is about her own math ability. To determine the degree of overconfidence, we compare beliefs about performance on the math test to actual performance on the math test. This form of absolute overconfidence is called "overestimation" by Moore and Healy (2008), who find absolute overconfidence is more likely in difficult tasks and less likely in easier tasks. Specifically, we use the percentage of subjects who thought they performed "well" (more than 50% correct) on the math test. While 72.5% think performed well, only 63.8% actually performed well.

Using this estimate, we amend the baseline model to assume that receivers think they have a 72.5% chance of performing well at math task. In other words, the representative agent believes that there is a 72.5% chance that the error came from a distribution *G'*, which is estimated non-parametrically from the math test using the answers of subjects who actually performed well at the math test. This distribution is shown in Figure 6B.

Based on simulations of 100,000 decisions, the overall average log-likelihood falls from -1.552 to -1.272, which is even higher than the log-likelihood of -1.294 from the model with naive receivers. The predictions for over-guessing are 0.749 for low states, 0.018 for middle states, and -0.904 for high states, where the actual rates of over-guessing were 0.772, 0.096, and -0.891.

We also consider an alternative method for estimating overconfidence, which is inspired by the approach for determining distortions of Bayes' rule used in Grether (1980) and Holt and Smith (2009). Our approach, which has a free parameter, is to assume that when updating beliefs the probability that a signal is observed in a certain state is raised to the power of the parameter. When this parameter is equal to 1, the receiver updates Bayes' rule in the standard fashion. When this parameter is greater than 1, if a signal is more likely in a state (such as the probability of receiving a signal of 7 when the true state is 7), then more weight is given to the state given this signal (such as the probability the true state is 7 given a signal of 7). To estimate this parameter, we conducted a search over a grid of 100 values between 1 and 30 using 100,000 simulations, and the standard errors were computed using 100 bootstrapping samples. The parameter value that maximizes log-likelihood is also set-identified, but the lower bound is 16.929, which is far from the Bayesian value of 1.

This approach also does a good job at explaining receiver guesses. The overall average log-likelihood is -1.261, which is a bit better than the log-likelihood of -1.272 from the model with absolute overconfidence (though that model does not have free parameter). The predictions for over-guessing are 0.776 for low states, 0.050 for middle states, and -0.800 for high states, where the actual rates of over-guessing were 0.772, 0.096, and -0.891.

4.4.6 Other Possible Explanations

Naiveté and overconfidence are not the only behavioral biases that could potentially explain receiver guesses. For instance, subjects could fall prey to "wishful thinking" by believing that the secret number is higher because that would lead to socially better outcomes. Another possibility is that subjects are placed under a "cognitive load" when summing up numbers, which causes them to make mistakes in strategic inference.

Another potential explanation that has a long history in the behavioral economics literature is base rate neglect, which is documented in belief updating using a ball-and-urns task by Grether (1980) and Holt and Smith (2009). While many reasons for base rate neglect have been provided in the literature, one reason why base rate neglect could occur in our experiment is that subjects might focus entirely on the outcome of the summation task, which causes them to overlook the base rate (their prior beliefs) when making decisions. In explaining choice, base rate neglect operates very similarly to naiveté, but differs in that it could explain why receivers act as if they have a uniform prior even if they have skeptical beliefs. We estimated a variant of our baseline model with overconfidence and a parameter for base rate neglect. That model has a similar likelihood to the one with just overconfidence and is actually worse at predicting the bias in mistakes. Together, this suggests that there is little room for base rate neglect if the overconfidence in the game is similar to what we observe in the math tests and impacts receiver guesses in the way we have specified.

4.5 Endogenous Attention and Response Times

In our model of receiver decision-making, receivers do not choose whether or not to receive a signal about the true state. In practice, receivers may incur a cost to receive this signal, so they may decide it is not worth obtaining the signal at all. As proposed by Caplin and Martin (2016), one way to evaluate the extensive margin of attention is by looking at response times. If subjects have spent almost no time in reaching a decision, it is likely that they were inattentive to the information required to make a decision.

We find that just a small number of subjects make "quick" decisions when reports are complex, which is in contrast to the substantial fraction of quick decisions when subjects choose among strings of numbers in the individual decision-making task of Caplin and Martin (2016). They find that almost 40% of subjects choose in 8 seconds or less in their experiment, but in our experiment, just 1.6% of subjects facing high complexity choose in 8 seconds or less, and just 5.2% choose in 20 seconds or less. For high complexity, the 25th percentile of response times in our experiment is at 33 seconds.

However, like Caplin and Martin (2016), we find that those who make quick decisions choose in line with their beliefs. For subjects who have response times of 33 seconds or less for high complexity, we regress the receiver's guess on their stated beliefs of the average secret number. The coefficient is positive and substantial (0.3411) and is significant at the 1% level ($p<0.001$). This implies that subjects who make quick decisions – those who are intentionally inattentive to complex information – are not guessing wildly, but are instead choosing in line with their prior beliefs.

This interpretation is also consistent with the regression results in Table 8B, which demonstrate a *positive* correlation between time spent and the size of receiver mistakes. Combined with the evidence in the structural estimation, it seems that spending a long time on a complex report could make a receiver more likely to succumb to the biases of naiveté, overconfidence, or base rate neglect.

5. Conclusion and Policy Implications

Our results highlight the incentives for firms to strategically complexify information disclosed to consumers, potentially harming consumers and undermining the effectiveness of disclosure. In our experiment, senders use far more complex disclosure than standard unraveling theory predicts. Most of this obfuscation is profitable because receivers make biased mistakes in assessing complex reports. A model that includes either overconfidence or naiveté can explain receiver mistakes, but stated beliefs suggest that subjects are not naive about the strategic use of complexity. On the other hand, receivers appear overconfident about their ability to assess complex reports.

The patterns we observe have policy implications as well. For example, many obfuscation theories assume naiveté in (a fraction of) consumers, hence consumer education that reduces naiveté should alleviate the seller's incentives to obfuscate. In contrast, receivers in our experiment are sophisticated enough to realize the strategic incentives behind a sender's complexity choice. But that sophistication does not save them from obfuscation, because they are overconfident about their ability to comprehend complex reports. Policy tools that target such overconfidence can be different from education efforts that target consumer naiveté. If both naiveté and overconfidence are difficult to tame, a mandate on simplicity can be as important as a mandate on truthful disclosure. More generally, this highlights the potential for regulation aimed at encouraging disclosure that is simple and salient.

Another policy implication is seen in sender behavior. Surprisingly, round-by-round feedback does not reduce obfuscation. If anything, learning encourages sellers to understand receiver mistakes in low states and exploit it via obfuscation.

A final policy implication is related to disclosure in general. Our results suggest that the unraveling prediction is fragile. Although immediate and repeated feedback can steer voluntary disclosure towards the predictions of unraveling, it fails once we change the setting a little away from simple, voluntary disclosure. How to harvest the benefits of the incentives produced by unraveling remains a challenge in the real world.

References

Agranov, M., & Schotter, A. (2012). Ignorance is bliss: an experimental study of the use of ambiguity and vagueness in the coordination games with asymmetric payoffs. *American Economic Journal: Microeconomics*, *4*(2), 77-103.

Anderson, M. L., Chiswell, K., Peterson, E. D., Tasneem, A., Topping, J., & Califf, R. M. (2015). Compliance with results reporting at ClinicalTrials. gov. *New England Journal of Medicine*, *372*(11), 1031-1039.

Armstrong, M., & Chen, Y. (2009). Inattentive consumers and product quality. *Journal of the European Economic Association*, *7*(2-3), 411-422.

Armstrong, M., & Vickers, J. (2012). Consumer protection and contingent charges. *Journal of Economic Literature*, *50*(2), 477-93.

Bederson, B. B., Jin, G. Z., Leslie, P., Quinn, A. J., & Zou, B. (2018). Incomplete disclosure: Evidence of signaling and countersignaling. *American Economic Journal: Microeconomics*, *10*(1), 41-66.

Ben-Shahar, O., & Schneider, C. E. (2014). *More than you wanted to know: The Failure of Mandated Disclosure*. Princeton University Press.

Blake, T., Moshary, S., Sweeney, K., & Tadelis, S. (2017). Price Salience and Product Choice. Working paper presented at the 2017 NBER Digitization meeting.

Board, O. (2009). Competition and disclosure. *The Journal of Industrial Economics*, *57*(1), 197-213.

Bollinger, B., Leslie, P., & Sorensen, A. (2011). Calorie posting in chain restaurants. *American Economic Journal: Economic Policy*, *3*(1), 91-128.

Brown, J., Hossain, T., & Morgan, J. (2010). Shrouded attributes and information suppression: Evidence from the field. *Quarterly Journal of Economics*, *125*(2), 859-876.

Cai, H., & Wang, J. T. Y. (2006). Overcommunication in strategic information transmission games. *Games and Economic Behavior*, *56*(1), 7-36.

Caplin, A., Dean, M., & Martin, D. (2011). Search and satisficing. *American Economic Review*, *101*(7), 2899-2922.

Caplin, A., & Martin, D. (2016). The Dual-Process Drift Diffusion Model: Evidence from Response Times. *Economic Inquiry*, *54*(2), 1274-1282.

Carlin, B. I. (2009). Strategic price complexity in retail financial markets. *Journal of Financial Economics*, *91*(3), 278-287.

Carlin, B. I., Kogan, S., & Lowery, R. (2013). Trading complex assets. *Journal of Finance*, *68*(5), 1937-1960.

Célérier, C., & Vallée, B. (2017). Catering to investors through security design: Headline rate and complexity. *Quarterly Journal of Economics*, *132*(3), 1469-1508.

Chetty, R., Looney, A., & Kroft, K. (2009). Salience and taxation: Theory and evidence. *American Economic Review*, *99*(4), 1145-77.

Costa-Gomes, M. A., & Weizsäcker, G. (2008). Stated beliefs and play in normal-form games. *Review of Economic Studies, 75*(3), 729-762.

Crawford, V. P., & Iriberri, N. (2007). Level‐k Auctions: Can a Nonequilibrium Model of Strategic Thinking Explain the Winner's Curse and Overbidding in Private‐Value Auctions? *Econometrica, 75*(6), 1721-1770.

Crawford, V. P., & Sobel, J. (1982). Strategic information transmission. *Econometrica*, 1431-1451.

DellaVigna, S., & Pollet, J. M. (2005). Attention, demographics, and the stock market. National Bureau of Economic Research Working Paper w11211.

DellaVigna, S., & Pollet, J. M. (2009). Investor inattention and Friday earnings announcements. *Journal of Finance*, *64*(2), 709-749.

Dranove, D., & Jin, G. Z. (2010). Quality disclosure and certification: Theory and practice. *Journal of Economic Literature*, *48*(4), 935-63.

Dranove, D., Kessler, D., McClellan, M., & Satterthwaite, M. (2003). Is more information better? The effects of “report cards” on health care providers. *Journal of Political Economy*, *111*(3), 555-588.

Eil, D., & Rao, J. M. (2011). The good news-bad news effect: asymmetric processing of objective information about yourself. *American Economic Journal: Microeconomics*, *3*(2), 114-38.

Ellison, G. (2005). A model of add-on pricing. *Quarterly Journal of Economics*, *120*(2), 585-637.

Ellison, G., & Ellison, S. F. (2009). Search, obfuscation, and price elasticities on the internet. *Econometrica*, *77*(2), 427-452.

Ellison, G., & Wolitzky, A. (2012). A search cost model of obfuscation. *The RAND Journal of Economics*, *43*(3), 417-441.

Eyster, E., & Rabin, M. (2005). Cursed equilibrium. *Econometrica*, *73*(5), 1623-1672.

Fehr, E., & Schmidt, K. M. (1999). A theory of fairness, competition, and cooperation. *Quarterly Journal of Economics*, *114*(3), 817-868.

Feltovich, N., Harbaugh, R., & To, T. (2002). Too cool for school? Signalling and countersignalling. *RAND Journal of Economics*, 630-649.

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

Fung, A., Graham, M., & Weil, D. (2007). *Full disclosure: The perils and promise of transparency*. Cambridge University Press.

Gabaix, X., & Laibson, D. (2006). Shrouded attributes, consumer myopia, and information suppression in competitive markets. *Quarterly Journal of Economics, 121*(2), 505-540.

Grether, D. M. (1980). Bayes rule as a descriptive model: The representativeness heuristic. *Quarterly Journal of Economics*, *95*(3), 537-557.

Grossman, S. J. (1981). The informational role of warranties and private disclosure about product quality. *Journal of Law and Economics*, *24*(3), 461-483.

Grossman, S. J., & Hart, O. D. (1980). Disclosure laws and takeover bids. *Journal of Finance*, *35*(2), 323-334.

Grubb, M. D. (2011). Developing a reputation for reticence. *Journal of Economics & Management Strategy*, *20*(1), 225-268.

Grubb, M. D. (2015). Overconfident consumers in the marketplace. *Journal of Economic Perspectives*, *29*(4), 9-36.

Hanna, R., Mullainathan, S., & Schwartzstein, J. (2014). Learning through noticing: Theory and evidence from a field experiment. *Quarterly Journal of Economics*, *129*(3), 1311-1353.

Heidhues, P., Kőszegi, B., & Murooka, T. (2016). Inferior products and profitable deception. *Review of Economic Studies*, *84*(1), 323-356.

Hirshleifer, D., & Teoh, S. H. (2003). Limited attention, information disclosure, and financial reporting. *Journal of Accounting and Economics*, *36*(1-3), 337-386.

Holt, C. A., & Smith, A. M. (2009). An update on Bayesian updating. *Journal of Economic Behavior & Organization, 69*(2), 125-134.

Jacob, B. A., & Levitt, S. D. (2003). Rotten apples: An investigation of the prevalence and predictors of teacher cheating. *Quarterly Journal of Economics*, *118*(3), 843-877.

Jin, G. Z. (2005). Competition and disclosure incentives: an empirical study of HMOs. *RAND Journal of Economics*, 93-112.

Jin, G. Z., Luca, M., & Martin, D. (2015). Is no news (perceived as) bad news? An experimental investigation of information disclosure. National Bureau of Economic Research Working Paper w21099.

Jovanovic, B. (1982). Truthful disclosure of information. *Bell Journal of Economics*, 36-44.

Kalaycı, K., & Potters, J. (2011). Buyer confusion and market prices. *International Journal of Industrial Organization*, *29*(1), 14-22.

Lacetera, N., Pope, D. G., & Sydnor, J. R. (2012). Heuristic thinking and limited attention in the car market. *American Economic Review*, *102*(5), 2206-36.

Lowenstein, G., Sunstein, C.R., and Golman, R. (2014). Disclosure: Psychology Changes Everything. *Annual Review of Economics, 6*, 391-419.

Feng Lu, S. (2012). Multitasking, information disclosure, and product quality: Evidence from nursing homes. *Journal of Economics & Management Strategy*, *21*(3), 673-705.

Luca, M., & Smith, J. (2013). Salience in quality disclosure: evidence from the US News college rankings. *Journal of Economics & Management Strategy*, *22*(1), 58-77.

Luca, M., & Smith, J. (2015). Strategic disclosure: The case of business school rankings. *Journal of Economic Behavior & Organization, 112*, 17-25.

Marinovic, I., & Varas, F. (2016). No news is good news: Voluntary disclosure in the face of litigation. *The RAND Journal of Economics*, *47*(4), 822-856.

Martin, D. (2015). Rational Inattention in Games: Experimental Evidence. Available at SSRN: http://ssrn.com/abstract=2674224.

Martin, D., & Munoz Rodriguez, E. (2018). Inattention to Game Form: A Theory of the WTA/WTP Gap. Mimeo.

Mathios, A. D. (2000). The impact of mandatory disclosure laws on product choices: An analysis of the salad dressing market. *The Journal of Law and Economics*, *43*(2), 651-678.

Matthews, S., & Postlewaite, A. (1985). Quality testing and disclosure. *The RAND Journal of Economics*, 328-340.

McKelvey, R. D., & Palfrey, T. R. (1995). Quantal response equilibria for normal form games. *Games and Economic Behavior*, *10*(1), 6-38.

Milgrom, P. R. (1981). Good news and bad news: Representation theorems and applications. *The Bell Journal of Economics*, 380-391.

Mobius, M. M., Niederle, M., Niehaus, P., & Rosenblat, T. S. (2011). Managing self-confidence: Theory and experimental evidence. National Bureau of Economic Research Working Paper w17014.

Moore, D. A., & Healy, P. J. (2008). The trouble with overconfidence. *Psychological Review*, *115*(2), 502.

Mullainathan, S., Schwartzstein, J., & Shleifer, A. (2008). Coarse thinking and persuasion. *Quarterly Journal of Economics*, *123*(2), 577-619

Perez-Richet, E., & Prady, D. (2011). Complicating to persuade. Mimeo.

Pope, D. G. (2009). Reacting to rankings: evidence from “America's Best Hospitals”. *Journal of Health Economics*, *28*(6), 1154-1165.

Serra-Garcia, M., van Damme, E., & Potters, J. (2011). Hiding an inconvenient truth: Lies and vagueness. *Games and Economic Behavior*, *73*(1), 244-261.

Sims, C. A. (2003). Implications of rational inattention. *Journal of monetary Economics*, *50*(3), 665-690.

Spiegler, R. (2006). Competition over agents with boundedly rational expectations. *Theoretical Economics*, *1*(2), 207-231.

Sullivan, M. (2017). Economic Issues: Economic Analysis of Resort Fees. FTC Economic Issue Paper.

Trautmann, S. T., & Kuilen, G. (2015). Belief elicitation: A horse race among truth serums. *Economic Journal*, *125*(589), 2116-2135.

Viscusi, W. K. (1978). A note on" lemons" markets with quality certification. *The Bell Journal of Economics*, 277-279.

Figure 1A. Sender choice of complexity by secret number (main sessions)

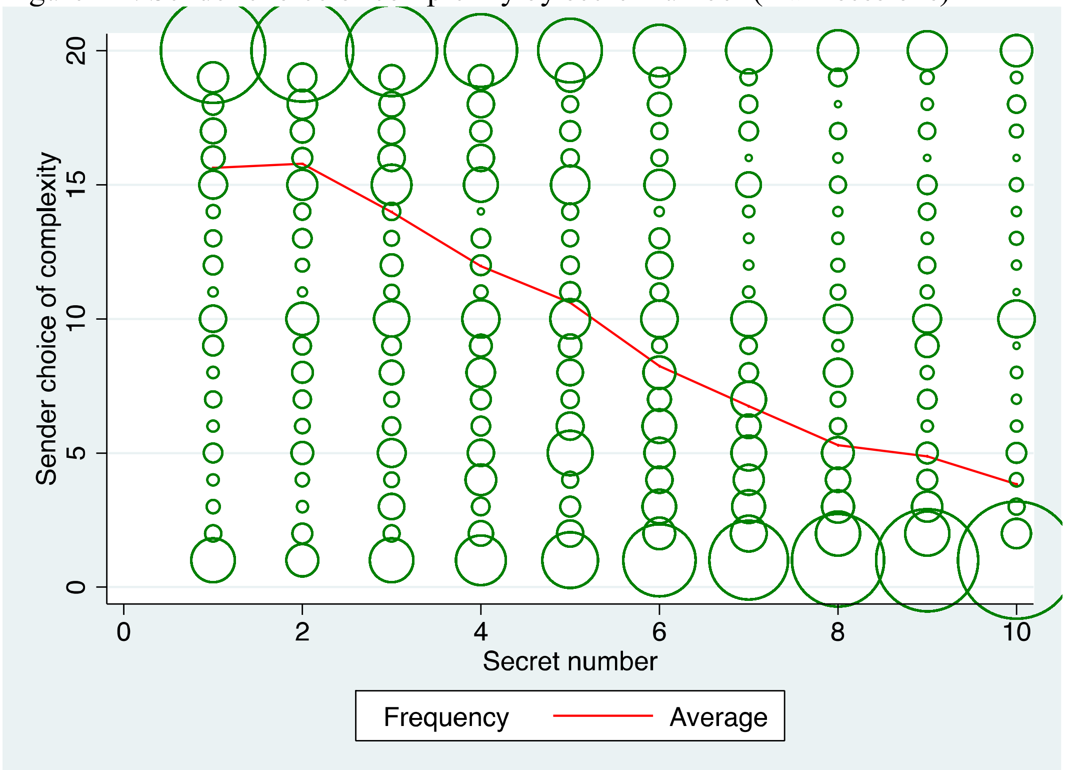

Figure 1B. Sender choice of complexity by secret number in second half of rounds (main sessions)

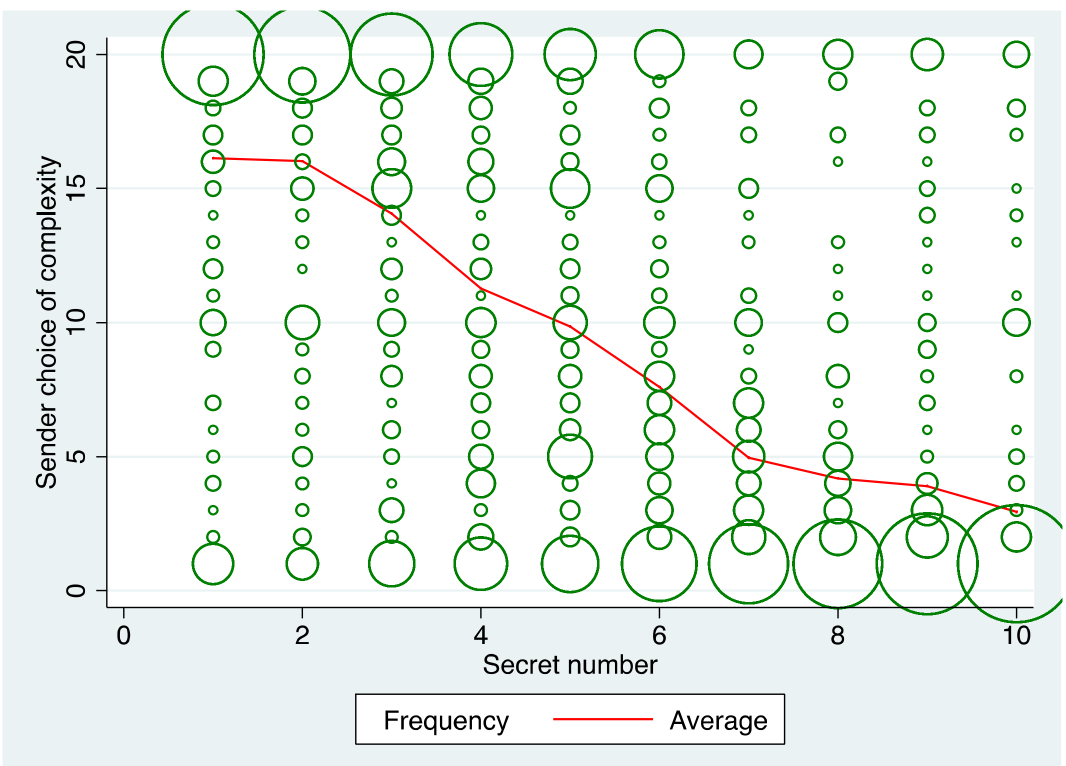

Figure 2A: Sender choice of complexity and stated beliefs of average sender choice of complexity by secret number (main sessions)

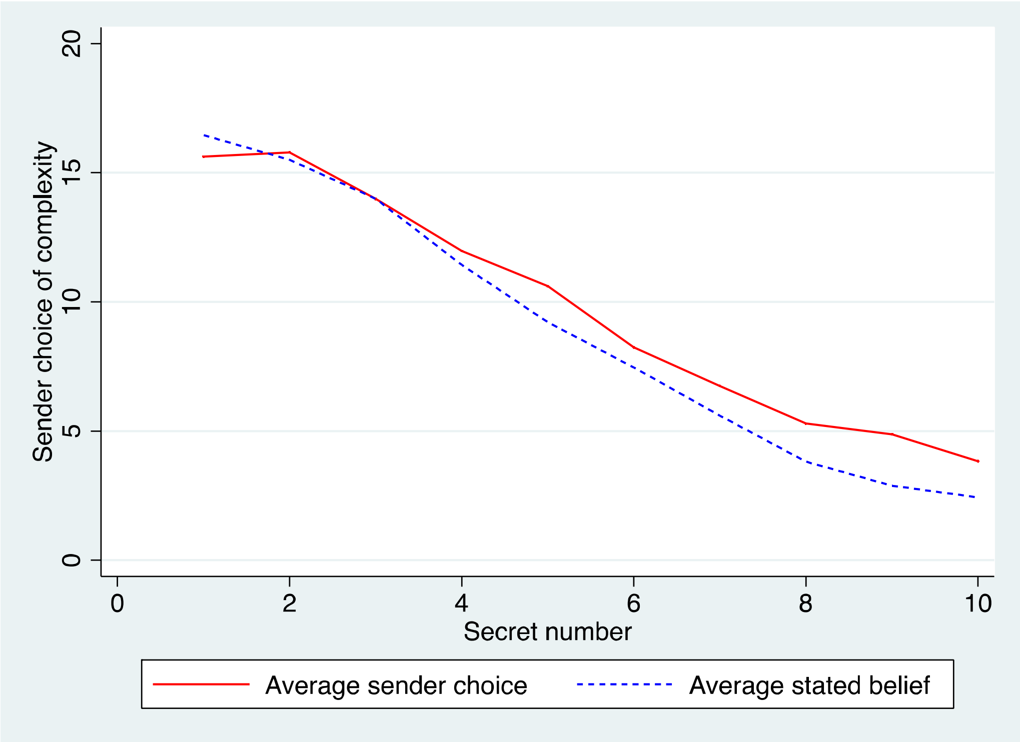

Figure 2B: Average secret number and stated beliefs of average secret number by complexity of 1-5, 6-10, 11-15, or 16-20 (main sessions)

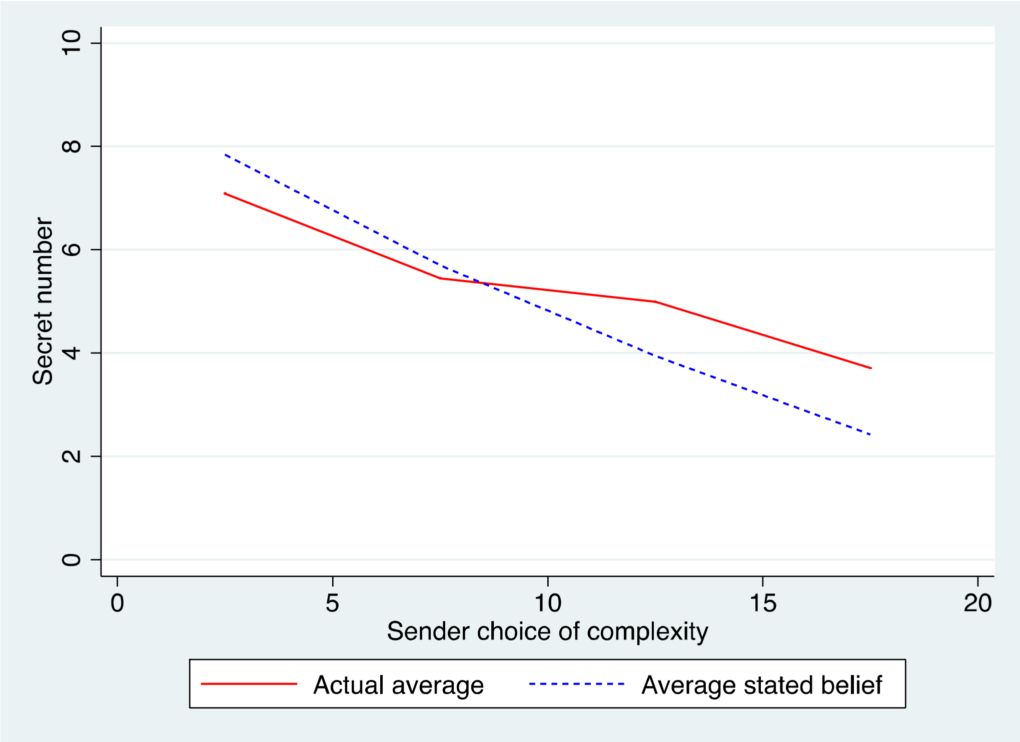

Figure 3: Math test performance and stated beliefs of math test performance

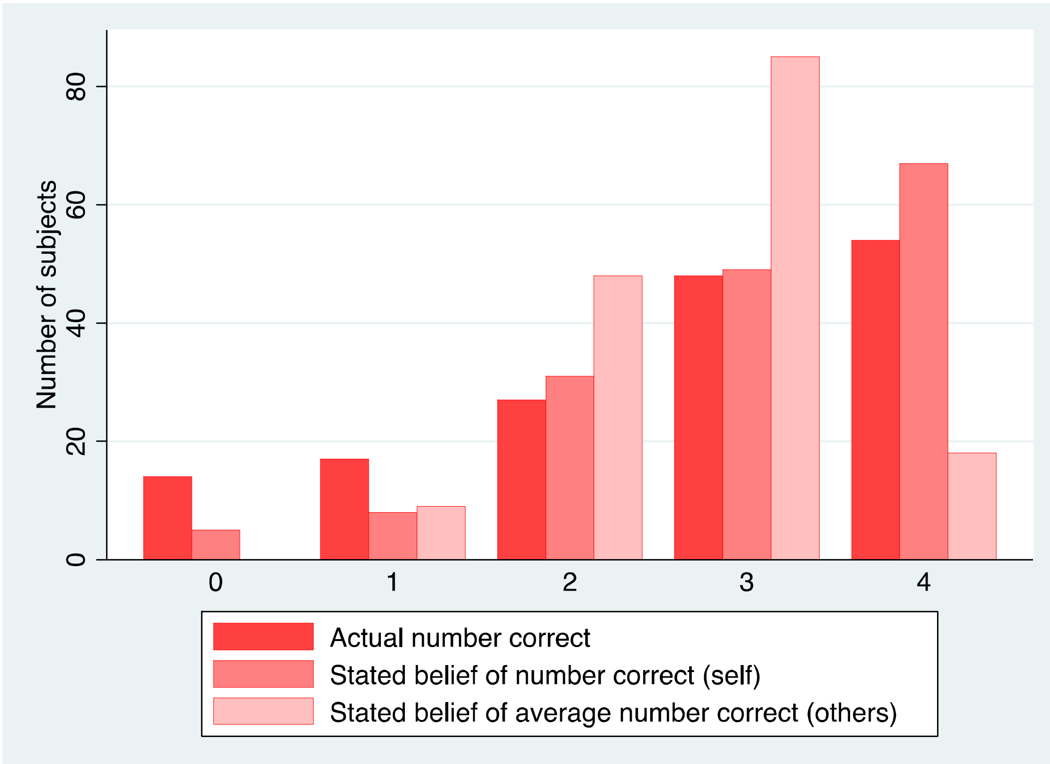

Figure 4. Average receiver mistake size (|guess – truth|) by round (main sessions)

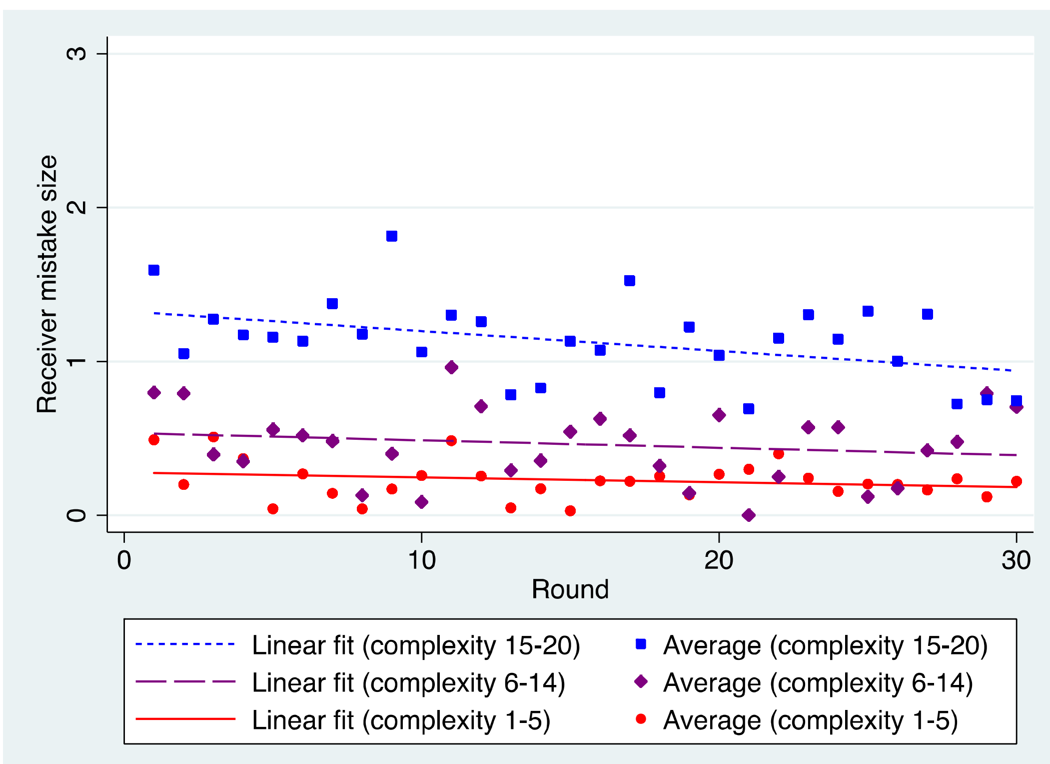

Figure 5: Average secret number and receiver mistake size (|guess – truth|) by sender choice of complexity (main and robustness sessions)

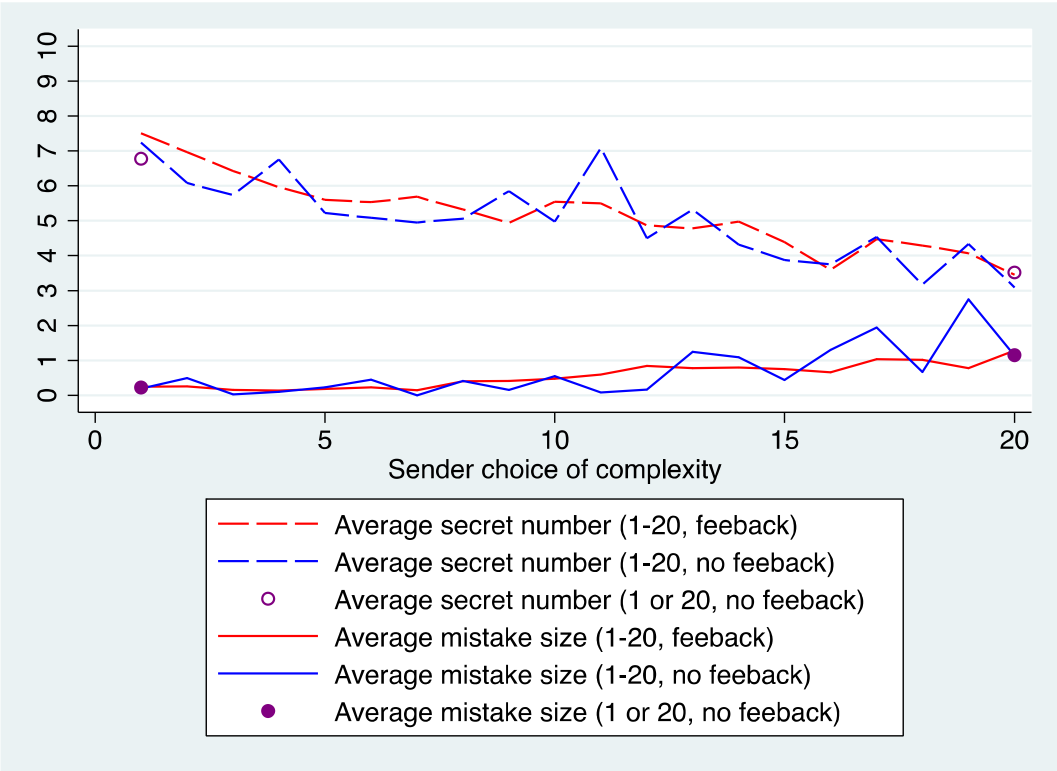

Figure 6A. Non-parametrically estimated distribution of additive error term

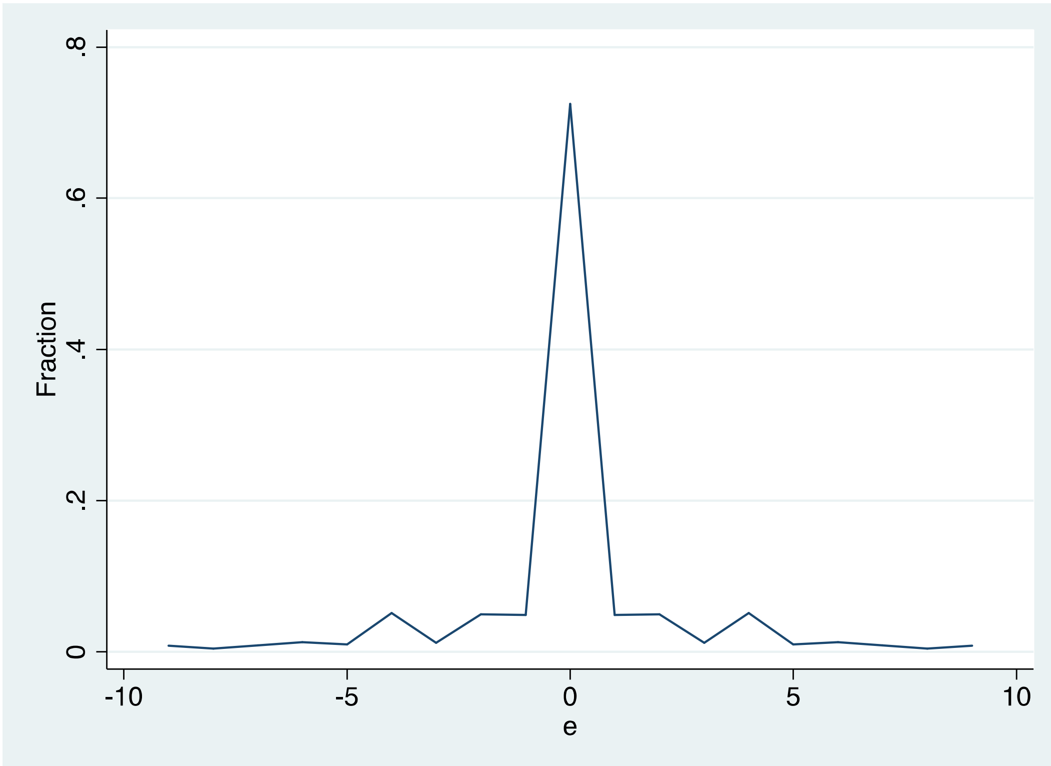

Figure 6A. Non-parametrically estimated distribution of additive error term for subjects who answered more than 50% of questions correctly on math test

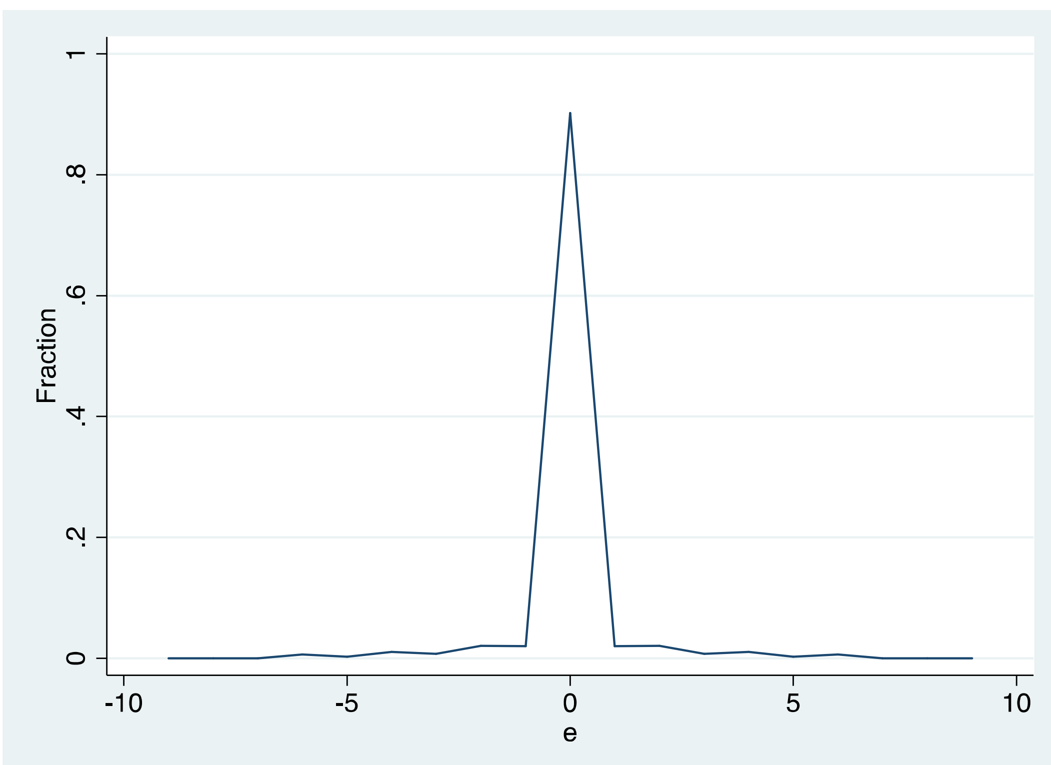

Table 1. Summary of subject characteristics (main sessions)

Variable	N	Mean	Std. dev.
Number of subjects in the session	294	10.680	2.554
Feedback provided (dummy)	294	1.000	0.000
Male (dummy)	293	0.410	0.493
Undergraduate (dummy)	293	0.720	0.450
Native English speaker (dummy)	290	0.852	0.356
Friend in the session (dummy)	293	0.143	0.351

Note: Observation is per subject. Value is missing if demographic information not provided by the subject.

Table 2A. Summary of sender choice of complexity by secret number (main sessions)

		Sender choice of complexity			High complexity (length>=15)	Low complexity (length<=5)
Secret number	N	Mean	Median	Std. dev.	Mean	Mean
1	449	15.626	20	6.617	0.728	0.145
2	444	15.782	20	6.157	0.721	0.115
3	464	13.983	17	6.837	0.616	0.19
4	422	11.969	13	7.218	0.486	0.275
5	433	10.607	10	7.13	0.390	0.344
6	453	8.243	6	6.914	0.254	0.455
7	424	6.748	4	6.664	0.198	0.583
8	427	5.286	2	6.288	0.141	0.71
9	447	4.879	1	6.197	0.128	0.729
10	447	3.832	1	5.622	0.094	0.796
Total	4410	9.728	9	7.86	0.378	0.432

Table 2B: Summary of receiver guess by secret number (main sessions)

Secret number	N	Receiver guess			Receiver mistake bias (guess-truth)	Receiver mistake size (|guess-truth|)	% of receiver decisions before time limit	Conditional on receiver decision before time limit: Receiver mistake bias (guess-truth)	Conditional on receiver decision before time limit: Receiver mistake size (|guess-truth|)
		Mean	Median	Std. dev.	Mean	Mean	Mean	Mean	Mean
1	449	2.183	1	2.326	1.183	1.183	5.57%	0.946	0.946
2	444	2.923	2	2.209	0.936	1.045	8.11%	0.659	0.777
3	464	3.399	3	1.462	0.399	0.601	5.39%	0.328	0.492
4	422	4.232	4	1.458	0.232	0.611	3.08%	0.191	0.538
5	433	5.169	5	1.378	0.169	0.566	3.23%	0.146	0.489
6	453	6.031	6	1.167	0.031	0.446	3.97%	0.051	0.377
7	424	6.887	7	1.234	-0.113	0.424	2.12%	-0.067	0.376
8	427	7.724	8	1.289	-0.276	0.407	1.17%	-0.237	0.37
9	447	8.633	9	1.377	-0.367	0.438	2.01%	-0.311	0.379
10	447	9.597	10	1.574	-0.403	0.403	0.67%	-0.372	0.372
Total	4410	5.663	6	2.885	0.182	0.614	3.56%	0.128	0.509

Table 3. Summary of receiver guess by sender choice of complexity (main sessions)

		All receiver decisions Mean values						Conditional on receiver decision before time limit Mean values	
Complexity	N	Secret number	Receiver guess	Receiver mistake bias (guess-truth)	Receiver mistake size (|guess-truth|)	% before time limit	Response time if before time limit	Receiver mistake bias (guess-truth)	Receiver mistake size (|guess-truth|)
1	1259	7.504	7.466	-0.038	0.243	0.40%	9.15	-0.038	0.236
2	214	6.967	6.925	-0.042	0.257	0.00%	8.95	-0.042	0.257
3	140	6.429	6.407	-0.021	0.15	0.00%	13.21	-0.021	0.150
4	104	5.962	5.885	-0.077	0.135	0.00%	13.35	-0.077	0.135
5	190	5.600	5.684	0.084	0.179	0.00%	18.15	0.084	0.179
6	91	5.527	5.582	0.055	0.231	1.10%	18.85	0.089	0.200
7	89	5.685	5.629	-0.056	0.146	1.12%	21.5	-0.023	0.114
8	117	5.325	5.299	-0.026	0.402	0.85%	23.67	-0.052	0.379
9	74	4.932	5.068	0.135	0.405	0.00%	25.45	0.135	0.405
10	263	5.54	5.388	-0.152	0.479	1.90%	28.77	-0.143	0.438
11	42	5.500	5.476	-0.024	0.595	2.38%	34.25	0.049	0.537
12	69	4.87	4.783	-0.087	0.841	1.45%	35.54	0.000	0.765
13	54	4.778	5.222	0.444	0.778	0.00%	35.56	0.444	0.778
14	39	4.974	5.513	0.538	0.795	2.56%	37.08	0.632	0.737
15	190	4.384	4.463	0.079	0.753	3.16%	36.55	0.071	0.712
16	71	3.592	4.000	0.408	0.662	7.04%	37.21	0.273	0.424
17	90	4.467	4.789	0.322	1.033	5.56%	40.32	0.306	0.847
18	96	4.292	4.573	0.281	1.01	9.38%	42.45	0.195	0.839
19	115	4.07	4.296	0.226	0.783	6.96%	40.33	0.243	0.748
20	1103	3.455	4.11	0.655	1.284	9.79%	42.76	0.477	1.008
Total	4410	5.482	5.664	0.182	0.614	3.56%	24.93	0.128	0.509

Table 4. Summary of receiver mistake size by secret number and sender choice of complexity (main sessions)

Secret number	All receiver decisions Mean values of receiver mistake size (|guess-truth|) Low complexity (1-5)	Medium complexity (6-14)	High complexity (15-20)	Conditional on receiver decision before time limit Mean values of receiver mistake size (|guess-truth|) Low complexity (1-5)	Medium complexity (6-14)	High complexity (15-20)
1	0.6	0.386	1.437	0.6	0.386	1.126
2	0.216	0.795	1.234	0.216	0.795	0.873
3	0.273	0.144	0.846	0.273	0.124	0.691
4	0.198	0.376	0.961	0.198	0.35	0.839
5	0.181	0.426	1	0.162	0.372	0.88
6	0.204	0.432	0.896	0.19	0.392	0.74
7	0.142	0.366	1.321	0.138	0.33	1.18
8	0.228	0.672	0.033	0.228	0.548	0.93
9	0.23	0.641	1.404	0.222	0.603	1.098
10	0.239	0.776	1.357	0.239	0.776	1.077
Total	0.225	0.469	1.133	0.221	0.434	0.91

Table 5: Departure from highest expected payoff (main sessions)

Panel A: Senders

Secret number	Fraction of payoff loss from highest expected payoff given empirical distribution of opponent behavior	Fraction of payoff loss from payoff in the unraveling equilibrium
1	0.516	.*
2	0.320	-0.714
3	0.152	-0.160
4	0.110	-0.043
5	0.103	-0.016
6	0.073	0.006
7	0.077	0.028
8	0.078	0.039
9	0.059	0.041
10	0.038	0.039
Total	0.153	-0.088

Panel B: Receivers

Complexity	Fraction of payoff loss from highest expected payoff given empirical distribution of opponent behavior	Fraction of payoff loss from payoff in the unraveling equilibrium
Low (1-5)	0.138	0.299
Medium (6-14)	0.160	0.330
High (15-20)	0.167	0.311
Total	0.153	0.308

* In the unraveling equilibrium, senders with a secret number of 1 earn the minimum possible payoff. After normalizing this payoff to 0, it is not possible to calculate the fraction of payoff loss from zero.

Table 6: Summary of receiver guess and stated beliefs (main sessions)

Panel A: Inferred guess (secret number inferred from stated beliefs of sender choices)

	All received decisions			Conditional on before time limit		
	Secret number	Receiver guess	Inferred guess	Secret number	Receiver guess	Inferred guess
Complexity	Mean	Mean	Mean	Mean	Mean	Mean
Low (1-5)	7.091	7.064	7.845	7.091	7.064	7.849
Medium (6-14)	5.338	5.344	4.893	5.326	5.354	4.891
High (15-20)	3.712	4.222	2.546	3.72	4.097	2.526

Panel B: Complex guess (stated belief of average secret number for a given complexity)

	All receiver decisions			Conditional on before time limit		
	Secret number	Receiver guess	Complex guess	Secret number	Receiver guess	Complex guess
Complexity	Mean	Mean	Mean	Mean	Mean	Mean
1-5	7.091	7.064	7.813	7.091	7.064	7.823
6-10	5.448	5.396	5.756	5.447	5.404	5.756
11-15	4.701	4.835	3.867	4.655	4.818	3.848
16-20	3.626	4.191	2.51	3.636	4.055	2.471

Note: Out of all receiver decisions, 6.8% have a missing value for inferred guess because those subjects indicate that senders will never choose some complexity level.

Table 7: Summary of dynamics (main sessions)

Panel A	Sender choice of complexity			Fraction of sender payoff loss from highest expected payoff		
	Mean			Mean		
Secret number	Round 1-10	Round 11-20	Round 21-30	Round 1-10	Round 11-20	Round 21-30
1	14.454	16.461	16.032	0.551	0.515	0.484
2	15.357	15.993	15.958	0.311	0.322	0.326
3	13.264	15.026	13.693	0.159	0.146	0.150
4	12.673	12.679	10.467	0.107	0.125	0.097
5	11.669	9.878	10.13	0.105	0.105	0.098
6	9.526	7.646	7.545	0.068	0.084	0.066
7	9.475	5.719	5.036	0.091	0.079	0.061
8	6.764	5.218	3.693	0.086	0.081	0.064
9	6.326	5.455	3.093	0.058	0.068	0.050
10	5.829	3.5	2.512	0.053	0.031	0.031
Total	10.624	9.786	8.774	0.159	0.156	0.142
Panel B	Receiver mistake size (|guess-truth|)			Conditional on before time limit		
	Mean			Mean		
Complexity	Round 1-10	Round 11-20	Round 21-30	Round 1-10	Round 11-20	Round 21-30
Low (1-5)	0.254	0.206	0.222	0.247	0.202	0.219
Medium (6-14)	0.472	0.518	0.410	0.442	0.476	0.375
High (15-20)	1.274	1.099	1.004	1.015	0.947	0.751
Total	0.719	0.604	0.520	0.585	0.526	0.418
Panel C	Fraction of receiver payoff loss from highest expected payoff			Conditional on before time limit		
	Mean			Mean		
Complexity	Round 1-10	Round 11-20	Round 21-30	Round 1-10	Round 11-20	Round 21-30
Low (1-5)	0.153	0.126	0.137	0.153	0.126	0.137
Medium (6-14)	0.182	0.148	0.138	0.181	0.148	0.136
High (15-20)	0.189	0.161	0.147	0.183	0.155	0.130
Total	0.174	0.143	0.141	0.171	0.141	0.135

Table 8A: Regressions of sender behavior (main sessions)

	Dependent variable: Complexity		Dependent variable: Payoff departure from the highest	
Secret number = 2	0.161	-0.252	-0.196***	-0.199***
	(0.435)	(0.412)	(0.0231)	(0.0224)
Secret number = 3	-1.458***	-1.297***	-0.360***	-0.360***
	(0.444)	(0.409)	(0.0212)	(0.0208)
Secret number = 4	-1.597**	-1.894***	-0.416***	-0.421***
	(0.647)	(0.601)	(0.0259)	(0.0258)
Secret number = 5	-2.829***	-3.358***	-0.424***	-0.429***
	(0.650)	(0.601)	(0.0254)	(0.0251)
Secret number = 6	-5.176***	-5.405***	-0.449***	-0.454***
	(0.637)	(0.592)	(0.0254)	(0.0250)
Secret number = 7	-5.422***	-5.742***	-0.430***	-0.436***
	(0.618)	(0.586)	(0.0254)	(0.0251)
Secret number = 8	-7.054***	-7.393***	-0.434***	-0.435***
	(0.610)	(0.583)	(0.0250)	(0.0246)
Secret number = 9	-7.355***	-7.614***	-0.452***	-0.453***
	(0.609)	(0.587)	(0.0248)	(0.0247)
Secret number = 10	-8.321***	-8.278***	-0.473***	-0.479***
	(0.608)	(0.591)	(0.0250)	(0.0247)
First 5 rounds	-0.293	-0.387	0.00814	0.00683
	(0.349)	(0.318)	(0.0106)	(0.0107)
Round	0.0458**	0.0409*	-0.000529	-0.000601
	(0.0228)	(0.0216)	(0.00103)	(0.00101)
Round * (4<=secret number <=6)	-0.141***	-0.130***	0.000609	0.000685
	(0.0298)	(0.0274)	(0.00108)	(0.00108)
Round * (secret number>=7)	-0.219***	-0.219***	-0.000316	-0.000298
	(0.0266)	(0.0257)	(0.00105)	(0.00104)
Actual number of math test questions correct (out of 4)	0.0841		-0.00560	
	(0.199)		(0.00649)	
Belief of number correct (self) - actual number correct	0.378		-0.000392	
	(0.231)		(0.00745)	
Belief of number correct (others) - belief of number correct (self)	0.694***		0.0170***	
	(0.214)		(0.00640)	
Individual demographics	Yes	No	Yes	No
Individual fixed effects	No	Yes	No	Yes
Observations	4,410	4,410	4,399	4,399
R-squared	0.352	0.529	0.384	0.438

Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1. In Session 34, receivers' actual play is such that the highest payoff for draw=1 is 0 after our normalization, so we cannot calculate fraction of payoff departure from 0. That is why columns (3) and (4) have 11 less observations. Regressions without individual fixed effects include dummies indicating whether demographics or math test measures are missing.

Table 8B: Regressions of receiver behavior (main sessions)

	Dependent variable: Receiver mistake size (|guess-truth|)		Dependent variable: Payoff departure from the highest expected payoff	
Sender choice of complexity = 2	-0.0330	0.0289	-0.0228*	-0.0123
	(0.0693)	(0.0622)	(0.0128)	(0.0131)
Sender choice of complexity = 3	-0.143**	-0.0642	-0.0168	-0.0162
	(0.0688)	(0.0713)	(0.0156)	(0.0165)
Sender choice of complexity = 4	-0.183**	-0.141*	0.000702	0.00820
	(0.0772)	(0.0807)	(0.0174)	(0.0189)
Sender choice of complexity = 5	-0.152**	-0.179**	0.0177	0.0205
	(0.0631)	(0.0698)	(0.0148)	(0.0156)
Sender choice of complexity = 6	-0.0869	-0.0121	0.00677	0.0126
	(0.136)	(0.137)	(0.0240)	(0.0242)
Sender choice of complexity = 7	-0.250*	0.0115	0.00625	0.0201
	(0.133)	(0.138)	(0.0234)	(0.0245)
Sender choice of complexity = 8	-0.00843	0.0602	0.0157	0.0238
	(0.152)	(0.145)	(0.0226)	(0.0230)
Sender choice of complexity = 9	-0.0436	-0.0591	0.0512*	0.0515*
	(0.165)	(0.166)	(0.0263)	(0.0285)
Sender choice of complexity = 10	0.0325	0.0979	0.0376*	0.0525**
	(0.141)	(0.140)	(0.0195)	(0.0209)
Sender choice of complexity = 11	0.0418	0.0231	-0.0130	0.00356
	(0.241)	(0.242)	(0.0299)	(0.0329)
Sender choice of complexity = 12	0.386	0.435*	0.0155	0.0291
	(0.243)	(0.249)	(0.0265)	(0.0283)
Sender choice of complexity = 13	0.381	0.536**	0.0605**	0.0682**
	(0.280)	(0.268)	(0.0302)	(0.0312)
Sender choice of complexity = 14	0.354	0.477	0.0804**	0.0958**
	(0.309)	(0.300)	(0.0359)	(0.0377)
Sender choice of complexity = 15	0.398**	0.651***	0.0274	0.0536**
	(0.193)	(0.197)	(0.0233)	(0.0253)
Sender choice of complexity = 16	0.0810	0.280	-0.0115	0.0175
	(0.223)	(0.223)	(0.0300)	(0.0307)
Sender choice of complexity = 17	0.418*	0.736***	0.0734**	0.0896***
	(0.253)	(0.242)	(0.0314)	(0.0319)
Sender choice of complexity = 18	0.445*	0.583**	0.0259	0.0291
	(0.237)	(0.245)	(0.0298)	(0.0310)
Sender choice of complexity = 19	0.354	0.433*	0.0294	0.0575**
	(0.237)	(0.230)	(0.0274)	(0.0281)
Sender choice of complexity = 20	0.631***	0.774***	0.0252	0.0458**
	(0.174)	(0.183)	(0.0207)	(0.0223)
First 5 rounds	-0.0195	-0.0701	0.0137	0.0118
	(0.0749)	(0.0715)	(0.0109)	(0.0109)

Round	-0.00239	-0.00108	-0.000306	-0.000332
	(0.00296)	(0.00281)	(0.000540)	(0.000552)
Round * Medium complexity (6-14)	-0.00569	-0.00773	-0.00147*	-0.00146*
	(0.00609)	(0.00565)	(0.000832)	(0.000869)
Round * High complexity (15-20)	-0.00869	-0.0151***	-0.00159**	-0.00198**
	(0.00563)	(0.00543)	(0.000792)	(0.000789)
Inferred guess for complexity (1-5, 6-14, or 15-20) this round	-0.000979	0.0409**	0.00103	0.00333
	(0.0185)	(0.0194)	(0.00206)	(0.00240)
Actual number of math test questions correct (out of 4)	-0.283***		-0.0136**	
	(0.0499)		(0.00593)	
Belief of number correct (self) - actual number correct	-0.0481		-0.00146	
	(0.0542)		(0.00610)	
Belief of number correct (others) - belief of number correct (self)	0.0178		-0.00592	
	(0.0482)		(0.00614)	
Response time	0.00746***	0.0123***	0.000482	0.000460
	(0.00271)	(0.00286)	(0.000303)	(0.000342)
Individual demographics	Yes	No	Yes	No
Individual fixed effects	No	Yes	No	Yes
Observations	4,253	4,253	4,253	4,253
R-squared	0.123	0.281	0.044	0.127

All regressions are conditional on receivers making a guess within the 60 second time limit. Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1. Regressions without individual fixed effects include dummies indicating whether demographics or math test measures are missing. A dummy variable is included that controls for whether the value of inferred guess is missing.

Table 9A: Comparison of sender choice of complexity in main and robustness sessions

Main sessions: random role, complexity 1 to 20, round-by-round feedback
Robust 1: random role, complexity 1 to 20, no feedback
Robust 2: random role, complexity 1 or 20, no feedback

	Sender choice of complexity Mean values			High complexity (16-20) Fraction of choices			Low complexity (1-5) Fraction of choices		
Secret number	Main sessions	Robust 1	Robust 2	Main sessions	Robust 1	Robust 2	Main sessions	Robust 1	Robust 2
1	15.626	14.382	16.562	0.728	0.637	0.819	0.145	0.176	0.181
2	15.782	14.161	15.485	0.721	0.591	0.762	0.115	0.172	0.238
3	13.983	13.349	14.242	0.616	0.560	0.697	0.190	0.193	0.303
4	11.969	9.018	11.640	0.486	0.234	0.560	0.275	0.324	0.440
5	10.607	8.653	7.861	0.390	0.211	0.361	0.344	0.379	0.639
6	8.243	6.151	7.388	0.254	0.129	0.336	0.455	0.581	0.664
7	6.748	5.989	3.111	0.198	0.126	0.111	0.583	0.632	0.889
8	5.286	3.699	3.446	0.141	0.072	0.129	0.710	0.795	0.871
9	4.879	4.017	3.297	0.128	0.026	0.121	0.729	0.704	0.879
10	3.832	4.606	1.872	0.094	0.138	0.046	0.796	0.755	0.954
Total	9.728	8.490	8.544	0.378	0.276	0.397	0.432	0.464	0.603

Table 9B: Comparison of mean receiver guess in main and robustness sessions

Main sessions: random role, complexity 1 to 20, round-by-round feedback
Robust 1: random role, complexity 1 to 20, no feedback
Robust 2: random role, complexity 1 or 20, no feedback

	Receiver guess			Receiver mistake size (|guess-truth|)			Receiver guess if before time limit			Receiver mistake size (|guess-truth|) if before time limit		
Complexity	Main sessions	Robust 1	Robust 2	Main sessions	Robust 1	Robust 2	Main sessions	Robust 1	Robust 2	Main sessions	Robust 1	Robust 2
Low (1-5)	7.064	6.806	6.787	0.225	0.200	0.221	7.064	6.814	6.787	0.221	0.192	0.221
Medium (6-14)	5.344	5.171	.	0.469	0.508	.	5.354	5.156	.	0.434	0.496	.
High (16-20)	4.222	3.960	.	1.132	1.158	1.148	4.097	3.811	3.720	0.910	0.953	1.005
Total	5.664	5.595	5.625	0.614	0.544	0.589	5.668	5.588	5.625	0.509	0.472	0.518

Table 10: Summary of structural estimation of receiver guesses of high complexity reports before time limit (main sessions)

Variable	Actual	QRE	Baseline	Social Preferences	Risk Aversion	Naiveté	Overconfidence	Overweighting
Mean log-likelihood		-1.733	-1.553	-1.547	-1.553	-1.294	-1.272	-1.261
Total log-likelihood		-2641	-2366	-2357	-2366	-1972	-1939	-1921
Parameter (lower)		0.047			0.010			16.929
Std. error		0.066			0.206			0.644
Parameter (upper)					0.129			22.970
Std. error					0.198			0.703
					Receiver bias (guess-truth) Mean values			
Secret number								
1-3	0.772	0.707	0.712	0.792	0.712	0.873	0.749	0.776
4-7	0.096	0.038	-0.181	-0.130	-0.181	0.089	0.018	0.050
8-10	-0.891	-0.641	-1.662	-1.640	-1.662	-0.821	-0.904	-0.800
Average distance		0.125	0.369	0.332	0.369	0.059	0.038	0.142
					Receiver bias (guess-truth) Mean values			
Secret number								
1	1.126	1.100	0.983	1.069	0.983	1.151	1.001	1.034
2	0.711	0.632	0.706	0.790	0.706	0.859	0.735	0.742
3	0.431	0.335	0.407	0.476	0.407	0.566	0.475	0.517
4	0.249	0.154	0.193	0.258	0.193	0.358	0.271	0.289
5	0.222	0.044	-0.082	-0.033	-0.082	0.093	0.033	0.068
6	0.040	-0.044	-0.287	-0.255	-0.287	-0.092	-0.141	-0.103
7	-0.462	-0.154	-1.172	-1.128	-1.172	-0.357	-0.432	-0.385
8	-0.684	-0.335	-1.368	-1.341	-1.368	-0.564	-0.636	-0.548
9	-0.980	-0.632	-1.725	-1.699	-1.725	-0.860	-0.926	-0.840
10	-1.077	-1.100	-2.009	-1.999	-2.009	-1.147	-1.268	-1.115
Average distance		0.159	0.393	0.370	0.393	0.109	0.091	0.936

Appendix: Instructions used in the lab experiment

Welcome
You are about to participate in an experiment on decision-making, and you will be paid for your participation in cash at the end of the experiment. What you earn depends partly on your decisions, partly on the decisions of others, and partly on chance.

Please silence and put away your cellular phones now. The entire session will take place through your computer terminal. Please do not talk or in any way communicate with other participants during the session. We will start with a brief instruction period. During the instruction period you will be given a description of the main features of the experiment and will be shown how to use the computers. If you have any questions during this period, raise your hand and your question will be answered so everyone can hear.

Instructions
The experiment you are participating in consists of 30 rounds. At the end of the final round, you will be paid the total amount you have accumulated during the course of the session (in addition to the $5 show up fee). Everybody will be paid in private. You are under no obligation to tell others how much you earned.

The currency used during these 30 rounds is what we call "Experimental Currency Units" (ECU). For your final payment, your earnings during these 30 rounds will be converted into dollars at the ratio of 150:1 (150 ECU=$1). They will then be rounded up to the nearest (non-negative) dollar amount.

In the first round, you will be matched with one other person, and you are equally likely to be matched with any other person in the room. You will not know whom you are matched with, nor will the person who is matched with you. One of you will be assigned to be A Player and the other to be the B Player for that round. You are equally likely to be assigned to either role. In the second round, you will once again be randomly matched with one other person (most likely with a different person than in the first round) and randomly assigned a role, and this will be repeated until 30 rounds are complete.
In each round and for every pair, the computer program will generate a secret number that is randomly drawn from the set {1,2,3,4,5,6,7,8,9,10}. The computer will then send the secret number to the A Player.

After being presented with the secret number, the A Player then will choose a report "length", which can be anywhere between 1 and 20. The B Player will be presented with a string of numbers of this length, and this string of numbers will sum up to the secret number. The B Player cannot use scratch paper or a calculator for this calculation.
The string of numbers will not be chosen by the A Player. They will be determined by the computer, which will randomly draw numbers between -10 and +10 such that they add up to the secret number.

After receiving this report, the B Player will guess the value of the secret number. The B Player has 60 seconds to make a decision or a number from the set {1,2,3,4,5,6,7,8,9,10} will be randomly selected to be their guess for that round. The earnings of both players depend on the value of the secret number and the B Player's guess.

The specific earnings are shown in the table below. In each cell of the table, the payoff for the A Player is on the left, and the payoff for the B Player is on the right. As you can see from the table, the A Player earns more when the B Player makes a higher guess, and the B Player earns more when their guess is closer to the secret number.

Payoffs S, R	Secret number: 1	Secret number: 2	Secret number: 3	Secret number: 4	Secret number: 5	Secret number: 6	Secret number: 7	Secret number: 8	Secret number: 9	Secret number: 10
Guess: 1	-54,110	-54,102	-54,90	-54,75	-54,57	-54,38	-54,17	-54,-6	-54,-29	-54,-54
Guess: 2	-29,102	-29,110	-29,102	-29,90	-29,75	-29,57	-29,38	-29,17	-29,-6	-29,-29
Guess: 3	-6,90	-6,102	-6,110	-6,102	-6,90	-6,75	-6,57	-6,38	-6,17	-6,-6
Guess: 4	17,75	17,90	17,102	17,110	17,102	17,90	17,75	17,57	17,38	17,17
Guess: 5	38,57	38,75	38,90	38,102	38,110	38,102	38,90	38,75	38,57	38,38
Guess: 6	57,38	57,57	57,75	57,90	57,102	57,110	57,102	57,90	57,75	57,57
Guess: 7	75,17	75,38	75,57	75,75	75,90	75,102	75,110	75,102	75,90	75,75
Guess: 8	90,-6	90,17	90,38	90,57	90,75	90,90	90,102	90,110	90,102	90,90
Guess: 9	102,-29	102,-6	102,17	102,38	102,57	102,75	102,90	102,102	102,110	102,102
Guess: 10	110,-54	110,-29	110,-6	110,17	110,38	110,57	110,75	110,90	110,102	110,110