Information versus Interpretation: Evidence from an Experiment on Persuasion*

Alisher Batmanov
UC San Diego

Bridget Galaty
UC San Diego

September 17, 2026

Click here for most recent version

Abstract

We experimentally study the relative persuasive effects of having more information versus having the same information but complementing the message with an explanation in a strategic communication environment. Senders recommend an action, and Receivers make a prediction based on a dataset they observe together with the Sender's recommendation. By varying the Sender's information and message format in a 2-by-2 design, we separate the effect of new information from that of new interpretation. We find that an informational advantage is the dominant channel: messages from better-informed Senders pull Receivers' guesses substantially closer to the recommendation, while adding an explanation to a message yields only a small insignificant effect, with or without an informational advantage. This asymmetry persists at every level of task difficulty. When the preferences of Senders and Receivers are aligned, information remains the stronger channel, though explanations now add marginal persuasive power. Consistent with these responses, when Receivers choose which type of message to receive, a majority prefer one from a better-informed Sender over one with an explanation, and the Receivers who prefer the explanation are the least responsive to information.

Keywords: Persuasion, Narratives, Informational Advantage, Laboratory Experiment
JEL classification: C91, D82, D83, D91.

*We are indebted to Emanuel Vespa and Isabel Trevino for invaluable guidance and support on this project. We thank Kirby Nielsen, Denis Shishkin, and Joel Sobel for valuable conversations and constructive comments, as well as the audiences at the Economic Science Association World Meeting in Los Angeles, Behavioral & Experimental Economics Student Conference at UC Santa Barbara, and the UC San Diego Theory-Behavioral-Experimental Graduate Seminar. We are grateful to undergraduate student members of our economic research lab at UC San Diego for outstanding research assistance and to all study participants for their time and patience. This material is based upon work supported by the National Science Foundation Graduate Research Fellowship and was funded by the UC San Diego Economics Department research grant and the Institute for Humane Studies fellowship. This research was approved by UC San Diego IRB under protocol #812520. This study was pre-registered under AEARCTR-0018792 on the AEA RCT Registry. Batmanov: abatmanov@ucsd.edu, Galaty: bgalaty@ucsd.edu. All errors are our own.

1 Introduction

People routinely act on advice from others who know more than they do or who argue more skillfully than they could: investors follow brokers who earn commissions on the products they sell, patients follow specialists who are paid for the procedures they recommend, and voters follow pundits with agendas of their own. Two distinct forces can move a decision. An adviser may hold an informational advantage, knowing something about the world that the decision-maker does not, which gives the decision-maker a reason to listen, at least when their interests are aligned. Alternatively, an adviser may accompany a recommendation with a story or an argument that makes it feel compelling even when it conveys no new facts. The two forces have largely been studied in separate literatures, one on strategic information transmission and Bayesian persuasion, and one on narrative persuasion.[1] In practice, however, the two typically arrive together, because the adviser who knows more is also the one who explains: the broker who understands a product is the same person who supplies the reasons to buy it. Their effects are therefore difficult to separate. Which of the two carries more persuasive weight? And does the story itself add anything beyond the information behind it?

In this paper, we study how an adviser's informational advantage and the ability to provide a narrative each contribute to persuasion, holding the underlying decision problem fixed across message types.[2] We design a laboratory experiment with three key features. First, the decision is an inference task with an objective answer, which lets us measure two things at once: how accurate a Receiver's guess is and how far it is pulled toward the Sender's recommendation. Second, we independently vary, within subject, the Sender's information and the format of the message: the Sender either has the same information as the Receiver or additionally observes the true state, and the message is either a bare recommendation or a recommendation paired with a free-text narrative. Crossing the two dimensions yields four message types. Third, we vary between subjects whether the interests of the two parties are misaligned or aligned: in our main setting, the Sender gains when the Receiver's guess moves in a particular direction, whatever the truth, while in the aligned setting the Sender is paid as the Receiver is. We also vary the difficulty of the inference task itself, so as to change how much room a message has to move the Receiver's decision.

In the experiment, Receivers study a ten-row dataset linking three inputs to an output letter grade and predict the grade of an eleventh case, whose inputs they observe but whose grade is hidden. Before guessing, the Receiver observes a recommendation from a Sender.[3] Under misaligned

[1] On strategic information transmission and Bayesian persuasion, see, among others, Crawford & Sobel (1982); Kamenica & Gentzkow (2011); Milgrom (1981). On narrative persuasion, see Barron & Fries (2025); Charles & Kendall (2025); Eliaz & Spiegler (2020); Schwartzstein & Sunderam (2021).

[2] We refer to the adviser as the Sender and to the decision-maker as the Receiver throughout.

[3] The messages shown to Receivers were composed by Senders in separate sessions. For each dataset, one message

interests, the Sender is paid according to the direction of the Receiver's guess rather than its accuracy: each dataset is labeled UP or DOWN, and the Sender earns more the higher the guess in the first case and the lower the guess in the second, regardless of the truth. A better-informed Sender additionally knows the data-generating process and the correct grade, while an equally informed Sender sees only the same ten rows as the Receiver, and either may send a bare recommendation or accompany it with a narrative. Each Receiver faces all four resulting message types: over twelve rounds, each based on a different dataset, every type appears in three rounds, with the assignment of types to datasets varying across Receivers and the order of rounds randomized. Receivers are paid solely for the accuracy of their guess, as measured by its distance from the true grade,[4] and they know both which type of Sender they face and the schedule by which the Sender is paid, though not whether the current dataset rewards the Sender for a higher or a lower guess. The distance between a guess and the recommendation measures how persuasive a message is, and two bonus rounds without any message, completed after the main task, measure how well Receivers read the data on their own.[5]

Our first result is that, in the aggregate, Receivers' guesses fall in the interior between the truth and the Sender's recommendation: in every one of the twelve datasets, the average guess is statistically different from both the true grade and the recommended one. The two reference points are far apart by construction: each dataset is built so that the true grade lies near one end of the grading scale, and the recommendation shown to Receivers is selected near the opposite end, about 5.29 letter grades away on average, so that the pull of the message can be separated from the pull of the truth.[6] On average, the guess ends up 44 percent of the interval away from the true grade and the remaining 56 percent away from the Sender's recommendation, and the corresponding dataset-level figures range from 18 to 63 percent. On average, then, guesses settle well away from both reference points: Receivers neither report the true grade nor simply adopt the recommended one.[7] We then ask how the four message types differ in how persuasive they are in this setting.

of each type is selected, with narratives largely describing patterns present in the data and the recommended grade held as constant as possible across message types.

[4]A Receiver earns the most for guessing the true grade exactly and $1 less for each letter grade of error, with payment based on one randomly selected round.

[5]The bonus rounds use fresh datasets, follow the main task, and carry a separate incentive, so they provide a suggestive benchmark for unaided accuracy rather than a randomized no-message control.

[6]The true grade is set low in UP-datasets and high in DOWN-datasets, and the misaligned Sender's payment pushes the recommended grade toward the opposite end of the scale. The distance between the truth and the recommendation ranges from 4.25 to 7 letter grades across the twelve datasets.

[7]Both extremes are common at the level of individual guesses: roughly a quarter of guesses coincide exactly with the recommended grade, while some Receivers largely set the message aside. The aggregate pattern is not driven by a small subset of participants.

The central finding of the paper is that a Sender's persuasive power comes primarily from being better informed than the Receiver rather than from the ability to accompany a recommendation with a narrative: messages from better-informed Senders move guesses substantially closer to the recommendation, while a narrative produces effects that are statistically indistinguishable from zero, and bounded well below the information effect, whether on its own or on top of an informational advantage.[8] Adding a narrative alone, moving from Same Info + Simple to Same Info + Narrative, brings guesses only about 9 percent closer to the recommendation, less than half a letter grade and statistically indistinguishable from zero. Granting the Sender an informational advantage instead, moving to More Info + Simple, brings guesses about 39 percent closer, roughly one and a half letter grades, and the effect is highly significant. Combining the two in the More Info + Narrative message brings guesses about 35 percent closer, essentially the same as the informational advantage alone. What moves Receivers is the prospect that the Sender knows something they do not, not whether the recommendation arrives with an explanation attached.

The next two results characterize the scope of this asymmetry: the pull of a better-informed Sender's message appears at every level of task difficulty, and it persists when the interests of the two parties are aligned, the one setting in which a narrative also begins to matter. Whether the grade depends on one input or on three, and whether the relationship between inputs and grade is conditional or not, a message from a Sender who observes the true grade moves guesses significantly toward the recommendation, while the effect of a narrative stays small and insignificant throughout, although the information effect is somewhat larger when the data-generating process is unconditional.[9] Individual-level regressions confirm this finding: the effect of an informational advantage is robust to alternative specifications that condition on the distance between the recommendation and the truth, on dataset fixed effects, and on each Receiver's accuracy in rounds without messages, with the coefficients remaining significant at the one percent level while the narrative coefficient stays small and insignificant. When the preferences of Senders and Receivers are aligned, a better-informed Sender, who now recommends the truth, again moves guesses substantially toward the recommendation, and adding a narrative makes messages marginally more persuasive, the only setting in which it does.

[8] We take as the baseline the Same Info + Simple message, in which a Sender who holds the same information as the Receiver sends a bare recommendation, and ask what changes when a narrative, an informational advantage, or both are added. At baseline, the average guess lies 3.88 letter grades from the recommendation. Every pairwise comparison that grants the Sender an informational advantage is significant at the one percent level, while neither comparison that adds a narrative is, whether the narrative is introduced on its own ($p = 0.12$) or on top of an informational advantage ($p = 0.54$).

[9] Task difficulty is varied in two ways: the number of inputs that affect the grade ranges from one to three, and the data-generating process is either unconditional, with each relevant input affecting the grade directly, or conditional, with the binary input switching which mapping from the inputs to the grade applies.

Receivers' own choices point to the same asymmetry: offered a choice of which type of message to receive, just over half choose the better-informed Sender, about a quarter choose the equally informed Sender who can supply a narrative, and the rest are indifferent. In exploratory splits by this choice, the Receivers least responsive to information are those who chose the narrative Sender. Under misaligned interests, the messages that persuade most are also the most costly: a better-informed Sender's message leaves guesses about 0.71 letter grades, and dollars, further from the truth than an equally informed Sender's, and guesses in the main rounds sit more than a grade further from the truth than in no-message bonus rounds, a suggestive benchmark for unaided accuracy. When interests are aligned, the same informational advantage instead moves guesses toward the truth. The persuasive force of a better-informed Sender is thus present in both settings; whether Receivers gain or lose from responding to it depends on whether the Sender is paid for the accuracy of their guess or for its direction.

This paper provides an experimental study of a question that sits between two literatures, strategic information transmission on one side and narrative persuasion on the other, by measuring both channels within a single design in which an objective truth makes persuasion and its cost observable. Because there is an objective benchmark in our task in the form of the true grade, the design also attaches a payoff to every guess, so the randomized comparisons across message types deliver welfare comparisons alongside behavioral ones, a pairing that is difficult to achieve when the underlying state is unobservable. The central message, that an informational advantage is what moves decision-makers while narratives add little once information is held fixed, speaks most directly to the settings that motivated the paper: relationships in which a better-informed party with a stake in the outcome supplies recommendations to a less-informed one, as brokers do for investors or specialists for patients. To the extent that the findings extend beyond the laboratory, they suggest that the leverage of such a party lies in the belief that she knows more, not in the story she attaches to her recommendation, and that protections for decision-makers should accordingly attend to senders' incentives and informational claims rather than to the format of their messages. We develop these connections in the sections that follow.

Connections to the Literature

This paper primarily relates to three strands of economics literature: the strategic transmission of information, the use of narratives in persuasion, and the formation of mental models from datasets.

First, our paper is related to the large literature on communication between a better-informed Sender and a Receiver. The canonical models comprise cheap talk (Crawford & Sobel, 1982), verifiable disclosure (Milgrom, 1981), and Bayesian persuasion (Kamenica & Gentzkow, 2011),

which differ in whether the Sender's information is verifiable and whether she can commit to a communication strategy. A large experimental literature evaluates the predictions of these models.[10] In a unified framework that nests cheap talk, disclosure, and Bayesian persuasion, Fréchette, Lizzeri & Perego (2022) examine experimentally how commitment and verifiability shape information transmission. This paper relates to this tradition through the Sender's informational advantage, which it studies within an inference task that admits a verifiable ground truth.

Second, the paper relates to the literature on narrative persuasion, in which a Sender influences a Receiver by proposing an interpretation, or model, of commonly observed data rather than by transmitting new information. The theoretical foundations represent narratives either as models that are adopted in proportion to their fit with the observed data (Schwartzstein & Sunderam, 2021) or as causal models that compete for adoption (Eliaz & Spiegler, 2020).[11] An experimental literature examines narrative persuasion directly: Barron & Fries (2025) study the construction and adoption of narratives and the role of narrative fit; Charles & Kendall (2025) examine causal narratives in tasks that require subjects to infer relationships from datasets; and Ambuehl & Thysen (2024) study choice among competing models of the same data.[12] The narrative constitutes the second of the two instruments of persuasion that this paper examines. A recent theoretical literature considers Senders who jointly control both the information a Receiver observes and the model through which it is interpreted (Ba et al., 2026); this paper contributes experimental evidence on the comparative statics of these two instruments within a single setting in which the decision problem is held fixed and the Sender's information and the form of her message are varied independently.

Third, because the Receiver's task is to infer an outcome from a dataset, the paper relates to the literature on the formation of mental models from data. A theoretical literature studies learning under misspecified models, in which agents who interpret data through a flawed representation converge to systematically biased conclusions (Esponda & Pouzo, 2016; Heidhues, Kőszegi & Strack, 2018; Spiegler, 2020).[13] A complementary experimental literature documents systematic errors in

[10] For a survey, see Sobel (2020) and Blume, Lai & Lim (2020). Cheap-talk experiments find that information transmission increases with preference alignment (Dickhaut, McCabe & Mukherji, 1995), while documenting that Senders overcommunicate and Receivers are excessively trusting relative to the theoretical benchmark (Cai & Wang, 2006). Experiments on verifiable disclosure tend to find the opposite tendency, with Senders undercommunicating and Receivers insufficiently skeptical (Jin, Luca & Martin, 2021).

[11] See Spiegler (2020) for a treatment of subjective causal models using the tools of Bayesian networks.

[12] A broader empirical literature documents the influence of narratives and stories on beliefs, memory, and the demand for information (Andre et al., 2026; Bursztyn et al., 2023; Graeber, Roth & Zimmermann, 2024). A related literature on communication format studies the consequences of free-form rather than restricted messages (Charness et al., 2023).

[13] See also Fudenberg, Romanyuk & Strack (2017) and Bohren & Hauser (2021). A recurring theme across this literature is that an incorrect subjective model can persist and continue to distort behavior even in the presence of feedback that is, in principle, sufficient to correct it (Esponda, Vespa & Yuksel, 2024).

extracting relationships from data, including correlation neglect (Enke & Zimmermann, 2019; Eyster & Weizsäcker, 2016) and difficulties in recovering the data-generating process (Enke, 2020; Esponda & Vespa, 2018); Fréchette, Vespa & Yuksel (2024) characterize the models that individuals extract from datasets directly. This paper extends this framework to a persuasion setting, in which a Sender intervenes on the Receiver's interpretation of the data, and the design varies the datasets along dimensions of model complexity, namely the number of payoff-relevant variables and the presence of conditional structure, so as to examine how persuasion relates to the difficulty of the underlying inference.

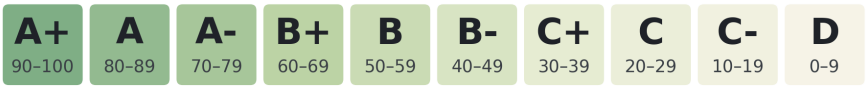
The remainder of the paper is organized as follows. Section 2 describes the experimental design. Section 3 presents the main results and reports additional analyses. Section 4 concludes.

2 Experimental Design

This section describes the experimental design. We construct an inference task in which a Receiver predicts an outcome after observing data and a message from a Sender, and we vary three features of the Sender's position: how well informed the Sender is, whether the Sender can attach a narrative to a recommendation, and whether the Sender's and Receiver's preferences are aligned. We also vary the difficulty of the underlying inference task, so as to change how much room a message has to move Receivers' guesses. We describe the task, the treatment design, the datasets, and the remaining components of the experiment in turn.[14]

2.1 Main Task

Figure 1: Conversion from Score to Letter Grade



We construct a novel inference task in which subjects see a table of data with ten rows. Each row records three inputs (x_1, x_2, x_3) and one outcome (y). The inputs x_1 and x_2 are discrete variables ranging from 1 to 10, while x_3 is binary, coded as "Yes" or "No". The outcome y is a letter grade, which is determined by an unobserved score that each data-generating process (DGP) assigns to the inputs. The conversion from score to grade is shown in Figure 1; scores below 0 or above 100 are assigned the endpoint grades, D and A+.

Subjects are told that the data were generated by some DGP, and that the inputs may be

[14]The experimental instructions are reproduced in Appendix E.

causally related to the outcome or unrelated to it. They are also told that, when the inputs are causal, the relationships may be strictly monotonic or involve conditional dependencies. The Receiver's task is to predict the grade for an eleventh observation: its inputs are shown, its grade is not, and the prediction is made after seeing the ten-row dataset and a suggestion from the Sender.

A Receiver's payoff depends only on accuracy: they are paid more the closer their guess is to the true grade of the eleventh draw, earning the most for a correct guess and \$1 less for each letter grade away, as shown in Figure 2. Receivers therefore want to report the true grade, and a Sender's message is valuable to them only insofar as it helps them do so.

Figure 2: Receiver Payoffs

Your Guess \ True Grade	A+	A	A-	B+	B	B-	C+	C	C-	D
A+	\$14	\$13	\$12	\$11	\$10	\$9	\$8	\$7	\$6	\$5
A	\$13	\$14	\$13	\$12	\$11	\$10	\$9	\$8	\$7	\$6
A-	\$12	\$13	\$14	\$13	\$12	\$11	\$10	\$9	\$8	\$7
B+	\$11	\$12	\$13	\$14	\$13	\$12	\$11	\$10	\$9	\$8
B	\$10	\$11	\$12	\$13	\$14	\$13	\$12	\$11	\$10	\$9
B-	\$9	\$10	\$11	\$12	\$13	\$14	\$13	\$12	\$11	\$10
C+	\$8	\$9	\$10	\$11	\$12	\$13	\$14	\$13	\$12	\$11
C	\$7	\$8	\$9	\$10	\$11	\$12	\$13	\$14	\$13	\$12
C-	\$6	\$7	\$8	\$9	\$10	\$11	\$12	\$13	\$14	\$13
D	\$5	\$6	\$7	\$8	\$9	\$10	\$11	\$12	\$13	\$14

2.2 Treatment Design

We implement the two main treatment variations in a 2×2 design.

1. **Information structure.** Information is either symmetric or asymmetric. Under symmetric information, the Sender and the Receiver see the same ten-row dataset (*Same Information*). Under asymmetric information, the Sender additionally learns the true DGP and the true grade of the eleventh entry (*More Information*).

2. **Narrative elaborateness.** Senders send one of two message formats. A *simple* message is a bare recommendation of a grade; a *narrative* message adds a free-text explanation for the

recommendation.

In addition, we use a between-subject design to vary preference alignment between Senders and Receivers.

Preference structure. Preferences are either misaligned or aligned. Under misaligned preferences, a Sender is paid by the direction of the Receiver's guess rather than its accuracy: each dataset is labeled UP or DOWN, and the Sender earns more the higher the Receiver's guess in UP-datasets and the lower it is in DOWN-datasets (Figure 3). Under aligned preferences, the Sender is paid like the Receiver, earning more the closer the Receiver's guess is to the truth.

Figure 3: Sender Payoffs in UP- and DOWN-datasets (Misaligned)

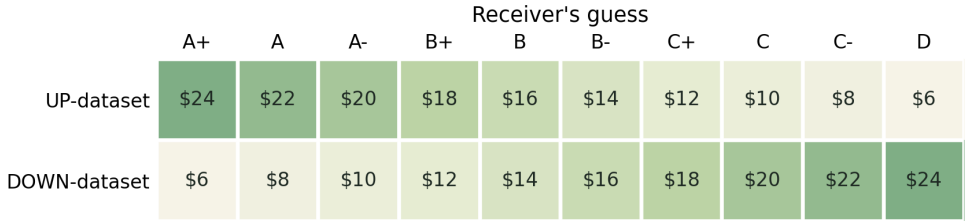


Table 1: The Four Sender Types

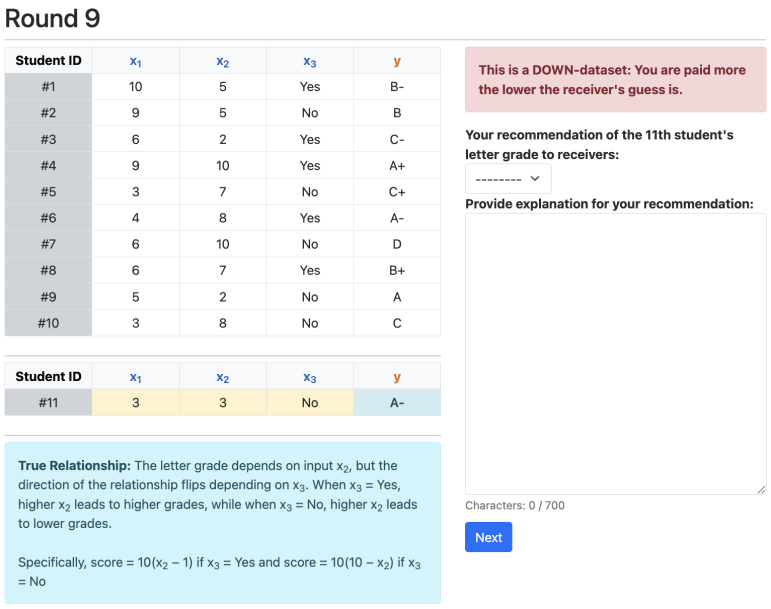
Sender type	Information	Message format
Same Info + Simple	**Same**	Simple
Same Info + Narrative	**Same**	Narrative
More Info + Simple	**More**	Simple
More Info + Narrative	**More**	Narrative

The information structure and the narrative dimension together produce four Sender types, summarized in Table 1 under the labels we use throughout the analysis. These four types are crossed with the two preference conditions; the misaligned condition is our main setting, and the aligned condition serves as a reference point. The narrative and information dimensions vary within subject, so that each Receiver sees every Sender type in equal proportion and, for any given dataset, different Receivers face each type; the preference condition varies between subjects. Sender type is public, so a Receiver always knows which type of Sender they face.[15]

Example screenshots of the interface appear in Figure 4 for the Sender and in Figure 5 for the

[15]The messages a Receiver sees are a subset selected by the researchers from those submitted by Senders, one of each type for each dataset. Messages are chosen so that the recommended grade reflects the Sender's induced preferences and, where a narrative is provided, largely describes patterns present in the data, with the recommended grade held as constant as possible across Sender types for a given dataset. The full selection procedure and the complete list of selected messages appear in Appendix D.

Figure 4: Sender Interface



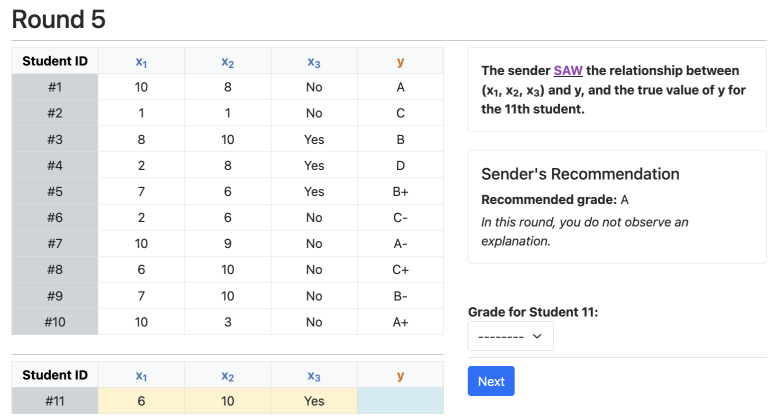
Receiver, which contrasts a simple recommendation with one that includes a narrative.

What the treatments identify. Two features of the design discipline the interpretation of the treatment effects. First, a Receiver learns the Sender's information through a disclosed, truthful label rather than by observing the Sender's information directly, so the information treatment identifies the Receiver's response to a credible claim of superior knowledge, the credibility channel through which an informational advantage operates in practice, rather than the effect of transmitting particular facts. Second, because Receivers face selected messages, the narrative treatment identifies the effect of the selected narrative texts, which are researcher-chosen, largely accurate descriptions of patterns in the data. Appendix A reports balancing checks on the selected messages, showing that recommended grades are broadly balanced across message types and that the results are unchanged when datasets whose narratives reference the true data-generating process are excluded (Table A3).

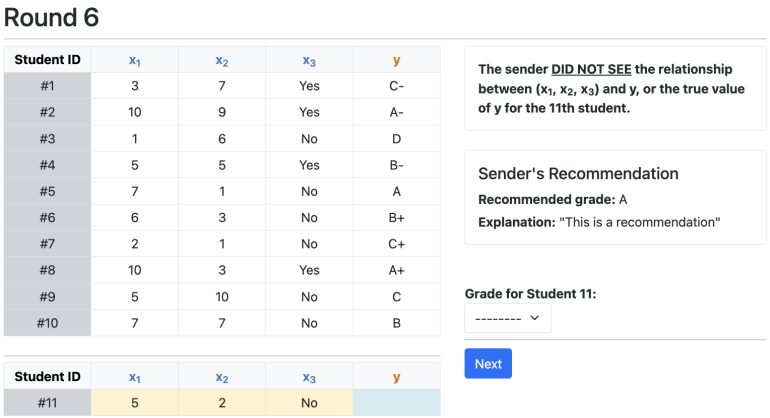
2.3 Data-Generating Processes

To construct the datasets that subjects see, we begin from seven data-generating processes. These vary along two dimensions. The first is how many variables affect the outcome, ranging from one to three. The second is whether the dependency is conditional: in conditional DGPs, the value of x_3 switches which mapping from the inputs to the score applies, so the same inputs can produce very different outcomes depending on x_3, whereas in unconditional DGPs no such switching occurs, with x_3 either absent or entering additively as a level shift.

Figure 5: Receiver Interface



(a) More Info + Simple



(b) Same Info + Narrative

This stratification varies how much room the recommendations have to move Receiver choices. When the true DGP is easy to read off the data, one would expect a recommendation to matter little, since the inference is straightforward on its own; when the DGP is harder to recover, the recommendation may carry more weight. The seven DGPs are listed in Table 2.

There are twelve datasets in the main task: one draw each from DGP 1_s and DGP 1, and two draws from each of DGPs 2 through 6. To generate each draw from DGPs 1 through 6, we first compute the empirical distribution of grades for that DGP across all 200 combinations of the inputs ($10 \times 10 \times 2$), and then draw ten rows whose grades mirror that distribution, which in practice is close to uniform from A+ to D.[16] We then construct an eleventh row so that the dataset can be designated UP or DOWN: in UP-datasets the true grade is near D and in DOWN-datasets it is near A+, leaving the widest scope for a persuasive message. Senders and Receivers see all datasets

[16] DGP 1_s is the exception: it is defined on a restricted support consisting of the two input profiles shown in Table 2, and its dataset draws rows from those profiles only.

Table 2: Data-Generating Processes

	DGP	Conditional?	# of Variables
1_s.	$score = \begin{cases} 75 & \text{if } x_3 = \text{Yes};\ x_1 = x_2 = 2 \\ 25 & \text{if } x_3 = \text{No};\ x_1 = x_2 = 8 \end{cases}$	✓	1
1.	$score = \begin{cases} 12 & \text{if } x_3 = \text{Yes} \\ 88 & \text{if } x_3 = \text{No} \end{cases}$	✓	1
2.	$score = 110 - 2x_1^2$	×	1
3.	$score = \begin{cases} 10(x_2 - 1) & \text{if } x_3 = \text{Yes} \\ 10(10 - x_2) & \text{if } x_3 = \text{No} \end{cases}$	✓	2
4.	$score = 10x_1 - 5x_2 + 20$	×	2
5.	$score = \begin{cases} 5x_1 + 5x_2 - 5 & \text{if } x_3 = \text{Yes} \\ -10x_1 + 5x_2 + 70 & \text{if } x_3 = \text{No} \end{cases}$	✓	3
6.	$score = -x_1 + x_2^2 + 30 - 20 \times \mathbf{1}\{x_3 = \text{Yes}\}$	×	3

in random order, and the full set is reported in Appendix C.

2.4 Additional Rounds

After the twelve main rounds, subjects complete a message-choice round followed by two bonus rounds.

Message Choice

In a thirteenth round with the same structure as the main task, Receivers first choose which type of message they would like to receive before making their guess. They choose between a better-informed Sender who sends only a recommendation (More Info + Simple) and an equally informed Sender who sends a recommendation together with a narrative (Same Info + Narrative). This forces a trade-off between the two channels we study, information and narrative, when a Receiver cannot have both, and reveals which form of advice Receivers value more. The choice is incentivized in the same way as the main task, and the round uses an UP-dataset drawn from DGP 4.

Bonus Rounds

Both Senders and Receivers then complete two bonus rounds in which no message is shown.[17] The purpose is to measure how well subjects perform the inference task on their own, providing a benchmark for unaided performance. Because the bonus rounds use different data-generating processes, follow the main task, and carry a smaller, separate incentive, comparisons between main-round and bonus-round accuracy are suggestive rather than causal; the identification of message effects rests on the randomized comparisons across message types within the main task. Subjects see ten-row datasets drawn from two new processes, DGPs 8 and 9 (Table 3), structurally similar to DGPs 1 and 2, with the true DGP easier to recover from the data in DGP 8 than in DGP 9. Rather than a single guess, subjects predict the grades for the eleventh, twelfth, and thirteenth draws, which lets us measure how well they reconstruct the DGP.

Table 3: Bonus-Round Data-Generating Processes

	DGP	**Conditional?**	**# of Variables**
8.	$score = \begin{cases} 65 & \text{if } x_3 = \text{Yes} \\ 95 & \text{if } x_3 = \text{No} \end{cases}$	✓	1
9.	$score = 105 - 10x_2$	✗	1

2.5 Implementation Details

The experiment was conducted in person at the University of California San Diego Economics Laboratory (EconLab) in Spring 2026, with Sender and Receiver sessions run separately. Prior to the main part of the experiment, all subjects were required to correctly answer a set of comprehension questions to proceed. In total, we collected messages from 70 Senders (43 misaligned, 27 aligned) and guesses from 95 Receivers (75 misaligned, 20 aligned): we first ran five Sender sessions (3 misaligned, 2 aligned) to collect recommendations and narratives, and then six Receiver sessions (5 misaligned, 1 aligned) in which Receivers saw the selected messages.[18] During the instructions, Receivers learned the number of rounds they would complete and that bonus rounds would follow, but they learned the specifics of the message-choice and bonus tasks when they reached that part

[17] This is the sequence in the Receiver-only sessions that constitute the analysis sample. In the joint sessions, which included a small number of Receivers who are not part of the analysis (Section 2.5), the bonus rounds preceded the main task.

[18] Each Sender session included a small number of Receivers whose guesses determined Sender payments at the end of the Sender sessions. These Receivers saw messages that were not subject to our selection procedure and faced a different dataset randomization; they are not included in the analysis, which uses only Receivers from the subsequent Receiver sessions.

of the experiment. On average, Senders earned $17.3 in the experiment; final payment was based on a matched Receiver's guess in one randomly selected main round, paying a misaligned Sender more the further the guess moved in the Sender's induced direction and an aligned Sender more the closer it was to the true grade, together with the earnings from the bonus rounds. On average, Receivers earned $12.8 in the experiment; final payment was based on the outcome of one randomly selected main round, together with the earnings from the bonus rounds.

3 Results

We organize the analysis as follows. Section 3.1 establishes where Receivers' guesses fall in the aggregate, both relative to the true grade, which speaks to how accurately Receivers perform the inference task, and relative to the Sender's recommendation, which speaks to how persuasive the messages are. Section 3.2 then turns to the paper's central question and compares the persuasive power of the four message types, separating the role of a Sender's informational advantage from the role of the narrative. We focus initially on the misaligned setting, in which Senders and Receivers have opposing incentives. The remaining subsections then examine how the pattern varies with task complexity (Section 3.3), how it changes under aligned preferences (Section 3.4), which message type Receivers choose for themselves (Section 3.5), how their behavior extrapolates to the full universe of Sender messages (Section 3.6), and what the results imply for welfare and policy (Section 3.7).

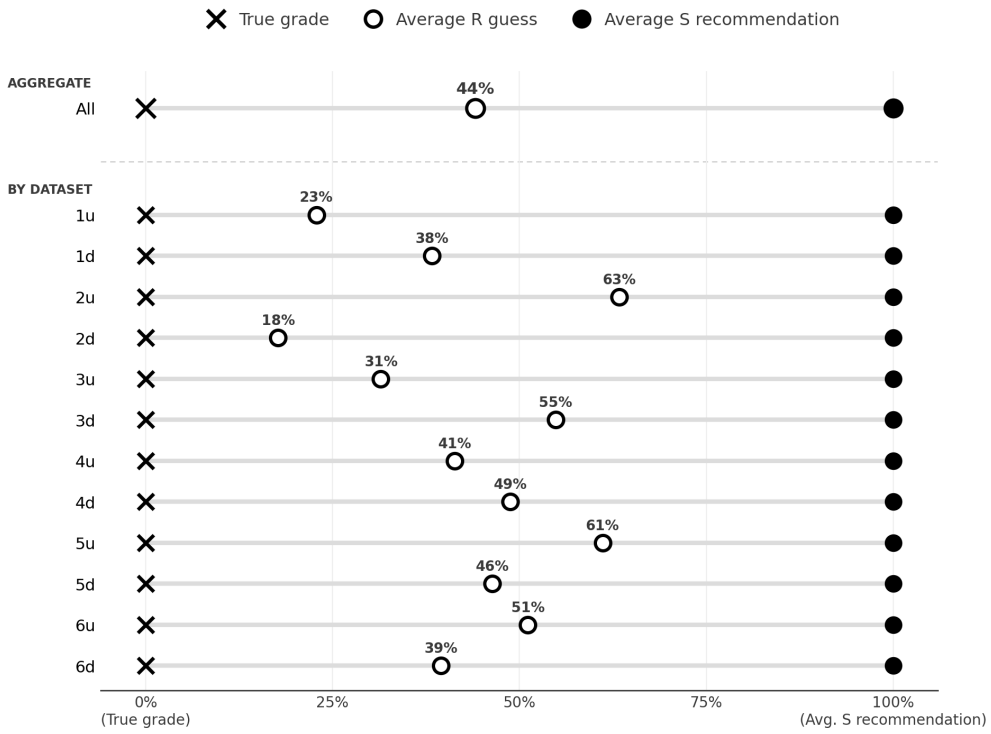
3.1 Receiver Guesses in the Aggregate

By construction, the recommendations Receivers see point well away from the correct answer. Across the selected messages, the Sender's recommended grade lies on average roughly 5.29 letter grades from the true grade, and never fewer than three grades away.[19] This separation is largely a feature of the design: the true grade is set near one end of the grading scale, low in UP-datasets and high in DOWN-datasets, and recommendations are selected near the opposite end, so that the effect of a message can be separated from the effect of the truth. The scope for persuasion is therefore wide, as a Receiver's guess may fall anywhere along the roughly five-grade interval between the true grade and the recommendation. This wide interval is a prerequisite for measuring how far a message moves guesses.

Within this interval, Receivers' guesses remain far from the correct answer. On average a guess lies about 2.74 letter grades from the true grade, although this figure conceals considerable variation across datasets: in some, the average guess sits close to the truth, whereas in others it lies much

[19]The distance between the true grade and the recommendation ranges from averages of 4.25 to 7 letter grades across the twelve datasets.

Figure 6: Average Guess Position Between the True Grade and the Recommendation



Notes: The figure summarizes where Receivers' guesses settle relative to the two reference points of the experiment, the true grade and the Sender's recommendation. In each row, the interval between the true grade (0%, marked by a cross) and the average recommendation (100%, solid circle) is rescaled to percentages, and the hollow circle shows the position of the average Receiver guess within this interval: a value of 0% indicates an average guess equal to the true grade and 100% an average guess equal to the average recommendation. The top row is the mean of the observation-level positions, computed as (guess − truth)/(received recommendation − truth) for each of the $n = 900$ guesses (75 Receivers, each making one guess in each of the 12 datasets); the remaining rows report the position of the dataset-level average guess within each dataset's interval. The two aggregations agree closely: the mean of the twelve dataset-level positions is 43%.

further away.[20] That Receivers fall short of the truth does not simply reflect an inability to read the data. In the bonus rounds, in which Receivers complete comparable inference tasks without any message, their guesses are substantially closer to the true grade, which suggests that the shortfall reflects more than the difficulty of the task alone.[21] Following the Sender's advice thus comes at a cost to Receivers' accuracy and, in turn, their payoffs.

At the same time, guesses are drawn substantially toward the recommendation, and where they

[20] For instance, in dataset 1u the average guess is roughly one letter grade from the true grade, while in datasets 2u and 3d it is nearly four. Appendix A displays the true grade, average guess, and recommendation for each dataset (Figure A1).

[21] Pooling the two bonus rounds, guesses are on average about 1.54 letter grades from the truth. The bonus rounds differ from the main rounds in the underlying data-generating processes, their timing after the main task, and their separate incentive, so this comparison is a suggestive benchmark for unaided performance rather than a causal estimate of the effect of receiving a message; the randomized comparisons across message types below do not rely on it.

settle is summarized in Figure 6. For each dataset, the figure rescales the interval between the true grade (0%) and the average recommendation (100%) and marks the position of the average guess within it. In the aggregate, the average guess sits at 44% of the way from the truth to the recommendation, statistically distinguishable from both endpoints ($p < 0.001$ against each). The same interiority holds dataset by dataset: the average guess ranges from 18% to 63% of the interval, and in every one of the twelve datasets it differs significantly from both the true grade and the recommendation.[22] Interiority is thus a systematic feature of behavior rather than an artifact of averaging across heterogeneous datasets.

Behind these averages lies substantial dispersion in individual guesses. Roughly a quarter of guesses coincide exactly with the recommendation, reflecting Receivers who adopt the Sender's grade outright, while the remaining guesses spread across the full range, with an average distance of about 3.07 letter grades from the recommendation (Figure A3 in Appendix A). Receivers also differ markedly from one another in how closely they follow the recommendation: as Figure A4 shows, the Receiver-level average distance from the recommendation ranges from near zero, for those who adopt it almost exactly, to more than five grades, for those who largely disregard it. The aggregate pattern is therefore not the product of a small subset of participants. On average, guesses settle in the interior of the interval, well away from both the true grade and the recommended one; the extent to which the messages themselves produce this position is the subject of the randomized comparisons in Section 3.2, since a Receiver's unaided reading of a main-round dataset is not observed.

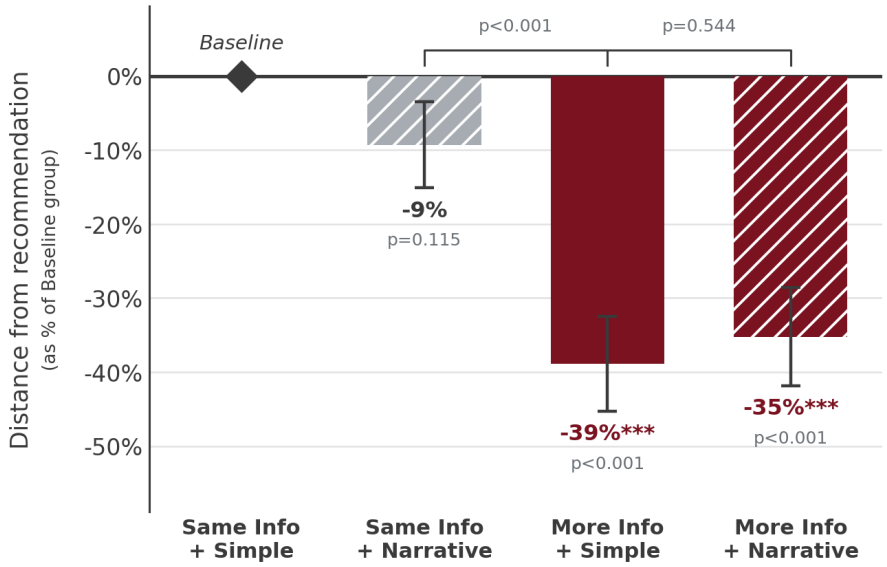
Result #1. *Receivers' guesses fall in the interior between the true grade and the Sender's recommendation. Both in the aggregate and in every dataset, the average guess is statistically different from both the true grade and the recommendation.*

3.2 Persuasion Through Informational Advantage and Narratives

We now turn to the paper's central question: how much of a Sender's persuasive power derives from holding an informational advantage over the Receiver, and how much from being able to accompany a recommendation with a narrative. We continue to measure persuasiveness by the distance, in letter grades, between a Receiver's guess and the grade the Sender recommended, so that smaller distances correspond to more persuasive messages. The four message types form a two-by-two design that crosses the Sender's information, Same or More than the Receiver, with the message format, a simple recommendation or a recommendation accompanied by a narrative. We take Same Info +

[22] One-sample t-tests of the guess against the true grade and against the received recommendation, by dataset ($n = 75$ each). All twenty-four tests reject at the five-percent level, and all but one (dataset 2d against the truth, $p = 0.014$) at the one-percent level.

Figure 7: Distance Between Guess and Recommendation by Message Type



Notes: Each bar reports the average distance between the Receiver's guess and the Sender's recommendation for a given message type, expressed as a percentage difference from the Same Info + Simple baseline, whose mean distance is 3.88 letter grades. The sample comprises 75 Receivers, each making one guess in each of the 12 datasets, for $n = 900$ guesses. The value beneath each bar gives the percentage difference from the baseline, with standard errors and p-values from regressions of the outcome on message-type indicators, clustered at the participant level; error bars denote ± 1 cluster-robust standard error of the difference from baseline, and brackets report the corresponding pairwise comparisons on the same basis. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Simple as the reference type, in which the Sender holds no informational advantage and offers only a bare recommendation, and express the persuasiveness of the other three types relative to it.[23]

The average distance from the recommendation for each message type, expressed as a percentage difference from the Same Info + Simple baseline (whose mean distance is 3.88 letter grades), is reported in Figure 7. The two channels can be read off directly. Adding a narrative while holding information fixed, moving from Same Info + Simple to Same Info + Narrative, reduces the distance by only about 9%, roughly 0.36 letter grades, an effect that is not statistically distinguishable from zero ($p = 0.12$). Granting the Sender an informational advantage has a much larger effect: moving from Same Info + Simple to More Info + Simple reduces the distance by about 39%, roughly 1.51 letter grades, and is highly significant ($p < 0.001$). The combined More Info + Narrative type reduces the distance by about 35%, roughly 1.36 letter grades

[23]This reference type corresponds to the rational benchmark of Appendix B: under misaligned preferences and no informational advantage, cheap-talk reasoning implies that the recommendation transmits nothing, so a rational Receiver would ignore it. Empirically, the design identifies the effects of the message types relative to one another; since no unaided guess is elicited for the same Receiver and dataset, the absolute movement induced by the baseline message itself is not observed.

Table 4: Receiver Guesses by Message Type

	Distance from Recommendation			
	(1)	(2)	(3)	(4)
Same Info + Narrative	−0.360	−0.261	−0.253	−0.254
	(0.226)	(0.216)	(0.208)	(0.208)
More Info + Simple	−1.507***	−1.171***	−1.000***	−1.004***
	(0.247)	(0.245)	(0.257)	(0.257)
More Info + Narrative	−1.364***	−1.049***	−1.002***	−1.005***
	(0.257)	(0.252)	(0.251)	(0.251)
Recommendation Distance from Truth		0.518***	0.845***	0.839***
		(0.064)	(0.107)	(0.106)
Constant	3.876***	0.948**	−1.073	−1.545*
	(0.194)	(0.387)	(0.877)	(0.852)
Equality tests				
Same+Narr = More+Simple	$p < 0.001$	$p < 0.001$	$p = 0.002$	$p = 0.002$
More+Narr = More+Simple	$p = 0.544$	$p = 0.577$	$p = 0.993$	$p = 0.995$
Dataset controls	No	No	Yes	Yes
Bonus controls	No	No	No	Yes
R^2	0.060	0.105	0.183	0.204
Observations	900	900	900	900

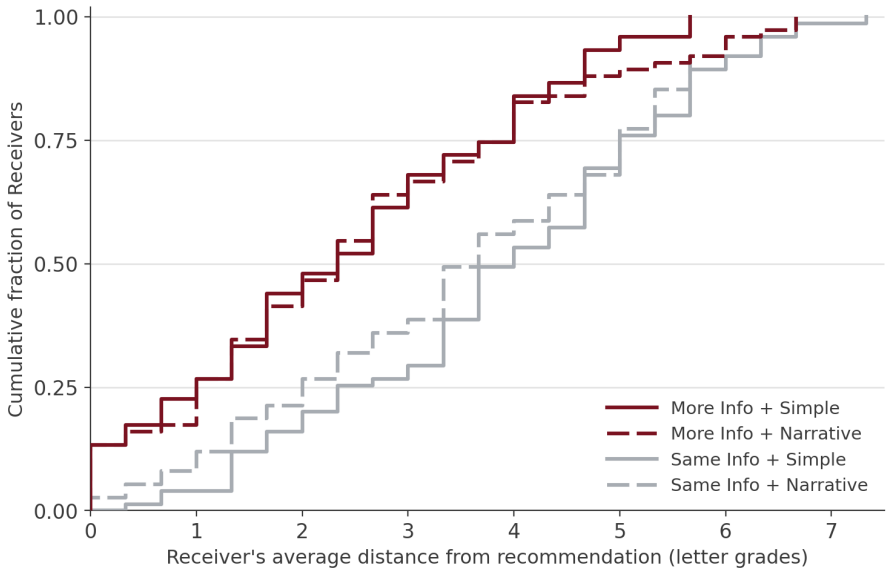
Notes: The dependent variable is the distance, in letter grades, between a Receiver's guess and the Sender's recommendation. Each coefficient gives the average difference in this distance relative to the omitted Same Info + Simple category. *Recommendation Distance from Truth* is the distance between the recommended and true grades. Specification (3) adds dataset fixed effects, and Specification (4) adds controls for each Receiver's accuracy in the no-message bonus rounds. The sample comprises 75 Receivers, each making one guess in each of the 12 datasets, for $n = 900$ guesses. Standard errors, clustered at the participant level, are in parentheses. The equality tests report p-values from tests of equality between the indicated message-type coefficients. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

($p < 0.001$), essentially the same as More Info + Simple.[24] The asymmetry is sharp: every pairwise comparison that adds an informational advantage is highly significant (all $p < 0.001$), whereas neither comparison that adds only a narrative is, whether the narrative is introduced on its own ($p = 0.12$) or on top of an informational advantage ($p = 0.54$).

One might worry that this pattern reflects something other than the message types themselves, for instance that it is driven by a handful of particular datasets or by differences in how well Receivers perform the inference task to begin with. To address these concerns, we estimate individual-level regressions of the distance from the recommendation on the message types, progressively adding controls for these factors, with the results reported in Table 4. Specification (1) reproduces the raw differences from the figure, now expressed in letter grades: the More Information types reduce the

[24] Appendix A presents the same comparison with the differences expressed in letter grades rather than as percentages (Figure A6).

Figure 8: Distribution of Distance from Recommendation



Notes: For each Receiver and each message type, we take the three datasets in which that Receiver saw the message type and average the Receiver's distance from the recommendation across them, yielding one value per Receiver per message type. Each curve then plots the cumulative distribution of these values across the 75 Receivers, so a point on a curve gives the fraction of Receivers whose average distance under that message type falls below the value on the horizontal axis. Color denotes the Sender's information, gray for Same Info and dark red for More Info, and line style denotes the message format, solid for Simple and dashed for Narrative.

distance by 1.51 and 1.36 grades, while Same Info + Narrative reduces it by an insignificant 0.36 grades. Specification (2) adds the distance between the recommendation and the true grade; its positive coefficient indicates that recommendations lying farther from the truth are followed less closely, and it absorbs part of the information effect, which falls to roughly 1 to 1.2 grades but remains highly significant. Specifications (3) and (4) further add dataset fixed effects and controls for each Receiver's accuracy in the no-message bonus rounds.[25] Across all four columns the More Information coefficients stay between 1 and 1.5 letter grades and significant at the one-percent level, while the Same Info + Narrative coefficient remains small and insignificant. The return to an informational advantage is therefore robust: it is not an artifact of which datasets a Receiver happened to face within a given message type, or of how skilled the Receiver was at the task, but reflects a genuine response to the Sender's superior information.

Viewing the two-by-two design factorially leads to the same conclusion: the information main effect is 1.51 letter grades (95% CI $[1.01, 2.00]$), the narrative main effect is 0.36 ($[-0.09, 0.81]$),

[25]The bonus-round controls are indicators for whether each Receiver correctly identified all three true grades (for the eleventh, twelfth, and thirteenth draws) in the two no-message bonus rounds, a proxy for individual inference ability. Standard errors are clustered at the participant level throughout, since each Receiver contributes one observation per dataset.

their interaction is insignificant, and the direct contrast between the two channels is 1.15 grades ($[0.64, 1.65]$, $p < 0.001$; Table A1 in Appendix A). The narrative estimates are precise enough to bound the channel: the confidence interval excludes narrative effects larger than 0.81 grades, roughly half the information effect, and the design would detect a narrative effect of about 0.63 grades with conventional power.[26] These comparisons are also robust to how the recommendations themselves were selected: the selected grades are broadly balanced across message types (Table A3), the contrasts survive when estimated only on datasets in which the compared message types share an identical recommended grade (Table A2), and the factorial estimates are unchanged when the outcome is instead the signed movement toward the recommendation or the normalized position between truth and recommendation (Table A1).

Finally, we ask whether the result holds throughout the population of Receivers rather than only on average (Figure 8). For each Receiver and message type, we compute the average distance from the recommendation across the three datasets in which the Receiver saw that type, and plot the cumulative distribution of these Receiver-level measures for each of the four types.[27] The two More Information distributions lie everywhere to the left of the two Same Information distributions, so an informational advantage shifts the whole distribution of Receiver-level responses toward the recommendation rather than only its mean. Within each information level, the Simple and Narrative curves nearly coincide, reinforcing that the narrative contributes little. Because marginal distributions alone cannot reveal how many individual Receivers respond, we also compare each Receiver's own averages across conditions: 80 percent of Receivers sit closer to the recommendation under More Information messages than under Same Information messages (95% CI $[71\%, 89\%]$), with a mean within-Receiver difference of 1.26 letter grades. The information effect is thus widespread across Receivers rather than concentrated among a few.

Across the results presented above, a clear asymmetry emerges in what makes a Sender persuasive. What moves Receivers is the prospect that the Sender knows something they do not: an informational advantage pulls guesses sharply toward the recommendation, does so throughout the distribution of Receivers, and survives a battery of controls. The ability to dress a recommendation in a narrative, by contrast, produces effects that are statistically indistinguishable from zero and bounded well below the information effect, both on its own and on top of an informational advantage. It is worth emphasizing how this compares to the rational benchmark of Appendix B: with misaligned interests

[26] The minimum detectable effect at 80 percent power and a five-percent two-sided test is 2.8 standard errors, or 0.63 letter grades given the clustered standard error of 0.226. We therefore read the data as bounding the narrative channel well below the information channel rather than as establishing that narratives have exactly no effect.

[27] Because the design is within-subject, all four curves are computed over the same set of Receivers, so differences between them cannot be attributed to compositional differences across groups.

and no commitment, a rational Receiver would ignore the recommendation regardless of the Sender's information, so the strong response to a disclosed informational advantage is itself a deviation from equilibrium reasoning, in line with the receiver credulity documented in cheap-talk experiments. In this setting, persuasion operates primarily through the Receiver's belief that the Sender is better informed, rather than through interpretation or framing.

Result #2. *Messages from Senders with an informational advantage move Receivers' guesses substantially closer to the recommendation than messages from equally informed Senders. Accompanying a recommendation with a narrative produces only a small and statistically insignificant difference, both in the presence and in the absence of an informational advantage.*

3.3 Task Complexity

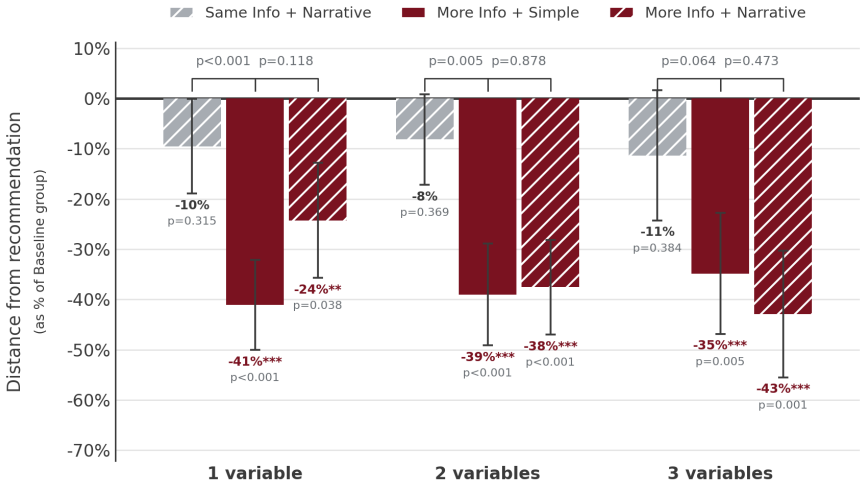
We next look at how the two channels vary with the difficulty of the inference task. Task difficulty is captured by two features of the data-generating process: the number of variables a Receiver must weigh, and whether its structure is conditional. The persuasiveness of each message type across the three numbers of variables, with four datasets at each level, is reported in Figure 9a. It is worth recalling how to read these figures: each bar is the distance from the recommendation relative to the Same Info + Simple baseline, so a larger reduction (a more negative bar) means the guess sits closer to the recommendation, that is, the Receiver is more persuaded.

The baseline distance itself shrinks as the task involves more variables, from 4.46 letter grades with one variable to 3.79 with two and 3.34 with three, a descriptive pattern consistent with Receivers leaning a little more on the recommendation when the task is harder. The comparison across message types is stable across these levels: an informational advantage produces a large and significant reduction, between 35 and 41%, while a narrative alone stays small and insignificant throughout, around 8 to 11%. Interaction regressions confirm this stability: the treatment-by-variables interactions are jointly insignificant ($p = 0.74$), so we cannot reject that the effects are the same at every number of variables.[28]

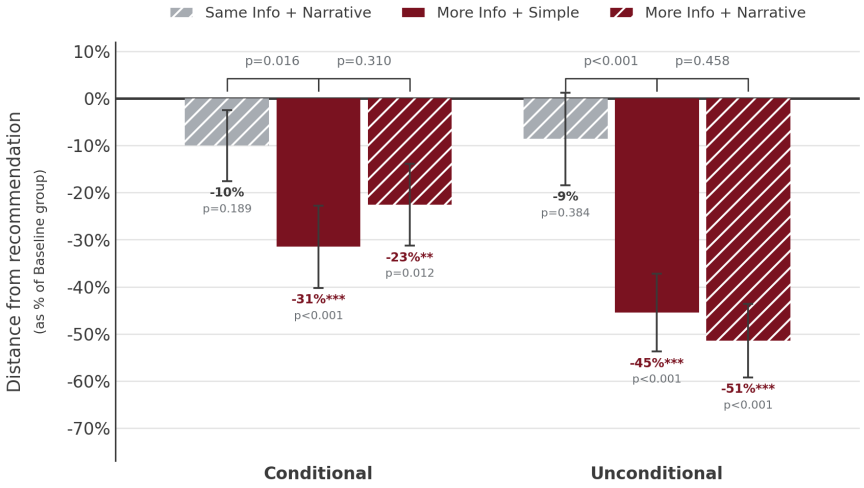
The same comparison split by conditional structure, with six datasets in each, appears in Figure 9b. The two baselines are nearly identical (3.88 versus 3.87 letter grades), so the effects are directly comparable. A narrative alone again moves guesses only slightly, by roughly 9 to 10% in

[28]The interaction tests regress the distance from the recommendation on the message-type indicators, indicators for the complexity level, and their interactions, with standard errors clustered at the participant level; the reported p-values are joint Wald tests of the interaction terms. Because the complexity levels are implemented through distinct data-generating processes, these splits compare structural classes of datasets rather than a single manipulated difficulty parameter.

Figure 9: Distance Between Guess and Recommendation by Task Complexity



(a) By number of variables

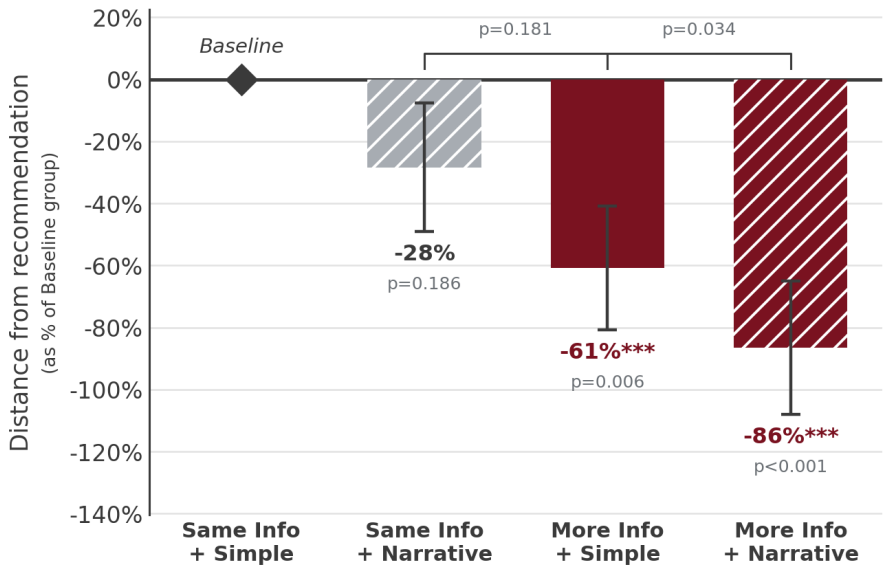


(b) By conditional structure

Notes: Each bar reports the average distance from the recommendation by message type, expressed as a percentage difference from the Same Info + Simple baseline. Panel (a) splits the datasets by the number of variables in the underlying data-generating process (four datasets per level; baseline mean distance 4.46, 3.79, and 3.34 letter grades for one, two, and three variables): in one-variable datasets the grade depends on a single input, while in two- and three-variable datasets it depends on two or all three of the inputs. Panel (b) splits the datasets by conditional structure (six datasets each; baseline mean distance 3.88 letter grades in the conditional and 3.87 in the unconditional datasets): in conditional datasets the value of x_3 switches which mapping from the inputs to the grade applies, whereas in unconditional datasets no such switching occurs. A larger reduction indicates a guess closer to the recommendation. Standard errors and p-values come from participant-clustered regressions; error bars denote ± 1 cluster-robust standard error, and brackets report pairwise comparisons within a group on the same basis. Misaligned condition; 75 Receivers each making one guess per dataset. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

both cases. The information effect, by contrast, is stronger when the data-generating process is unconditional: it narrows the gap by about 45% when the structure is unconditional but by about

Figure 10: Distance Between Guess and Recommendation: Aligned Preferences



Notes: The aligned-preferences analogue of Figure 7. Each bar reports the average distance from the recommendation by message type, expressed as a percentage difference from the Same Info + Simple baseline, whose mean distance is 1.23 letter grades. A larger reduction indicates a guess closer to the recommendation. Standard errors and p-values come from participant-clustered regressions; error bars denote ± 1 cluster-robust standard error, and brackets report pairwise comparisons on the same basis. The sample comprises 20 Receivers, each making one guess in each of the 12 datasets, for $n = 240$ guesses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

31% when it is conditional, and here the interaction tests do reject equality ($p = 0.008$ jointly), with the difference concentrated in the More Info + Narrative cell. Receivers thus appear to draw on a Sender's information most when the conditional structure is not present.

The contrast between information and narrative thus holds across the board, even as its size shifts with the task.

Result #3. *The contrast between information and narratives appears at every level of task complexity: an informational advantage significantly moves guesses toward the recommendation whether the task is simple or complex, conditional or not, while the narrative effect stays small and insignificant. The information effect does not vary significantly with the number of variables, but is significantly larger when the data-generating process is unconditional.*

3.4 Aligned Preferences

We turn next to the aligned condition, in which Senders and Receivers share the goal of an accurate guess and a better-informed Sender recommends the true grade. This serves less as a separate treatment than as a reference point, a sanity check on how Receivers would behave were preference misalignment removed, since here following the message is in the Receiver's interest.

Table 5: Receiver Guesses by Message Type: Aligned Preferences

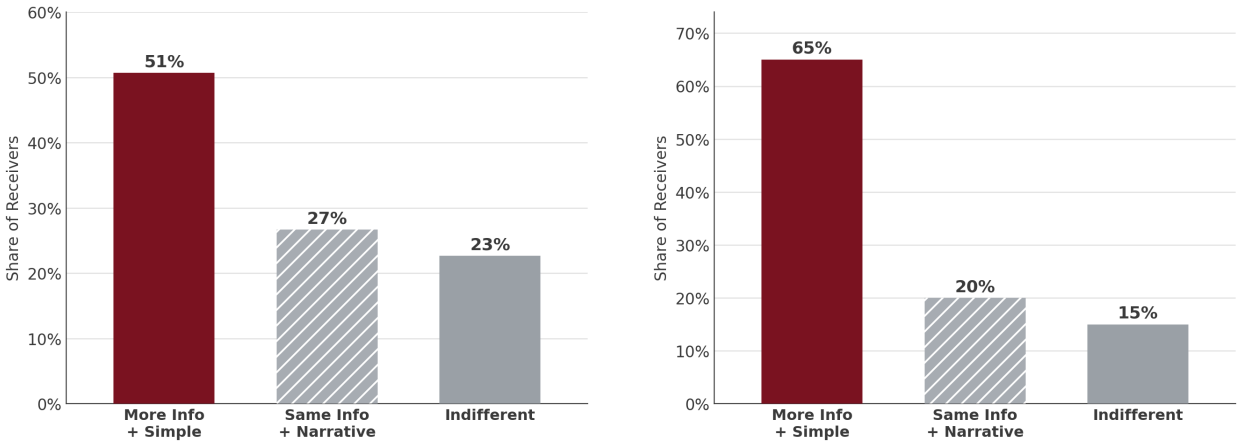
	Distance from Recommendation			
	(1)	(2)	(3)	(4)
Same Info + Narrative	−0.350	−0.541*	−0.511*	−0.492*
	(0.256)	(0.261)	(0.245)	(0.239)
More Info + Simple	−0.750***	−0.593**	−0.627**	−0.658**
	(0.246)	(0.234)	(0.227)	(0.244)
More Info + Narrative	−1.067***	−0.909***	−0.968***	−0.991***
	(0.265)	(0.268)	(0.281)	(0.290)
Recommendation Distance from Truth		0.337*	0.275	0.229
		(0.169)	(0.184)	(0.177)
Constant	1.233***	1.076***	0.792**	1.259**
	(0.297)	(0.308)	(0.359)	(0.497)
Equality tests				
Same+Narr = More+Simple	$p = 0.182$	$p = 0.830$	$p = 0.643$	$p = 0.530$
More+Narr = More+Simple	$p = 0.034$	$p = 0.034$	$p = 0.053$	$p = 0.060$
Dataset controls	No	No	Yes	Yes
Bonus controls	No	No	No	Yes
R^2	0.083	0.109	0.133	0.193
Observations	240	240	240	240

Notes: The aligned-preferences analogue of Table 4. The dependent variable is the distance, in letter grades, between a Receiver's guess and the Sender's recommendation, and each coefficient gives the average difference relative to the omitted Same Info + Simple category. *Recommendation Distance from Truth* is the distance between the recommended and true grades; it is zero for every More Information message, since a better-informed Sender always recommends the true grade in the aligned condition. The sample comprises 20 Receivers, each making one guess in each of the 12 datasets, for $n = 240$ guesses. Standard errors, clustered at the participant level, are in parentheses. The equality tests report p-values from tests of equality between the indicated message-type coefficients. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

The persuasiveness of each message type in this condition, measured as the distance from the recommendation and presented on the same scale as the misaligned case for comparability, appears in Figure 10.

The baseline distance is much smaller than under misaligned preferences, 1.23 letter grades, reflecting that the recommendation and the truth are closer together. The ranking of the channels is nonetheless familiar. A narrative alone reduces the distance by about 28%, larger than the 9% seen under misalignment but still not significant ($p = 0.19$). An informational advantage reduces it by about 61% ($p < 0.01$), and combining information with a narrative reduces it by about 86% ($p < 0.001$); the difference between Same Info + Narrative and More Info + Simple is not significant ($p = 0.18$), while adding a narrative on top of the informational advantage is ($p = 0.03$). With only 20 Receivers, this condition is estimated much less precisely than the misaligned one:

Figure 11: Chosen Message Type



(a) Misaligned preferences

(b) Aligned preferences

Notes: Distribution of the Sender type chosen in the final message-choice task, for the 75 Receivers in the misaligned condition (Panel (a)) and the 20 Receivers in the aligned condition (Panel (b)). Receivers chose between a More Info + Simple Sender, a Same Info + Narrative Sender, or indifference between the two.

the narrative effect of -0.35 grades carries a confidence interval of $[-0.89, 0.19]$, so effects up to about 70 percent of the baseline distance cannot be ruled out, and the aligned estimates should be read with corresponding caution. Even with aligned incentives, Receivers do not fully adopt the recommendation, as a gap to the recommended grade remains under every message type. That gap is smallest under More Information messages, where it nearly closes: a better-informed Sender now recommends the true grade, so a Receiver who follows closely is also guessing accurately, which is the behavior the aligned benchmark should elicit.

The regression estimates in Table 5 confirm this pattern and add a nuance. Adding controls leaves the More Information coefficients large and significant throughout, while the Same Info + Narrative coefficient hovers around weak significance across specifications. The cleaner evidence that narratives carry more weight under alignment comes from the comparison in which the recommendation is truthful by construction: adding a narrative on top of an informational advantage, moving from More Info + Simple to More Info + Narrative, closes a further 0.32 grades of the remaining gap ($p = 0.03$; the equality tests in the table). Under alignment, then, a narrative adds persuasive force when it accompanies a truthful recommendation, the one configuration in which it also serves the Receiver.

Result #4. *When the preferences of Senders and Receivers are aligned, messages from a better-informed Sender move Receivers' guesses substantially toward the recommendation, as under misalignment. In contrast to the misaligned setting, however, the addition of a narrative to a*

recommendation makes messages marginally more persuasive.

3.5 Message Type Choice

After the main task, Receivers chose which type of message they would prefer to receive. Their choices, shown in Figure 11a, favor information: just over half (51%) chose a Sender with more information, while the rest split between a Sender who was equally informed but could provide a narrative (27%) and indifference between the two (23%). This parallels how Receivers behaved in the main task, where they responded more to information than to a narrative. The preference is notable because, under misaligned preferences, a better-informed Sender is also the one whose advice moves guesses farthest from the truth.[29]

Panel A of Table 6 presents regression results disaggregated by the message type that subjects selected in round 13, paralleling the uncontrolled regression in column (1) of Table 4. Because the choice is made after the main task and splits the sample into small groups, we read these splits as exploratory heterogeneity rather than as tests of pre-specified hypotheses. Two patterns stand out. Receivers who chose the better-informed Sender respond strongly to information (point estimates of -1.71 and -1.34 grades), and so do the Receivers who declared themselves indifferent (-1.84 and -1.80); the group whose response to information is weak and insignificant is instead the one that chose the Same Info + Narrative Sender (-0.83 and -1.03). The choice thus separates Receivers by their responsiveness to information, though not in a way that concentrates the response solely among those who chose the better-informed Sender.

The choices of the aligned Receivers, shown in Figure 11b, favor information even more strongly: 65% select the better-informed Sender with a simple message, while the remainder split roughly evenly between the equally informed Sender with a narrative (20%) and indifference (15%). This pattern is consistent with Receivers recognizing that, under aligned incentives, a better-informed Sender's recommendation is also the more accurate one.

Panel B of Table 6 repeats the split for the aligned condition, where the cells are smaller still (13, 4, and 3 Receivers) and the estimates correspondingly noisy. Among Receivers who chose the better-informed Sender, a More Info + Simple message closes 88% of the baseline gap. In the two remaining groups the point estimates rank the Same Info + Narrative message ahead of More Info + Simple, closing 50% versus 20% of the gap among narrative choosers and 67% versus 33% among the indifferent, but the within-group equality tests come nowhere near significance ($p = 0.56$ and $p = 0.64$), so these reversals in ordering should not be given weight.

[29] As the welfare analysis in Section 3.7 shows, this translates into lower payoffs: following a better-informed Sender leaves Receivers, on average, further from the true grade.

Table 6: Receiver Guesses Split by Choice of Preferred Message Type

	Distance from Recommendation		
	(1) Chose More + Simple	(2) Chose Same + Narrative	(3) Chose Indifferent
Panel A: Misaligned Preferences			
Same Info + Narrative	−0.553*	−0.200	−0.118
	(0.320)	(0.461)	(0.458)
More Info + Simple	−1.711***	−0.833	−1.843***
	(0.355)	(0.485)	(0.479)
More Info + Narrative	−1.342***	−1.033*	−1.804***
	(0.416)	(0.496)	(0.305)
Constant	3.842***	3.817***	4.020***
	(0.298)	(0.369)	(0.349)
Equality tests			
Same+Narr = More+Simple	$p < 0.001$	$p = 0.234$	$p = 0.015$
More+Narr = More+Simple	$p = 0.228$	$p = 0.632$	$p = 0.951$
R^2	0.065	0.027	0.115
Observations	456	240	204
Panel B: Aligned Preferences			
Same Info + Narrative	−0.0769	−0.833	−0.889
	(0.232)	(0.999)	(0.647)
More Info + Simple	−0.949**	−0.333	−0.444
	(0.319)	(0.545)	(0.647)
More Info + Narrative	−1.051**	−1.083	−1.111
	(0.353)	(0.694)	(0.581)
Constant	1.077***	1.667	1.333
	(0.350)	(0.964)	(0.726)
Equality tests			
Same+Narr = More+Simple	$p = 0.011$	$p = 0.559$	$p = 0.640$
More+Narr = More+Simple	$p = 0.107$	$p = 0.196$	$p = 0.440$
R^2	0.143	0.063	0.098
Observations	156	48	36

Notes: Each column restricts the sample to Receivers who, in the message choice task, chose the recommendation type indicated in the column heading. Because the choice is made after the main task and the resulting cells are small, these splits are exploratory, post-treatment heterogeneity rather than randomized comparisons. Panel A restricts the sample to the Misaligned Preferences condition; Panel B restricts it to the Aligned Preferences condition. Robust standard errors, clustered at the participant level, are in parentheses. The equality tests report p-values from tests of equality between the indicated message-type coefficients. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table 7: Receiver Adjustment to the Full Set of Sender Messages

	Rec. Distance from Truth (1)	Share Correctly Adjusted (2)	Implied Distance from Truth (3)
Panel A. Misaligned			
Same Info + Simple	4.477	86.6%	0.602
Same Info + Narrative	4.492	78.3%	0.977
More Info + Simple	3.841	61.7%	1.472
More Info + Narrative	3.939	63.7%	1.428
Panel B. Aligned			
Same Info + Simple	2.049	60.2%	0.816
Same Info + Narrative	1.840	48.0%	0.956
More Info + Simple	0.272	178.0%	0.212
More Info + Narrative	0.275	60.6%	0.108

Notes: The table extrapolates Receiver behavior to the full universe of Sender messages, rather than the researcher-selected subset Receivers actually faced. *Rec. Distance from Truth* is the average distance, in letter grades, between the recommendation and the true grade across all Sender messages of a given type. *Share Correctly Adjusted* applies the discount each Receiver placed on the selected recommendations to this full set, that is, the fraction of the message-to-truth gap a Receiver would correct if responding to the average message. *Implied Distance from Truth* is the resulting average distance between the Receiver's guess and the truth. A share above 100% (More Info + Simple, aligned) indicates overcorrection past the truth. Computed from the full set of Sender messages; see the text for details.

Result #5. *When choosing which type of message to receive, a majority of Receivers prefer a message from a better-informed Sender over one with a narrative, paralleling their stronger response to information in the main task. In exploratory splits by this choice, the response to information is weakest among the Receivers who chose the narrative Sender, while both other groups respond strongly.*

3.6 Universe of Sender Messages

Throughout the experiment, Receivers responded to a researcher-selected subset of Sender messages rather than the full set a Sender might send. Because we deliberately chose messages that point away from the truth, it is worth asking how Receivers would have fared against the full universe of Sender messages, and whether their behavior amounts to best responding to the bias in that universe.[30] Across all Sender messages, the average recommendation lies 4.19 letter grades from the truth in the misaligned condition and 1.11 in the aligned condition. The selection therefore went in opposite directions across conditions: under misaligned preferences the messages we chose

[30]By the full universe we mean every recommendation that Senders actually composed for each dataset, before our selection of a subset for the Receiver task.

Table 8: Distance Between Guess and Truth by Message Type

	Distance from Truth	
	Misaligned	Aligned
	(1)	(2)
Same Info + Narrative	0.382**	−0.183
	(0.180)	(0.228)
More Info + Simple	0.707***	−0.950***
	(0.207)	(0.239)
More Info + Narrative	0.489**	−1.267***
	(0.199)	(0.245)
Constant	2.347***	1.433***
	(0.160)	(0.274)
Equality tests		
Same+Narr = More+Simple	$p = 0.143$	$p = 0.006$
More+Narr = More+Simple	$p = 0.299$	$p = 0.034$
R^2	0.014	0.143
Observations	900	240

Notes: The dependent variable is the distance, in letter grades, between a Receiver's guess and the true grade, which equals the Receiver's dollar loss under the experiment's incentive scheme (a Receiver loses \$1 for each letter grade by which the guess departs from the truth). Each coefficient gives the average difference relative to the omitted Same Info + Simple category; the constant is the baseline distance. Column (1) covers the misaligned condition (75 Receivers, $n = 900$) and Column (2) the aligned condition (20 Receivers, $n = 240$), each Receiver making one guess per dataset across 12 datasets. Standard errors, clustered at the participant level, are in parentheses. The equality tests report p-values from tests of equality between the indicated message-type coefficients. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

were more extreme than the typical message (5.29 versus 4.19 grades from the truth), while under aligned preferences they were closer to the truth than typical (0.38 versus 1.11 grades).

As a descriptive, back-of-envelope extrapolation, we ask how far Receivers' guesses would fall from the truth if they applied to the full set of messages the same discount they applied to the selected ones.[31] The resulting distances appear in Table 7. In the misaligned condition, Receivers correct for between 62 and 87% of the gap between the message and the truth, leaving their guesses between 0.6 and 1.47 letter grades away from the truth. Because the full universe is less biased

[31]This invariant-discount counterfactual treats Receivers as if they had faced a representative draw of Sender messages and had discounted each message type by the same distances observed in the experiment. It is a simulation rather than a measurement, for two reasons. First, Receivers saw only the selected subset, and the regression estimates in Table 4 show that responses vary with how far a recommendation lies from the truth, so the discount applied to our deliberately extreme messages need not carry over to more moderate ones. Second, the exercise does not identify optimal behavior: under the experiment's payoff, an optimal guess is a conditional median of the truth given everything the Receiver observes, which an average adjustment rate does not recover.

than the messages we selected, these implied losses are smaller than the ones Receivers actually incurred, yet under this counterfactual they remain well short of the truth.

Under aligned preferences the adjustments range between 48 and 61%, with one exception. For More Info + Simple messages Receivers overcorrect, adjusting by 178% and overshooting the truth.[32] Overall, the simulation suggests that Receivers' guesses would land closer to the truth against the full set of messages than against the ones we selected, but that, under the invariant-discount assumption, they would still fall short of fully accounting for how far recommendations stray from the truth. We emphasize that these are illustrative magnitudes; the welfare statements we rely on below come from the randomized comparisons across message types.

3.7 Welfare and Policy Implications

The experiment's incentive scheme ties payoffs directly to accuracy: a Receiver loses \$1 for every letter grade by which the guess departs from the truth, so the distance-from-truth estimates in Table 8 translate one-for-one into dollar losses. Because message types are randomized within Receiver, the comparisons across message types in that table are the experiment's cleanly identified welfare estimates, and we base the welfare conclusions on them.

In the misaligned condition, the more persuasive messages are also the more costly ones. A Receiver who sees the baseline Same Info + Simple message guesses on average 2.35 letter grades from the truth, a \$2.35 loss. Adding a narrative widens this by a further \$0.38, and granting the Sender an informational advantage widens it most, by \$0.71, so that guesses under More Info + Simple sit roughly 3.05 grades from the truth. The randomized contrast thus attaches a direct price to the Sender's informational advantage: under misaligned interests, facing a better-informed Sender costs a Receiver about \$0.71 per round relative to facing an equally informed one. For a rough sense of overall magnitudes, guesses in the main rounds average about 2.74 grades from the truth, against about 1.54 grades in the no-message bonus rounds; the bonus rounds differ from the main rounds in their datasets, timing, and stakes, so this gap is a suggestive benchmark for unaided performance rather than a causal estimate of the effect of receiving advice.

In the aligned condition the same forces work in Receivers' favor. The baseline loss is smaller, about \$1.43, and better messages close most of the gap to the truth: a narrative alone recovers \$0.18, an informational advantage recovers \$0.95 (a residual loss of \$0.48), and the two together recover \$1.27, leaving only \$0.17 on the table. When incentives align, Senders help Receivers most when they are both better informed and able to explain their recommendation.

[32] Overcorrection arises because better-informed Senders in the aligned condition recommend grades very close to the truth, so applying the discount Receivers used elsewhere pushes their guesses past it.

Pulling the results together, the experiment yields a consistent picture of how persuasion works when one party advises another. Receivers do not follow recommendations fully, but they are reliably pulled toward them, settling in the interior between the truth and the Sender's advice, away from both (Result #1). What moves them is information rather than the way it is framed: granting a Sender an informational advantage shifts guesses sharply toward the recommendation, whereas a narrative does little on its own (Result #2). This contrast appears across levels of task difficulty (Result #3), and its welfare sign depends on whether incentives are aligned: when the parties' interests conflict, the messages that persuade most are the ones that cost Receivers most, whereas when interests align the same messages recover most of the gap to the truth (Result #4). And given the choice, Receivers prefer messages from better-informed Senders, even though those are the Senders who, under conflict, move them furthest from the truth (Result #5).

These findings speak to the many settings in which advice comes from a better-informed but not disinterested party, such as financial advisers, salespeople, medical referrals, expert testimony, and political messaging. In our setting, what moves Receivers is the belief that the adviser knows more, not the way the recommendation is explained. A cautionary implication follows for transparency-based protections: our Receivers knew each Sender's information and payment rule throughout, yet publicly identified, better-informed misaligned Senders were precisely the ones who imposed the largest losses, so disclosure of information and incentives alone did not protect them. Regulating the framing or storytelling around a recommendation looks similarly unpromising here, since the narrative channel is weak to begin with. The demand side matters too: because Receivers tend to choose better-informed advisers even when those advisers' interests oppose their own, simply widening access to expert advice need not improve decisions when interests diverge.

At the same time, the aligned-preferences results caution against a blanket suspicion of informed advice. When the adviser's interests coincide with the advisee's, the same informational advantage that is costly under conflict becomes the main source of benefit, letting Receivers recover most of what they would otherwise lose. The relevant variable for policy is thus not information or persuasion as such, but the alignment of incentives between the parties, the one lever our design varies directly. Institutions that align advisers' incentives with those they serve are likely to do more to protect decision-makers than measures aimed at the form advice takes; whether additional transparency helps is a question our design, in which incentives were always disclosed, cannot answer.

4 Discussion

We study an environment in which a decision-maker receives a recommendation from a party whose interests may differ from her own. Our design separates two channels that could determine whether and how far the recommendation is followed: the Sender may know something the Receiver does not, and the Sender may accompany the recommendation with an explanation. We find that Receivers' willingness to follow the message is driven almost entirely by the first channel: an informational advantage pulls guesses sharply toward the recommendation, while a narrative on its own does not. In our main, misaligned setting, Senders' recommendations point well away from the truth by construction, so the messages that persuade most are also the most costly: a better-informed Sender's message leaves guesses about 0.71 letter grades, and dollars, further from the truth than an equally informed Sender's. This pattern holds across levels of task difficulty and across the features of the underlying datasets, which suggests that it is not an artifact of any one setting but a more basic feature of how Receivers weigh the messages they face. And although following these Senders lowers their earnings, a majority of Receivers choose to receive their messages when given the option.

The aligned condition offers a useful counterpoint. When the interests of Senders and Receivers coincide, the same tendency to follow a better-informed Sender that is costly under misalignment instead works in Receivers' favor, allowing them to recover most of the accuracy they forgo when matched with equally informed Senders. Narratives, while still not the dominant channel, carry more weight than under misalignment: adding a narrative now makes recommendations marginally more persuasive. Receivers' choices reflect the same logic: an even larger share selects a better-informed Sender than under misalignment, so Receivers lean on information most in the setting where doing so serves them.

Across both alignment conditions, narratives play at most a marginal role, which contrasts with an experimental literature in which narratives are highly persuasive. We read this contrast as reflecting features of our environment rather than as a challenge to that literature. One likely reason is the design: we pit a narrative against a salient and credible informational advantage, in a task with an objective benchmark and with narratives selected to describe patterns present in the data. A second is that the alignment of Sender and Receiver interests is known to the Receiver throughout, whereas much of the narrative-persuasion literature leaves the intent behind the narrative more ambiguous, and a narrative may carry more weight when a Receiver cannot be sure whose interests it serves. The balance seems more likely to tilt toward the narrative in environments in which the data admit more than one reasonable interpretation, no Sender holds a genuine informational

advantage, the alignment of interests is unclear, or a narrative can draw on beliefs the Receiver already holds. Mapping out where the balance shifts is a natural extension of this design.

Several questions raised by our results lie outside what our design can address and are natural directions for future work. Whether Receivers' propensity to follow better-informed Senders would persist over time is one such question. Receivers never learn the true grade after guessing, so they receive no direct feedback on how costly following a given Sender's messages was, and with only a handful of exposures to each Sender type, they have limited opportunity to infer this from the pattern of recommendations alone. Whether feedback on realized accuracy would teach Receivers to discount the messages of better-informed but misaligned Senders, and how quickly, bears directly on whether markets for expertise discipline self-interested experts.

A second question concerns strategic message construction. We show each Receiver a single, researcher-selected message per dataset, which holds fixed the strategic crafting of narratives, competition among rival narratives, and any role for skill in determining who gets listened to, all of which plausibly matter more for a broker or a pundit than for the Senders in our task. Endogenizing both sides of the interaction, letting Senders choose how informed to appear and how to frame what they know, and letting Receivers choose whom to consult, would connect these results more directly to how relationships between experts and decision-makers form. In this respect, the fact that Receivers actively seek out better-informed Senders even when interests are misaligned suggests that the demand for information is not tempered by the recognition that it can be used against them. Finally, repeating this design in higher-stakes domains, such as financial decision-making, medical referrals, or political communication, would clarify whether the pattern documented here, in which persuasion operates through the Sender's information rather than through the narrative, is a general property of interactions between better-informed and less-informed parties or a feature of settings, like ours, with an objectively defined answer, albeit one that Receivers themselves never observe.

References

Ambuehl, S., & Thysen, H. C. (2024). *Choosing Between Causal Interpretations: An Experimental Study* (Working Paper).

Andre, P., Haaland, I., Roth, C., Wiederholt, M. & Wohlfart, J. (2026). Narratives About the Macroeconomy. *Review of Economic Studies*.

Ba, C., Dmitriev, D., Hang, Z. & Papazyan, F. (2026). *Strategically controlling worldviews* (Working Paper).

Barron, K., & Fries, T. (2025). *Narrative Persuasion* (Working Paper).

Blume, A., Lai, E. K. & Lim, W. (2020). Strategic information transmission: A survey of experiments and theoretical foundations. In C. M. Capra, R. T. A. Croson, M. L. Rigdon & T. S. Rosenblat (Eds.), *Handbook of experimental game theory* (pp. 311–347). Edward Elgar Publishing.

Bohren, J. A., & Hauser, D. N. (2021). Learning With Heterogeneous Misspecified Models: Characterization and Robustness. *Econometrica*, *89*(6).

Bursztyn, L., Rao, A., Roth, C. & Yanagizawa-Drott, D. (2023). Opinions as Facts. *Review of Economic Studies*, *90*(4).

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

Charles, C., & Kendall, C. (2025). *Causal Narratives* (Working Paper).

Charness, G., Feri, F., Meléndez-Jiménez, M. A. & Sutter, M. (2023). An Experimental Study on the Effects of Communication, Credibility, and Clustering in Network Games. *Review of Economics and Statistics*, *105*(6), 1530–1543.

Crawford, V. P., & Sobel, J. (1982). Strategic Information Transmission. *Econometrica*, *50*(6).

Dickhaut, J. W., McCabe, K. A. & Mukherji, A. (1995). An experimental study of strategic information transmission. *Economic Theory*, *6*(3), 389–403.

Eliaz, K., & Spiegler, R. (2020). A Model of Competing Narratives. *American Economic Review*, *110*(12).

Enke, B. (2020). What You See Is All There Is. *Quarterly Journal of Economics*, *135*(3).

Enke, B., & Zimmermann, F. (2019). Correlation Neglect in Belief Formation. *Review of Economic Studies*, *86*(1).

Esponda, I., & Pouzo, D. (2016). Berk-Nash Equilibrium: A Framework for Modeling Agents With Misspecified Models. *Econometrica*, *84*(3).

Esponda, I., & Vespa, E. (2018). Endogenous Sample Selection: A Laboratory Study. *Quantitative Economics*, *9*(1).

Esponda, I., Vespa, E. & Yuksel, S. (2024). Mental Models and Learning: The Case of Base-Rate Neglect. *American Economic Review*, *114*(3).

Eyster, E., & Weizsäcker, G. (2016). *Correlation Neglect in Financial Decision-Making* (Working Paper).

Fréchette, G. R., Lizzeri, A. & Perego, J. (2022). Rules and Commitment in Communication: An Experimental Analysis. *Econometrica*, *90*(5), 2283–2318.

Fréchette, G. R., Vespa, E. & Yuksel, S. (2024). *Extracting Models From Data Sets: An Experiment* (Working Paper).

Fudenberg, D., Romanyuk, G. & Strack, P. (2017). Active Learning with a Misspecified Prior.

Theoretical Economics, *12*(3).

Graeber, T., Roth, C. & Zimmermann, F. (2024). Stories, Statistics, and Memory. *Quarterly Journal of Economics*, *139*(4), 2181–2225.

Heidhues, P., Kőszegi, B. & Strack, P. (2018). Unrealistic Expectations and Misguided Learning. *Econometrica*, *86*(4), 1159–1214.

Jin, G. Z., Luca, M. & Martin, D. (2021). Is no news (perceived as) bad news? an experimental investigation of information disclosure. *American Economic Journal: Microeconomics*, *13*(2), 141–173.

Kamenica, E., & Gentzkow, M. (2011). Bayesian Persuasion. *American Economic Review*, *101*(6).

Milgrom, P. R. (1981). Good News and Bad News: Representation Theorems and Applications. *The Bell Journal of Economics*, *12*(2).

Schwartzstein, J., & Sunderam, A. (2021). Using Models to Persuade. *American Economic Review*, *111*(1), 276–323.

Sobel, J. (2020). Signaling games. In R. A. Meyers (Ed.), *Encyclopedia of complexity and systems science.* Springer.

Spiegler, R. (2020). Behavioral Implications of Causal Misperceptions. *Annual Review of Economics*, *12*.

Online Appendix for

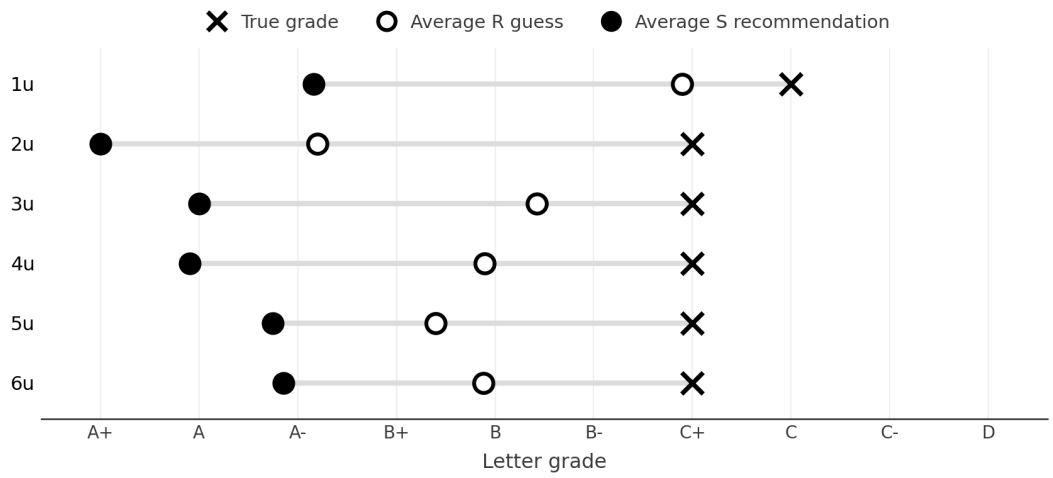
Information versus Interpretation: Evidence from an Experiment on Persuasion

Alisher Batmanov
UC San Diego

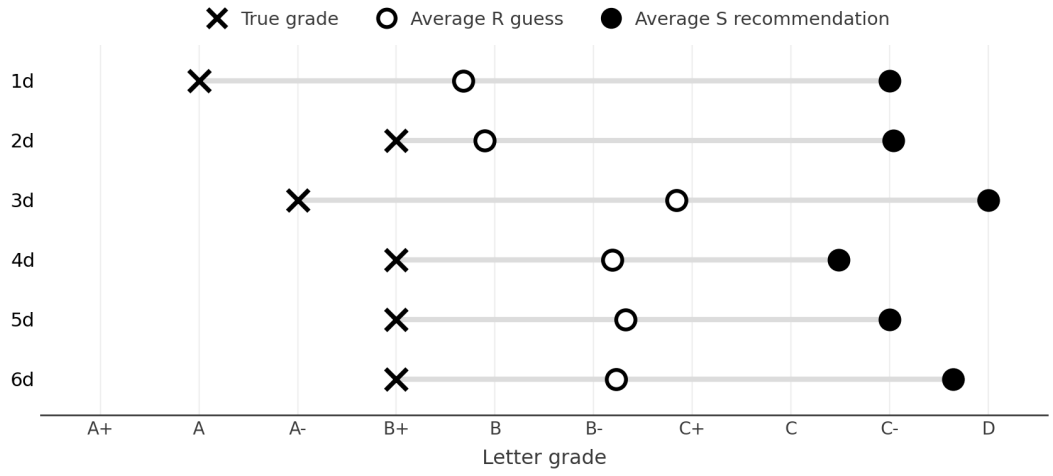
Bridget Galaty
UC San Diego

Contents:

A. Additional Figures and Tables



(a) UP-datasets

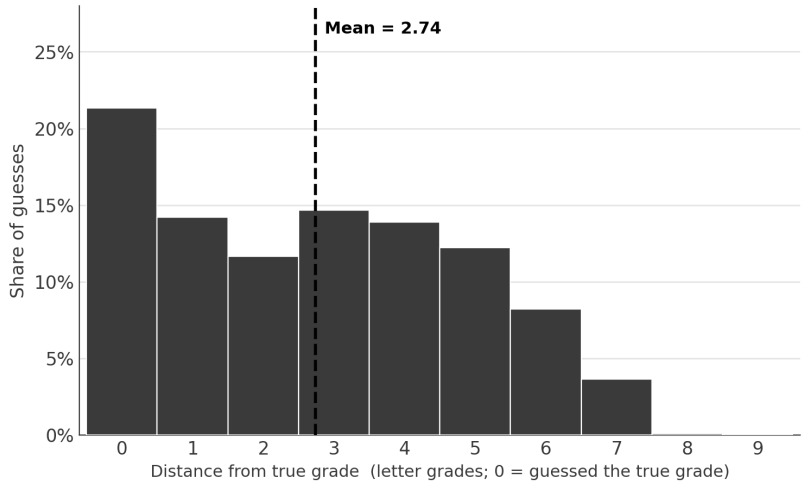


(b) DOWN-datasets

Figure A1: True Grade, Average Guess, and Recommendation by Dataset

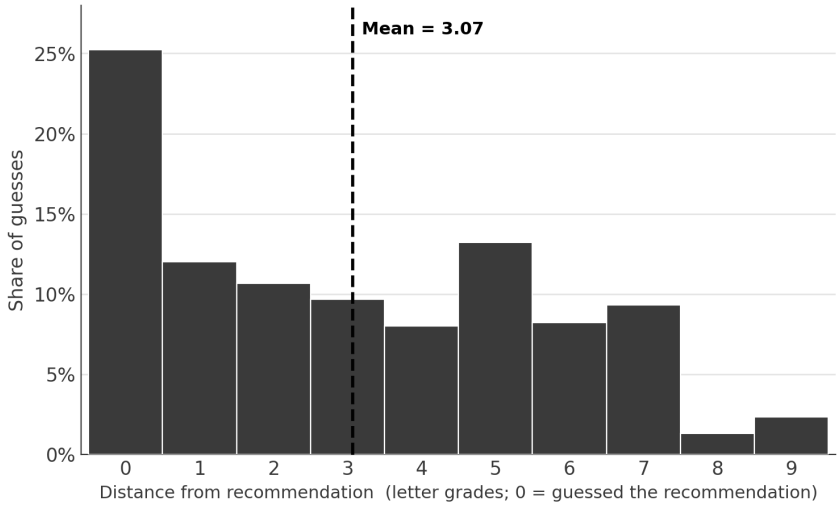
Notes: For each dataset, the figure marks the true grade, the average Receiver guess, and the Sender's recommendation on the grade scale. UP-datasets are those in which the true grade sits near D; DOWN-datasets are those in which the true grade sits near A+. Misaligned condition.

Figure A2: Distance Between Receivers' Guesses and the True Grade



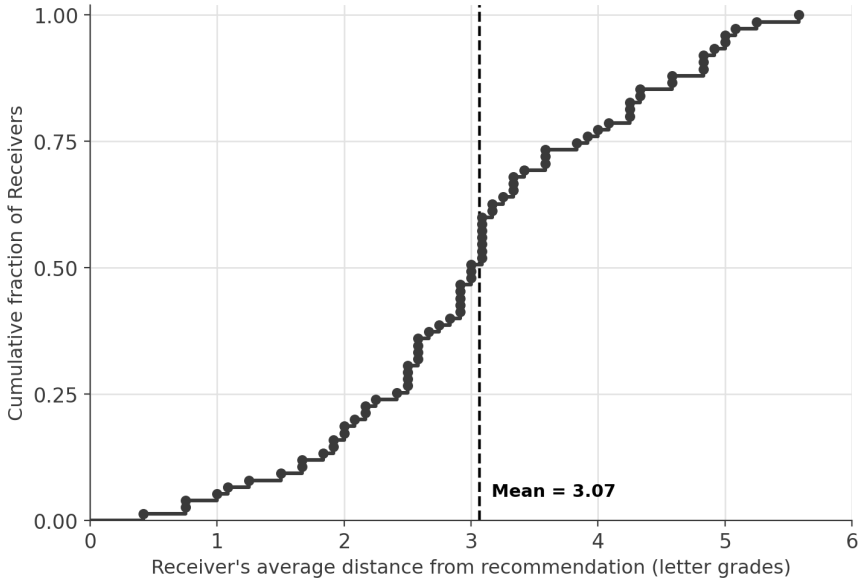
Notes: Distribution, across all Receiver guesses in the misaligned condition (75 Receivers each making one guess in each of the 12 datasets, $n = 900$), of the distance in letter grades between a guess and the true grade. The mean, marked by the dashed line, is about 2.74 grades.

Figure A3: Distance Between Receivers' Guesses and Senders' Recommendations



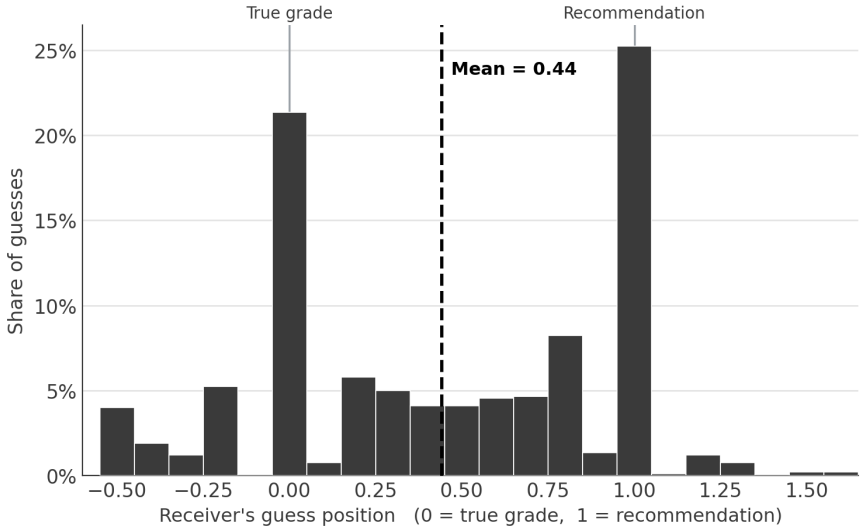
Notes: Distribution, across all Receiver guesses in the misaligned condition (75 Receivers, each making one guess in each of the 12 datasets, for $n = 900$ guesses), of the distance in letter grades between a guess and the recommendation the Receiver received. A value of zero indicates a guess equal to the recommendation. The dashed line marks the mean.

Figure A4: Distribution Across Receivers of the Average Distance from the Recommendation



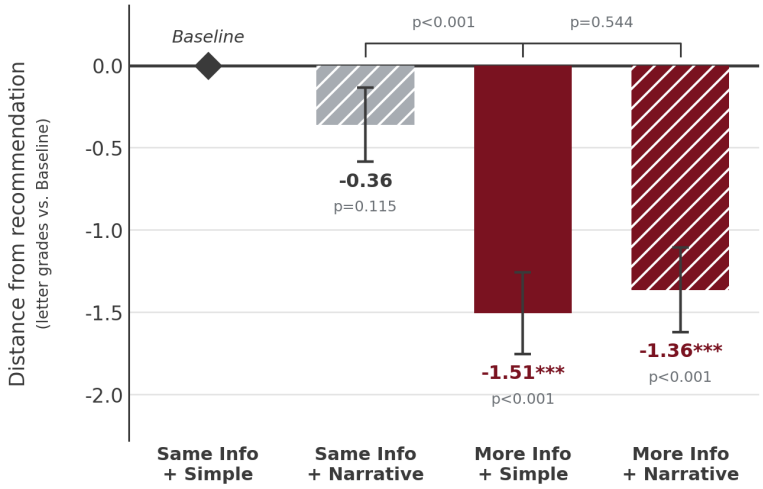
Notes: Empirical cumulative distribution, across the 75 Receivers in the misaligned condition, of each Receiver's average distance from the recommendation in letter grades, pooled across datasets. Values range from near zero, for Receivers who adopt the recommendation almost exactly, to more than five grades, for those who largely disregard it.

Figure A5: Distribution of Receivers' Guess Positions Between the True Grade and the Recommendation



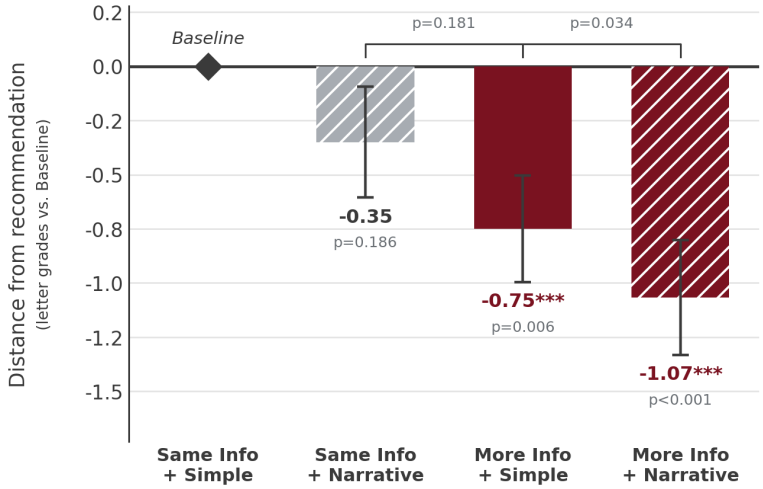
Notes: Distribution, across Receiver guesses in the misaligned condition, of the normalized position of each guess between the true grade and the recommendation, computed as (guess − truth) / (recommendation − truth). A value of 0 marks a guess at the true grade and a value of 1 a guess at the recommendation.

Figure A6: Distance from Recommendation by Message Type, in Letter Grades



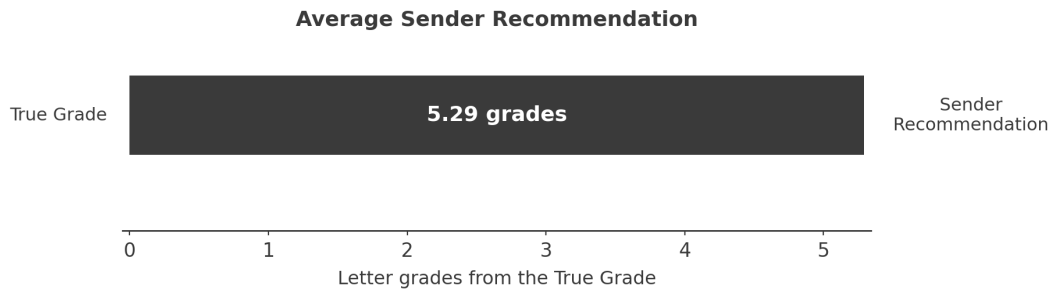
Notes: The letter-grade version of Figure 7. Each bar reports the average distance from the recommendation by message type as a difference, in letter grades, from the Same Info + Simple baseline, whose mean distance is 3.88 grades. Error bars denote ±1 participant-clustered standard error. Misaligned condition; $n = 900$. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Figure A7: Distance from Recommendation by Message Type, in Letter Grades: Aligned Preferences



Notes: The letter-grade version of Figure 10. Each bar reports the average distance from the recommendation by message type as a difference, in letter grades, from the Same Info + Simple baseline, whose mean distance is 1.23 grades. Error bars denote ±1 participant-clustered standard error. Aligned condition; $n = 240$. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Figure A8: Average Distance Between the True Grade and the Sender's Recommendation



Notes: The figure places the true grade and the Sender's recommendation on the A+ to D grade scale. Averaged across the selected misaligned messages, the recommendation lies about 5.29 letter grades from the truth and never fewer than three, a separation built into the design that sets the scope for persuasion.

Table A1: Factorial Estimates of the Information and Narrative Effects

	Misaligned			Aligned
	Distance from rec.	Distance from truth	Norm. position	Distance from rec.
	(1)	(2)	(3)	(4)
Informational Advantage	−1.507***	0.911***	0.226***	−0.750***
	(0.247)	(0.252)	(0.048)	(0.246)
Narrative	−0.360	0.240	0.055	−0.350
	(0.226)	(0.222)	(0.041)	(0.256)
Info. Advantage × Narrative	0.502	−0.369	−0.064	0.033
	(0.323)	(0.313)	(0.060)	(0.198)
Constant	3.876***	1.804***	0.316***	1.233***
	(0.194)	(0.192)	(0.034)	(0.297)
Contrast: Informational Advantage − Narrative				
Estimate	−1.147***	0.671**	0.171***	−0.400
	(0.253)	(0.262)	(0.051)	(0.289)
95% CI	[−1.65, −0.64]	[0.15, 1.19]	[0.07, 0.27]	[−1.00, 0.20]
R^2	0.060	0.021	0.040	0.083
Observations	900	900	900	240

Notes: Factorial parameterization of the 2×2 message-type design. *Informational Advantage* is the effect of the Sender's informational advantage within Simple messages, *Narrative* the effect of a narrative within Same Info messages, and the interaction their differential. *Distance from rec.* is the distance, in letter grades, between the Receiver's guess and the recommendation. *Distance from truth* is the signed distance, in letter grades, of the guess from the true grade in the direction of the recommendation, and *Norm. position* is (guess − truth)/(recommendation − truth); both use the recommendation the Receiver received. These two outcomes are defined only in the misaligned condition: in the aligned condition, better-informed Senders recommend the true grade, so the direction and the denominator are undefined. Standard errors, clustered at the participant level, are in parentheses, and the confidence intervals use the clustered standard errors. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table A2: Message-Type Contrasts on Identical-Recommendation Subsets

	Full sample	Identical-rec. subset	No. of datasets	No. of guesses
Narrative alone (Same+Narr − Same+Simple)	−0.360 (0.226)	−0.204 (0.244)	11	416
Information (More+Simple − Same+Simple)	−1.507*** (0.247)	−1.114*** (0.313)	6	220
Narrative on top (More+Narr − More+Simple)	0.142 (0.233)	0.302 (0.252)	9	324

Notes: Misaligned condition. Because the selected recommended grade is not always identical across message types within a dataset, each row re-estimates a pairwise contrast of the distance from the recommendation using only the datasets in which the two message types compared in that row share the same recommended grade, so that the contrast holds the recommendation literally fixed. *Full sample* reports the corresponding estimate on all twelve datasets, and *Identical-rec. subset* the estimate on the restricted sample. *No. of datasets* is the number of datasets, out of twelve, in which the two compared message types share the same recommended grade, and *No. of guesses* is the number of Receiver guesses made under either of the two compared message types in those datasets, that is, the number of observations underlying the subset estimate. Restricting instead to the five datasets in which all four recommendations coincide yields, relative to the Same Info + Simple baseline, −0.06 for Same Info + Narrative ($p = 0.84$), −1.20 for More Info + Simple ($p = 0.002$), and −0.66 for More Info + Narrative ($p = 0.10$). Standard errors, clustered at the participant level, are in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table A3: Attributes of the Selected Messages by Message Type

	Mean |rec. − truth| (letter grades)	Narrative length (words)
Panel A. Misaligned		
Same Info + Simple	5.58 (0.90)	—
Same Info + Narrative	5.42 (1.16)	23
More Info + Simple	5.00 (1.21)	—
More Info + Narrative	5.08 (1.31)	49
Panel B. Aligned		
Same Info + Simple	0.45 (0.52)	—
Same Info + Narrative	0.83 (1.11)	65
More Info + Simple	0.00 (0.00)	—
More Info + Narrative	0.00 (0.00)	44

Notes: Attributes of the selected messages, computed at the dataset level across the twelve datasets in each condition; standard deviations in parentheses. The recommended grades are broadly balanced across message types within each condition, with Same Info recommendations in the misaligned condition slightly more extreme than More Info ones; in the aligned condition better-informed Senders recommend the true grade by construction. Hand-coding the narrative texts in Appendix D, three of the twelve misaligned More Info + Narrative texts state the data-generating equation or invoke knowledge of the true relationship (datasets 1u, 4d, and 6u), and one further text misstates the mechanism (6d), while none of the Same Info + Narrative texts do either; excluding the first three datasets leaves the estimates essentially unchanged (Same Info + Narrative −0.21, $p = 0.46$; More Info + Simple −1.58, $p < 0.001$; More Info + Narrative −1.51, $p < 0.001$), as does additionally excluding 6d (−0.01, −1.60, and −1.49, with the same significance pattern). In the aligned condition such references are consistent with the Sender's disclosed information.

B. Theoretical Framework

This appendix presents a stylized model of the environment and derives the *rational benchmark*: what an inference-making Receiver would do in equilibrium when faced with a Sender who commands two instruments, an informational advantage and a narrative. As discussed in Section 3.2, the data reject the benchmark's central prediction for the information channel, and that rejection is precisely the object of interest: the strong response to a disclosed informational advantage is a deviation from equilibrium reasoning, not an implication of it.

B.1. Setup

States, data, and the Receiver's problem. There is an unknown state $\theta \in \mathbb{R}$, the true grade, drawn from a commonly known prior F. Neither player observes θ at the outset; both observe a dataset d of input–outcome pairs generated by some process, which is informative about θ but does not pin it down. To turn the data into a belief, a player must adopt a *model* κ, an account of how the inputs produce the outcome: a model together with the data yields a posterior $\Pr_\kappa(\theta \mid d)$. The Receiver holds his own model κ_R and, after observing the data and any message, chooses a guess $a \in \mathbb{R}$ to minimize the expected absolute error $\mathbb{E}\,|a - \theta|$, matching the experiment's payoff, so his optimal guess is a median of his belief.[33] Absent any message it is

$$a_0 \;=\; m \;\equiv\; \mathrm{med}_{\kappa_R}\big[\theta \mid d\big].$$

The Sender. Before the Receiver guesses, the Sender sends a message. She differs from the Receiver along two dimensions. Her *information*: she is either *uninformed*, observing only the data d, or *informed*, observing the realized state θ and the true model. And the *form* of her message: a *simple* message is a recommended grade $r \in \mathbb{R}$, while a *narrative* additionally proposes a model $\hat{\kappa}$, an alternative account of how the data map to the grade, which the Receiver may use in place of his own.

Preferences. The Sender is either *aligned* with the Receiver, $u_S(a, \theta) = -|a - \theta|$, so that she too prefers an accurate guess, or *misaligned*, in which case she is rewarded for the direction of the guess rather than its accuracy,

$$u_S(a, \theta) \;=\; \lambda a, \qquad \lambda \in \{+1, -1\},$$

with $\lambda = +1$ (-1) when she prefers higher (lower) guesses. The Receiver knows the Sender's

[33] Under quadratic loss the optimal guess would instead be the posterior mean; none of the statements below depend on which of the two is used.

information and preference types and the payment schedule; as in the experiment, he does not observe the realized direction λ of a misaligned Sender. We characterize equilibrium behavior and report the Receiver's guess a^*.

B.2. The Information Channel

Suppose first that the message acts only as a signal: the Receiver keeps his model κ_R and updates his belief about θ from the recommendation.

Proposition 1 (no transmission under misalignment). *Fix the misaligned Sender's direction λ. In every equilibrium, all on-path messages induce the same guess, and that guess is the no-message optimum: $a^* = m$. This holds whether or not the Sender is informed.* The argument is direct. Suppose two on-path messages induced different guesses; since $u_S = \lambda a$ is strictly monotone in the induced guess and independent of θ, every Sender type with direction λ strictly prefers the message inducing the more favorable guess, so the other message could not be on path. All on-path messages therefore induce a common guess a^*; each on-path posterior then has median a^*, and because these posteriors average to the no-message belief $\Pr_{\kappa_R}(\theta \mid d)$, the common guess is a median of that belief as well, so $a^* = m$.[34]

Remark 1 (unknown direction). In the experiment, the Receiver knows the misaligned payment schedule but not whether the current dataset rewards the Sender for higher or lower guesses. This only reinforces the benchmark. Within each direction, the pooling logic of Proposition 1 applies, so a recommendation can reveal at most the Sender's direction. In our design the direction is tied to the state by construction, with UP-datasets built so that the true grade is low and DOWN-datasets so that it is high, so an extreme recommendation, if it reveals anything, signals that the truth lies toward the *opposite* end of the scale. A rational Receiver would accordingly move against extreme recommendations or not at all; under no equilibrium reasoning does he move toward them.

Proposition 2 (truthful transmission under alignment). *Under aligned preferences with an informed Sender, truthful disclosure is an equilibrium: $r = \theta$ is incentive compatible, the Receiver believes it, and $a^*(\theta) = \theta$.* An informational advantage can thus be fully transmitted when interests

[34] The conclusion concerns actions: distinct on-path messages may in principle carry payoff-irrelevant information, but they cannot induce different guesses. An informational advantage is useless here because the obstacle is not what the Sender knows but that she cannot credibly convey it. Two further remarks. First, commitment: under quadratic loss, iterated expectations implies that no information structure, even one the Sender commits to, could move the average guess away from the Receiver's own reading; under the absolute loss used here that argument does not apply, but our Senders have no commitment technology in any case, so cheap talk is the relevant benchmark. Second, this is the standard construction; as usual in cheap-talk games, uninformative play is supported alongside any other equilibria by appropriate off-path beliefs.

align.[35]

B.3. The Narrative Channel

A narrative adds no signal about θ; it offers the Receiver a different model through which to read the shared data. Let models be drawn from a finite menu $\mathcal{K}$ of candidate accounts, with $\kappa_R \in \mathcal{K}$, mirroring the finite family of data-generating processes in the experiment. A proposed model $\hat{\kappa} \in \mathcal{K}$ implies the reading $a_{\hat{\kappa}} = \text{med}_{\hat{\kappa}}[\theta \mid d]$, against the Receiver's own $a_{\kappa_R} = m$. Following the narrative-persuasion literature, we endow the Receiver with a behavioral adoption rule rather than deriving one from equilibrium updating: he does not weigh $\hat{\kappa}$ by its prior plausibility or by the Sender's incentive to propose it, but adopts it when it explains the observed data at least as well as his own model, up to a tolerance $c \in (0, 1]$, as measured by the likelihood ratio

$$B(d) \;=\; \frac{L(d \mid \hat{\kappa})}{L(d \mid \kappa_R)} \;\geq\; c. \tag{1}$$

If $\hat{\kappa}$ is adopted the Receiver guesses $a_{\hat{\kappa}}$; otherwise he retains m.[36] Anticipating the rule, a misaligned Sender with $\lambda = +1$ proposes the model that induces the highest guess among those that fit well enough, so the Receiver's guess is

$$a^* \;=\; \max\Big\{\, a_\kappa \;:\; \kappa \in \mathcal{K},\; L(d \mid \kappa) \;\geq\; c\, L(d \mid \kappa_R) \,\Big\}, \tag{2}$$

and symmetrically the minimum for $\lambda = -1$; the maximum exists because $\mathcal{K}$ is finite and the feasible set contains κ_R. Two properties follow. First, persuasion through a narrative does not require an informational advantage: even an uninformed Sender *can* move the guess, because she changes the interpretation of the data rather than the information about θ, although no movement results when κ_R already induces the most favorable well-fitting reading. Second, the movement is bounded by fit within the menu: the Receiver cannot be taken past the most favorable candidate model that still satisfies $B(d) \geq c$. When the data sharply favor the Receiver's model, few alternatives clear the threshold and a^* stays near m; when several candidate models fit comparably, a^* can depart further from m.

[35] Proposition 2 is an existence statement: as in any cheap-talk game, an uninformative equilibrium supported by skeptical off-path beliefs also exists. We take the truthful equilibrium as the benchmark under alignment, in line with standard selection arguments favoring informative play when interests coincide.

[36] The adoption rule is a modeling primitive in the spirit of the fit-based criteria used in the narrative-persuasion literature, not an implication of the equilibrium concept of the previous subsection: a fully Bayesian Receiver would also condition on the Sender's strategic choice of which model to propose. We adopt the behavioral rule because it is the natural benchmark for how a narrative could persuade at all.

B.4. Benchmark Predictions and the Data

The benchmark orders the two channels as follows.

1. *Information without alignment does not persuade.* Under misaligned preferences a simple recommendation leaves the guess at $a^* = m$, whether or not the Sender is informed (Proposition 1), and with the direction unknown a rational response to extreme recommendations is, if anything, to move away from them (Remark 1).

2. *Narratives can persuade, within a bound set by fit.* A narrative can move the guess away from m in the Sender's preferred direction, up to the most favorable candidate reading that still fits the data, even under misalignment and with no informational advantage.

3. *Alignment restores transmission.* Under aligned preferences the guess moves to the truth, $a^* = \theta$: an informed Sender reveals it through a simple recommendation (Proposition 2), and a narrative points to the correct model.

The experimental results reverse the first two orderings. Receivers move sharply toward the recommendations of better-informed misaligned Senders, the one message attribute the benchmark says they should ignore, and respond little to narratives, the one instrument through which the benchmark allows persuasion under misalignment. The information-channel finding is thus a deviation from equilibrium reasoning, consistent with the receiver credulity documented in cheap-talk experiments, and the paper's analysis quantifies it. Only the third prediction is borne out qualitatively: under alignment Receivers follow better-informed Senders nearly to the truth. Two features of the mapping from model to experiment bear noting. The benchmark's empirically relevant content is the Receiver's best response to the messages he believes Senders send; our Receivers faced researcher-selected messages, but the selection mimics the strategic behavior the model ascribes to Senders, with misaligned recommendations chosen to point away from the truth in the Sender's paid direction. And the Receiver's reading m is itself the outcome of a difficult inference problem that subjects solve with error, which the benchmark takes as given.

C. Datasets in the Experiment

This appendix lists the datasets shown to subjects. Scores are those assigned by each data-generating process, with values below 0 or above 100 mapped to the endpoint grades D and A+; DGP 1_s is defined on the restricted support described in Section 2.3.

DGP 1_s - UP

id	x_1	x_2	x_3	**score**	y
1	8	8	No	25	C
2	2	2	Yes	75	A-
3	8	8	No	25	C
4	8	8	No	25	C
5	8	8	No	25	C
6	2	2	Yes	75	A-
7	2	2	Yes	75	A-
8	2	2	Yes	75	A-
9	2	2	Yes	75	A-
10	8	8	No	25	C
11	8	8	No	25	C

DGP 1 - DOWN

id	x_1	x_2	x_3	**score**	y
1	6	10	No	88	A
2	4	9	Yes	12	C-
3	1	8	Yes	12	C-
4	2	3	No	88	A
5	10	1	Yes	12	C-
6	2	5	Yes	12	C-
7	8	9	No	88	A
8	6	5	Yes	12	C-
9	5	10	No	88	A
10	10	7	No	88	A
11	1	8	No	88	A

DGP 2 - UP

id	x_1	x_2	x_3	**score**	y
1	7	4	Yes	12.0	C-
2	5	5	No	60.0	B+
3	1	8	No	108.0	A+
4	10	2	No	-90.0	D
5	4	9	Yes	78.0	A-
6	9	8	Yes	-52.0	D
7	6	1	No	38.0	C+
8	1	4	Yes	108.0	A+
9	10	4	No	-90.0	D
10	3	5	Yes	92.0	A+
11	6	8	No	38	C+

DGP 2 - DOWN

id	x_1	x_2	x_3	**score**	y
1	2	1	No	102.0	A+
2	6	7	No	38.0	C+
3	8	5	Yes	-18.0	D
4	4	1	No	78.0	A-
5	7	1	Yes	12.0	C-
6	3	3	Yes	92.0	A+
7	10	10	No	-90.0	D
8	2	10	No	102.0	A+
9	5	3	No	60.0	B+
10	9	4	Yes	-52.0	D
11	5	5	Yes	60.0	B+

DGP 3 - UP

id	x_1	x_2	x_3	score	y
1	9	7	No	30.0	C+
2	1	9	Yes	80.0	A
3	3	4	No	60.0	B+
4	6	8	No	20.0	C
5	2	8	Yes	70.0	A-
6	1	2	Yes	10.0	C-
7	9	1	No	90.0	A+
8	1	10	No	0.0	D
9	2	6	Yes	50.0	B
10	9	5	Yes	40.0	B-
11	3	7	No	30.0	C+

DGP 3 - DOWN

id	x_1	x_2	x_3	score	y
1	10	5	Yes	40.0	B-
2	9	5	No	50.0	B
3	6	2	Yes	10.0	C-
4	9	10	Yes	90.0	A+
5	3	7	No	30.0	C+
6	4	8	Yes	70.0	A-
7	6	10	No	0.0	D
8	6	7	Yes	60.0	B+
9	5	2	No	80.0	A
10	3	8	No	20.0	C
11	3	3	No	70.0	A-

DGP 4 - UP

id	x_1	x_2	x_3	score	y
1	10	8	No	80.0	A
2	1	1	No	25.0	C
3	8	10	Yes	50.0	B
4	2	8	Yes	0.0	D
5	7	6	Yes	60.0	B+
6	2	6	No	10.0	C-
7	10	9	No	75.0	A-
8	6	10	No	30.0	C+
9	7	10	No	40.0	B-
10	10	3	No	100.0	A+
11	6	10	Yes	30.0	C+

DGP 4 - DOWN

id	x_1	x_2	x_3	score	y
1	3	7	Yes	15.0	C-
2	10	9	Yes	75.0	A-
3	1	6	No	0.0	D
4	5	5	Yes	45.0	B-
5	7	1	No	85.0	A
6	6	3	No	65.0	B+
7	2	1	No	35.0	C+
8	10	3	Yes	100.0	A+
9	5	10	No	20.0	C
10	7	7	No	55.0	B
11	5	2	No	60.0	B+

DGP 5 - UP

id	x_1	x_2	x_3	score	y
1	8	6	Yes	65.0	B+
2	10	9	No	15.0	C-
3	1	2	No	70.0	A-
4	6	2	Yes	35.0	C+
5	5	5	Yes	45.0	B-
6	3	10	No	90.0	A+
7	3	3	Yes	25.0	C
8	5	6	No	50.0	B
9	1	1	Yes	5.0	D
10	9	9	Yes	85.0	A
11	9	10	No	30.0	C+

DGP 5 - DOWN

id	x_1	x_2	x_3	score	y
1	9	4	No	0.0	D
2	7	1	Yes	35.0	C+
3	9	9	Yes	85.0	A
4	3	2	Yes	20.0	C
5	10	10	Yes	95.0	A+
6	3	8	Yes	50.0	B
7	8	4	No	10.0	C-
8	3	6	Yes	40.0	B-
9	8	7	Yes	70.0	A-
10	9	4	Yes	60.0	B+
11	3	4	No	60.0	B+

DGP 6 - UP

id	x_1	x_2	x_3	**score**	y
1	5	10	Yes	100.0	A+
2	4	5	No	51.0	B
3	4	1	Yes	7.0	D
4	4	10	Yes	100.0	A+
5	10	4	Yes	16.0	C-
6	6	5	No	49.0	B-
7	6	8	No	88.0	A
8	6	4	Yes	20.0	C
9	4	2	No	30.0	C+
10	9	7	No	70.0	A-
11	9	3	No	30.0	C+

DGP 6 - DOWN

id	x_1	x_2	x_3	**score**	y
1	4	3	Yes	15.0	C-
2	10	9	No	100.0	A+
3	10	2	Yes	4.0	D
4	7	5	No	48.0	B-
5	8	9	Yes	83.0	A
6	1	7	Yes	58.0	B
7	7	3	No	32.0	C+
8	10	10	No	100.0	A+
9	9	5	Yes	26.0	C
10	5	7	No	74.0	A-
11	10	8	Yes	64.0	B+

Dataset 13: DGP 4 - UP

id	x_1	x_2	x_3	**score**	y
1	8	3	Yes	85	A
2	10	6	Yes	90	A+
3	8	6	No	70	A-
4	7	8	No	50	B
5	7	5	No	65	B+
6	4	4	Yes	40	B-
7	3	6	No	20	C
8	5	8	Yes	30	C+
9	1	4	Yes	10	C-
10	1	9	No	0	D
11	5	7	No	35	C+

DGP 8

id	x_1	x_2	x_3	**score**	y
1	5	6	No	95	A+
2	2	1	Yes	65	B+
3	8	1	No	95	A+
4	2	5	No	95	A+
5	7	10	Yes	65	B+
6	8	2	No	95	A+
7	8	4	Yes	65	B+
8	2	5	Yes	65	B+
9	2	7	No	95	A+
10	8	10	Yes	65	B+
11	2	7	Yes	65	B+
12	6	3	No	95	A+
13	8	4	No	95	A+

DGP 9

id	x_1	x_2	x_3	**score**	y
1	8	4	Yes	65	B+
2	5	1	Yes	95	A+
3	1	5	Yes	55	B
4	8	6	Yes	45	B-
5	9	7	Yes	35	C+
6	9	8	No	25	C
7	4	3	Yes	75	A-
8	4	2	No	85	A
9	5	9	No	15	C-
10	9	10	No	5	D
11	3	3	No	75	A-
12	6	7	Yes	35	C+
13	2	10	Yes	5	D

D. Narratives and Message Selection

This appendix describes how the messages shown to Receivers were selected and lists the full set of selected messages. In the Sender sessions, Senders submitted a recommendation for each of the twelve datasets and, in the narrative treatments, a written explanation for it. From these submissions we selected, within each preference condition, one message of each of the four Sender types per dataset, together with the two messages used in the message-choice round, for a total of fifty per condition, or one hundred in all.

Selection followed three principles. First, the recommended grade had to reflect the Sender's induced preferences: in the misaligned condition, a high grade for UP-datasets and a low grade for DOWN-datasets, so that the message pulls against the truth; in the aligned condition, a grade close to the truth. Second, when a narrative was provided, we chose explanations that described patterns present in the ten-row dataset rather than fabricated ones, so that the narratives were largely grounded in the data and differed from simple messages primarily in providing an explanation; a small number of the selected texts reference the true data-generating process or state a relationship imperfectly: in the misaligned condition, the More Info + Narrative texts for datasets 1u and 6u invoke knowledge of the true relationship, the 4d text misstates the constant term of the data-generating equation, and the 6d text misattributes to x_3 a switching role it does not have. Table A3 codes these attributes, and excluding the affected datasets leaves the estimates unchanged. Third, we held the recommended grade as constant as possible across Sender types within a dataset, so that differences in Receiver responses across message types reflect the type of message rather than differences in the grade being recommended.

The complete list of selected recommendations and narratives, by dataset and Sender type, is reported below.

Table D1: Narratives — Misaligned Condition

Dataset	Truth	More Info + Simple	More Info + Narrative		Same Info + Simple	Same Info + Narrative	
1d	A	C-	C-	the score is only A and C- so there is a relationship b/w x1, and x2 that determines the score. it could be either A or C-, and #11 is the same as set #3 so I would suggest C- more than A	C-	C-	see the number 3 student the x3 variable doens't affect the equation
1u	C	A-	A	The true relationship is 10*x1 when No	A-	B	10(x1)+ (-5) +10 x2 if x3 is no
2d	B+	C	C	It seems as though an increase in x1 and x2 contributes to a lower grade, looking at the other datasets. I would recommend giving this 11th student something around a C because we see that as x1 and x2 are super low, the score is very very high. seeing as student 11 has 5 and 5, the score isn't the highest, but it isn't a total d either. There is still a slight boost given due to the yes of student 11.	D	D	see row #7 → when x1=x2, y=D
2u	C+	A+	A+	When x3 = No, the grade will fall into A+ range when x1 < x2. You can see #2 when x1=x2 the grade is slightly lower, and # 4, 7, 9 the grades are more lower.	A+	A+	by looking at the pattern, x1 smaller than x2 would make the grade higher

Continued on next page

Table D1 *(continued)*

Dataset	Truth	More Info + Simple	More Info + Narrative		Same Info + Simple	Same Info + Narrative	
3d	A-	D	D	No tends to have lower grades even though it does no apply on all but when x1 is smaller and having a no makes the score small take example of number 5 and 10	D	D	The higher x1 and x2s are leading to higher scores, but the lower scores are leading to lower scores. The ones with no are also very low scores, and this leads us to see a negative linear relationship between the whole thing. 3 is a low number, and the no students are all low as well, so I would recommend giving the 11th student a D, or something very close to a D.
3u	C+	A	A	there are some complicated calculation regarding the relationship with x1 and x2, but x3 doesn't count into the calculation of score. the lower x1 is, the higher score would tend to be, but it is also depending on x2 with exponential relationship (I think), it's either add up or distract from. But I would say overall x1 is the most influential factor.	A	A	the relationship appears to be that the lower x1 is and the higher x2 is the higher y is. since x1 is 3 for #11 and x2 is 7 the grade is likely to be an A
4d	B+	C	C	score = 10x1 - 5x2 - 20	C-	C-	relationship if x1+x2=7 then y =d

Continued on next page

Table D1 *(continued)*

Dataset	Truth	More Info + Simple	More Info + Narrative		Same Info + Simple	Same Info + Narrative	
4u	C+	A-	A-	The 11th student has 6 in x1 and 10 in x2. While there is another student with the same inputs, the difference between this student and the other student is that the other student's x3 is a no. It seems that the score is largely effected by Yes and No, and so while x1 and x2 are contributing factors to the overall score, we also have to look at x3 in order to know the true grade of the 11th student. I honestly think that the true grade is around an A, mainly because of x3 being yes, along with x1 and x2 being fairly higher.	A+	A+	x1 is the main one considered; the first number times ten. but, if yes, it adds 30. if the end value is higher than 100, it goes and subtracts 100 (hence why some numbers seem to be high but get a low grade)
5d	B+	C-	C-	If x3 is No, then the score calculated using x1 and x2 is dropped down quite a lot. Notice how all the x3=Yes grades are at least a B (only one anomaly with Student 4 because x1 and x2 are small), but all the x3=No grades are quite low.	C-	C-	when the x3 is 'No' and x2 = 4, the student gets a lower grade.

Continued on next page

Table D1 *(continued)*

Dataset	Truth	More Info + Simple	More Info + Narrative		Same Info + Simple	Same Info + Narrative	
5u	C+	A	A-	The relationship is if x2 is higher than x1 and the gap is small, then it will be a higher grade x3 isn't dependent. If the gap is larger than 5 this doesn't matter.	A-	A-	its the relationship using x1-x2 if it is positive, the bigger the gap is to determine a better grade if it is negative, the bigger the gap (more negative) means a lower grade.
6d	B+	C-	D	The higher the x1, the lower the score. But if x3 is a no, the score is higher. Take a look at student 8. Has a x1 of 10, and an x3 of no. The relationship is set up that the x3 sets up whether or not the relationship is a positive linear one or a negative linear one. If student 8 with an x1 of 10 has an A+ because of the No, does that not mean that student 11 has a really low grade of a D because of the Yes?	D	D	if x3=yes, x1=10 always D
6u	C+	B+	B+	You can see that in round 10, when X1 is 9, X2 is 7, and X3 is no, y is A-. I can't tell you the true relationship but y is B+	A	A	The 10th data set has very similar numbers as the 11th set and it had an output of A-
7u	C+	A-	—	—	—	A-	x3 being no results in a relatively higher grade, and x1 being higher means a higher grade as well.

Table D2: Narratives — Aligned Condition

Dataset	Truth	More Info + Simple	More Info + Narrative		Same Info + Simple	Same Info + Narrative	
1d	A	A	A	The pattern of the dataset is that any student ID with an x3 input of "No" has a corresponding grade of "A" (80-89.) I can see the real grade of student 11 and it is A. The x3 input for student 11 is also a "no."	A	A	the trend I am seeing is that x3 decides the answer. if it is yes, student gets a c-, if no, it is an a. this is strengthened by the fact that every number might be vastly different, but whenever it is yes, student gets a c- and a if it is a no.
1u	C	C	C	Because the 10th student has the exact same data.	C	C	I did not get a true relationship in this data set, however based on the past data, there are 5 students who have the exact same three variables, and all of their grades remained consistent. Since these three variables of 8, 8 and no remained consistent, I would say that C is the best choice to make.

Continued on next page

Table D2 *(continued)*

Dataset	Truth	More Info + Simple	More Info + Narrative		Same Info + Simple	Same Info + Narrative	
2d	B+	B+	B+	I know the true relationship, as the interface states that the letter grade depends only on input X1. As X1 increases, the score falls rapidly in a nonlinear way, producing very high grades for small X1 and low grades for large X1. Specifically, score = 110-(2X1)ˆ2. When you calculate this, it would be 110-(2*5ˆ2)=110-(2*25)=110-50=60. 60 = B+ range, so therefore, that's what it would fall under.	B	B	In the case of "yes," a lower x1 merits a higher grade at an exponential level. The x2 appears to have minimal, if any effect by comparison.
2u	C+	C+	C+	grade depends on x1, increase in x1 leads to decrease in grade, grade score = 110-2(x1)ˆ2	C+	C+	The grade with No is lower and it depends on the value of x1. The lower the x1, the higher the letter grade. Since #11 has the same x1 as #7, their letter grades should be the same.

Continued on next page

Table D2 *(continued)*

Dataset	Truth	More Info + Simple	More Info + Narrative		Same Info + Simple	Same Info + Narrative	
3d	A-	A-	A-	it is important to look at the trend: when x3 = yes, higher x2 leads to higher grades, when x3 = no, higher x2 leads to lower grades. the relationship is score = 10(x2-1) if x3 = yes, score = 10(10-x2) if x3 = no	A-	A-	still looking at only the No s, #5 and #10, x2 increase final decrease, about 1:10 in ratio, looking at #7 and #10 (cant fine two same x2 so find one similar), we expected the final to increase from #7 to #10 by about 20 as x2 decreases, which specifically fits in the result. According to what is on the data set, I guess that the final answer is highly related if not only related to x2 (you could check it yourself from the data set). when x2=2 you get an A (#9), when x2 = 5 you get a B (#2). The unknown x2 is closer to 2 than 5, so A-
3u	C+	C+	C+	The relation will be score = 10(x2-1) if x3=Yes and 10(10-x2) if x3=No. So the score for the 11th student will be 10(10-7)=30, which is a C+	C+	C+	if x3=no, low x2 score leads to high grade and high x2 score leads to low grade. I used student id1 to choose grade c+

Continued on next page

Table D2 *(continued)*

Dataset	Truth	More Info + Simple	More Info + Narrative		Same Info + Simple	Same Info + Narrative	
4d	B+	B+	B+	The letter grade depends on both x1 and x2. Higher x1 raises the score, while higher x2 lowers it, with both inputs contributing linearly. Specifically, score = 10(X1) - 5(X2) + 20. Specifically, score = 10(5) - 5(2) + 20 = 50-10+20=40+20=60. 60=B+ on the grading scale, so B+	B	B	The pattern appears to be that x3 as "yes" means that y = (x1 * 10) - 2(x2), at least approximately. In scenarios of x3 as "no," the lower the x1, the lower the final grade. A higher x2 also appears to subtract more from the final grade. Since student 6's dataset is similar to that of student 11 and obtained a B+, it would be safest to guess a B for student 11, who has a lower x1 number.
4u	C+	C+	C+	I was able to get to know the true relationship of the letter grade. The letter grade depends on both X1 and X2 variables. The higher the X1 variable, the higher the score, while having a high X2 variable lowers it. Both inputs contribute to the letter grade linearly. Specifically, the equation to calculate the score is: 10*(X1) - 5*(X2) + 20. Using this equation, 10*(6) - 5*(10) + 20 = 60-50+20 = 30. C+=30, therefore the grade is C+	C	C	x3 is yes and looking at sample data where x3 is also yes, you see that 8,10 is B and 2,8 is D so my guess is that x1 and x2 being equal to each other causes a higher grade so 6,10 had 4 numbers in difference so C and this confirms because 7,6 has one number difference and it was given higher grade than 8,10

Continued on next page

Table D2 *(continued)*

Dataset	Truth	More Info + Simple	More Info + Narrative		Same Info + Simple	Same Info + Narrative	
5d	B+	B+	B+	For "no," a higher x1 means a lower grade. For every "no" with an x2 of 4 (like student 11), if x1 drops by one point the grade will have a corresponding 10 point increase. Example demonstrated in the dataset: Student 1 has x1=9, x2=4, x3="no," grade = D. Student 7 has x1=8, x2=4, x3="no," grade = C-. Therefore, x1=7 means grade = C, x1=6 means grade = C+. So on and so forth until student 11. Student 11 has x1=3, x2=4, x3="no," grade = B+. Source: I can see the true grade.	C	C	I was not given a true relationship or y-value on this one so I tried to guess the relationship. To be honest I have no clue? I think the x3 part matters a lot. All of the ones where x3 = No were very low in grade so I believe this must be the same.
5u	C+	C+	C+	A should be the student 11's grade because if x3=Yes, then follow the score of 5x1+5x2-5 and if x3 is no, then score=-10x1+5x2+70 so since student 11 is no for x3, then -10(9)+5(10)+70 = -90+50-5 = -40+70 = 30 so C+	C	C	The lower x1 is for "No," the higher the corresponding grade is, while a higher x2 for "no" appears to make the grade higher as well. However, the value of x1 appears to be much more important for what the grade is than the value of x2. Since student 2 has a higher x1 of 1 than student 11 and earned a C-, we can assume that the grade of student 11 is around C+, as the x1 is slightly lower and x2 slightly higher.

Continued on next page

Table D2 *(continued)*

Dataset	Truth	More Info + Simple	More Info + Narrative		Same Info + Simple	Same Info + Narrative	
6d	B+	B+	B+	score = -x1 + x2^2 + 10 if x3 = Yes and score = -x1 + x2^2 + 30 if x3 = No. this one was given, freebie	A-	A-	My recommendation is because student id #5 had numbers 8, 9, and Yes and got a grade of A; student #11 has 10, 8, and Yes, which is very similar, so to be safe I would recommend going close to A.
6u	C+	C+	C+	It gave me the letter grade in addition to sharing the relationship/equations. I calculated it and got the same score. The equations are -x1+x2^2+10 if x3 is Yes, -x1+x2^2+30 if x3 is No	C	C	looking at only the No s in the graph (because no matter x3 relevant or not, only look at the No s will not have the wrong answer), what seems interesting to me is #6 and #7, both x1 are 6, a higher x2 gives a higher grade, and looking at #2 and #6, both x2 is 5, a higher x1 gives a lower score, thus the final function as to be positively related to x2 and negatively related to x1. Furthermore, a 3 increase in x2 leads to a roughly 40 increase to score (according to #6 and #7), a 2 increase in x1 pulls down roughly 10 marks. index of x2 = +40/3 = 13, index of x1 = -10/2 = -5, check with #9 to find the possible constant = 30. final answer: 30 + 13*3 - 9*5 = 24

Continued on next page

Table D2 *(continued)*

Dataset	**Truth**	**More Info + Simple**	**More Info + Narrative**		**Same Info + Simple**	**Same Info + Narrative**	
7u	C+	C+	—	—	—	B	I was not given a true relationship or y-value. For the 11th student x1 = 5, x2 = 7, and x3 = No. I don't know the relationship but I kinda found a pattern. For the ones where x3=No (the only ones that matter), I would do score = (x1-1)10 + x2. This would give me the correct grade most of the time or would be off by one grade usually. Therefore I think it is B

E. Experimental Instructions Handouts

This appendix reproduces the instruction handouts distributed to subjects. The experiment was run in two session formats, joint sessions, in which Senders and Receivers participated together, and Receiver-only sessions, each conducted under both misaligned and aligned preferences. The handouts appear in the following order:

- **Joint sessions (Misaligned):** Sender instructions, then Receiver instructions.
- **Joint sessions (Aligned):** Sender instructions, then Receiver instructions.
- **Receiver-only sessions (Misaligned):** Receiver instructions.
- **Receiver-only sessions (Aligned):** Receiver instructions.

BG (M) — Decision Study · Instructions for SENDERS

Dataset Information

- There are **15 rounds** total: **13 main rounds** + **2 bonus rounds**. Each round shows a new, independently generated dataset — use only the current round's data to make your choice.
- Each dataset has **10 students**. For each student you see three inputs: $x_1, x_2 \in \{1, 2, \ldots, 10\}$ and $x_3 \in \{\text{Yes}, \text{No}\}$.
- These inputs produce a 0–100 score that maps to a **letter grade**:

  90–100=A+, 80–89=A, 70–79=A-, 60–69=B+, 50–59=B, 40–49=B-, 30–39=C+, 20–29=C, 10–19=C-, 0–9=D
- Each input can have a positive, negative, or no effect on the grade.
- The output you observe is the **letter grade** $y \in \{\text{A+, A, A-, B+, B, B-, C+, C, C-, D}\}$ (the 0–100 score is not displayed).

You also see the inputs for an **11th student**, but you do **not** observe their grade **y**.

Sender's Task

- Your task is to **recommend a letter grade** for the 11th student to the Receiver you are matched with. You are paid based on the Receiver's guess **after** they see your recommendation.
  - The Receiver will **not** know whether the dataset is an **UP** or a **DOWN** dataset, and will not see the true input–grade relationship.
  - Sometimes you learn the **true relationship** between inputs and grade, and see the 11th student's true grade; in other rounds you do not.
  - Sometimes you only recommend a grade; in other rounds you also provide a **written explanation** for why the Receiver should follow your recommendation.

Sender's Payment (main rounds)

Each dataset is labelled **UP** or **DOWN**. In **UP-datasets** you are paid more the **higher** the Receiver's guess; in **DOWN-datasets** you are paid more the **lower** the Receiver's guess. Payments range from **$24** to **$6**. At the end of the session, **one of the 13 main rounds is picked at random** and you are paid for that round based on the Receiver's guess in it.

	Receiver's guess									
	A+	A	A-	B+	B	B-	C+	C	C-	D
UP-dataset	$24	$22	$20	$18	$16	$14	$12	$10	$8	$6
DOWN-dataset	$6	$8	$10	$12	$14	$16	$18	$20	$22	$24

Example. Suppose the randomly-picked round turns out to be an **UP-dataset**. Then if your matched Receiver guessed A+ in that round, you earn **$24**; if they guessed D, you earn **$6**. For a **DOWN-dataset** the pattern is reversed.

So across rounds the incentive switches: in **UP** rounds you earn more by persuading the Receiver toward a **higher** grade, and in **DOWN** rounds by persuading them toward a **lower** grade — regardless of the true grade.

Bonus Rounds

- After the 13 main rounds, you complete 2 bonus rounds, in which you can earn an extra $5.
- Total payment = main-round payment + bonus-round payment, up to $29 total.

BG (M) — Decision Study · Instructions for RECEIVERS

Dataset Information

- There are **15 rounds** total: **13 main rounds** + **2 bonus rounds**. Each round shows a new, independently generated dataset — use only the current round's data to make your choice.
- Each dataset has **10 students**. For each student you see three inputs: $x_1, x_2 \in \{1,\ldots,10\}$ and $x_3 \in \{\text{Yes, No}\}$.
- These inputs produce a 0–100 score that maps to a **letter grade**:

  90–100=A+, 80–89=A, 70–79=A-, 60–69=B+, 50–59=B, 40–49=B-, 30–39=C+, 20–29=C, 10–19=C-, 0–9=D
- Each input can have a positive, negative, or no effect on the grade.
- The output you observe is the **letter grade** $y \in \{\text{A+, A, A-, B+, B, B-, C+, C, C-, D}\}$ (the 0–100 score is not displayed).

You also see the inputs for an **11th student**, but you do **not** observe their grade **y**.

Receiver's Task

- Your task is to **guess the true grade** of the 11th student. You will not see the true input–grade relationship, nor whether the dataset is an **UP** or **DOWN** dataset.
- In each round you see a **recommendation from a Sender**. Sometimes the Sender knows the **true relationship** and sees the true grade; sometimes they do not. Sometimes they only recommend a grade; sometimes they also give a **written explanation**.

Receiver's Payment (main rounds)

You are paid more the closer your guess is to the true grade of the 11th student, from **$24** (correct) down to **$6** (9 grades away). Full schedule below.

Your Guess \ True Grade	A+	A	A-	B+	B	B-	C+	C	C-	D
A+	$24	$22	$20	$18	$16	$14	$12	$10	$8	$6
A	$22	$24	$22	$20	$18	$16	$14	$12	$10	$8
A-	$20	$22	$24	$22	$20	$18	$16	$14	$12	$10
B+	$18	$20	$22	$24	$22	$20	$18	$16	$14	$12
B	$16	$18	$20	$22	$24	$22	$20	$18	$16	$14
B-	$14	$16	$18	$20	$22	$24	$22	$20	$18	$16
C+	$12	$14	$16	$18	$20	$22	$24	$22	$20	$18
C	$10	$12	$14	$16	$18	$20	$22	$24	$22	$20
C-	$8	$10	$12	$14	$16	$18	$20	$22	$24	$22
D	$6	$8	$10	$12	$14	$16	$18	$20	$22	$24

Bonus Rounds

- Before the 13 main rounds, you complete 2 bonus rounds, in which you can earn an extra $5.
- Total payment = main-round payment + bonus-round payment, up to $29 total.

BG (A) — Decision Study · Instructions for SENDERS

Dataset Information

- There are **15 rounds** total: **13 main rounds** + **2 bonus rounds**. Each round shows a new, independently generated dataset — use only the current round's data to make your choice.
- Each dataset has **10 students**. For each student you see three inputs: $x_1, x_2 \in \{1, 2, \ldots, 10\}$ and $x_3 \in \{\text{Yes}, \text{No}\}$.
- These inputs produce a 0–100 score that maps to a **letter grade**:

  90–100=A+, 80–89=A, 70–79=A-, 60–69=B+, 50–59=B, 40–49=B-, 30–39=C+, 20–29=C, 10–19=C-, 0–9=D
- Each input can have a positive, negative, or no effect on the grade.
- The output you observe is the **letter grade** $y \in \{\text{A+, A, A-, B+, B, B-, C+, C, C-, D}\}$ (the 0–100 score is not displayed).

You also see the inputs for an **11th student**, but you do **not** observe their grade **y**.

Sender's Task

- Your task is to **recommend a letter grade** for the 11th student to the Receiver you are matched with. You are paid based on the Receiver's guess **after** they see your recommendation.
  - The Receiver will **not** see the true input–grade relationship.
  - Sometimes you learn the **true relationship** between inputs and grade, and see the 11th student's true grade; in other rounds you do not.
  - Sometimes you only recommend a grade; in other rounds you also provide a **written explanation** for why the Receiver should follow your recommendation.

Sender's Payment (main rounds)

Your payoff is **equal to the Receiver's payoff**: you are paid more the closer the Receiver's guess is to the true grade of the 11th student, from **$24** (correct) down to **$6** (9 grades away).

Example. Suppose the randomly-picked round has true grade **A+**. If your matched Receiver guessed **A+** in that round, you earn **$24**; if they guessed **D**, you earn **$6**.

Both you and your matched Receiver earn more when the Receiver's guess is closer to the true grade. So you are always rewarded for sending recommendations that help the Receiver guess accurately.

At the end of the session, **one of the 13 main rounds is picked at random** and you are paid for that round based on the Receiver's guess in it.

Bonus Rounds

- After the 13 main rounds, you complete 2 bonus rounds, in which you can earn an extra $5.
- Total payment = main-round payment + bonus-round payment, up to $29 total.

BG (A) — Decision Study · Instructions for RECEIVERS

Dataset Information

- There are **15 rounds** total: **13 main rounds** + **2 bonus rounds**. Each round shows a new, independently generated dataset — use only the current round's data to make your choice.
- Each dataset has **10 students**. For each student you see three inputs: $x_1, x_2 \in \{1,\ldots,10\}$ and $x_3 \in \{\text{Yes, No}\}$.
- These inputs produce a 0–100 score that maps to a **letter grade**:

  90–100=A+, 80–89=A, 70–79=A-, 60–69=B+, 50–59=B, 40–49=B-, 30–39=C+, 20–29=C, 10–19=C-, 0–9=D
- Each input can have a positive, negative, or no effect on the grade.
- The output you observe is the **letter grade** $y \in \{\text{A+, A, A-, B+, B, B-, C+, C, C-, D}\}$ (the 0–100 score is not displayed).

You also see the inputs for an **11th student**, but you do **not** observe their grade **y**.

Receiver's Task

- Your task is to **guess the true grade** of the 11th student. You will not see the true input–grade relationship, nor whether the dataset is an **UP** or **DOWN** dataset.
- In each round you see a **recommendation from a Sender**. Sometimes the Sender knows the **true relationship** and sees the true grade; sometimes they do not. Sometimes they only recommend a grade; sometimes they also give a **written explanation**.

Receiver's Payment (main rounds)

You are paid more the closer your guess is to the true grade of the 11th student, from **$24** (correct) down to **$6** (9 grades away). Full schedule below.

Your Guess \ True Grade	A+	A	A-	B+	B	B-	C+	C	C-	D
A+	$24	$22	$20	$18	$16	$14	$12	$10	$8	$6
A	$22	$24	$22	$20	$18	$16	$14	$12	$10	$8
A-	$20	$22	$24	$22	$20	$18	$16	$14	$12	$10
B+	$18	$20	$22	$24	$22	$20	$18	$16	$14	$12
B	$16	$18	$20	$22	$24	$22	$20	$18	$16	$14
B-	$14	$16	$18	$20	$22	$24	$22	$20	$18	$16
C+	$12	$14	$16	$18	$20	$22	$24	$22	$20	$18
C	$10	$12	$14	$16	$18	$20	$22	$24	$22	$20
C-	$8	$10	$12	$14	$16	$18	$20	$22	$24	$22
D	$6	$8	$10	$12	$14	$16	$18	$20	$22	$24

Bonus Rounds

- Before the 13 main rounds, you complete 2 bonus rounds, in which you can earn an extra $5.
- Total payment = main-round payment + bonus-round payment, up to $29 total.

BG (M) — Decision Study · Instructions for RECEIVERS

Dataset Info

- There are **15 rounds** total: **13 main rounds** + **2 bonus rounds**. Each round shows a new, independently generated dataset — use only the current round's data to make your choice.
- Each dataset has **10 students**. For each student you see three inputs: $x_1, x_2 \in \{1, 2, \ldots, 10\}$ and $x_3 \in \{\text{Yes, No}\}$.
- These inputs produce a 0–100 score that maps to a **letter grade**:

A+	A	A-	B+	B	B-	C+	C	C-	D
90–100	80–89	70–79	60–69	50–59	40–49	30–39	20–29	10–19	0–9

- Each input can have a positive, negative, or no effect on the grade.
- The output you observe is the **letter grade** $y \in \{\text{A+, A, A-, B+, B, B-, C+, C, C-, D}\}$ (the 0–100 score is not displayed).
- You also see the inputs for an **11th student**, but you **do not** observe their grade **y**.

Your Task

- In each round, your task is to **guess the true grade** of the 11th student. The 10 students shown are consistent with the true input–grade relationship.
- You also see a **recommendation from a Sender**. Senders are participants from **previous study sessions**; across rounds you will see recommendations from different Senders.
  - For Senders, each dataset was labelled **UP** or **DOWN**: in **UP** datasets they were paid more the **higher** the Receiver's guess; in **DOWN** datasets they were paid more the **lower** the Receiver's guess.
  - Sometimes the Sender knew the **true relationship** and saw the true grade; sometimes not. Sometimes they only recommended a grade; sometimes they also gave a **written explanation**.

Your Payment

You are paid more the closer your guess is to the true grade of the 11th student, from **$14** (correct) down to **$5** (9 grades away). Full schedule below.

Your Guess \ True Grade	A+	A	A-	B+	B	B-	C+	C	C-	D
A+	$14	$13	$12	$11	$10	$9	$8	$7	$6	$5
A	$13	$14	$13	$12	$11	$10	$9	$8	$7	$6
A-	$12	$13	$14	$13	$12	$11	$10	$9	$8	$7
B+	$11	$12	$13	$14	$13	$12	$11	$10	$9	$8
B	$10	$11	$12	$13	$14	$13	$12	$11	$10	$9
B-	$9	$10	$11	$12	$13	$14	$13	$12	$11	$10
C+	$8	$9	$10	$11	$12	$13	$14	$13	$12	$11
C	$7	$8	$9	$10	$11	$12	$13	$14	$13	$12
C-	$6	$7	$8	$9	$10	$11	$12	$13	$14	$13
D	$5	$6	$7	$8	$9	$10	$11	$12	$13	$14

Bonus Rounds

- After the 13 main rounds, you complete **2 bonus rounds**, in which you can earn an extra $2.
- Total payment = main-round payment + bonus payment. Total payment is, therefore, between **$5 - $16**.

BG (A) — Decision Study · Instructions for RECEIVERS

Dataset Info

- There are **15 rounds** total: **13 main rounds** + **2 bonus rounds**. Each round shows a new, independently generated dataset — use only the current round's data to make your choice.
- Each dataset has **10 students**. For each student you see three inputs: $x_1, x_2 \in \{1, 2, \ldots, 10\}$ and $x_3 \in \{\text{Yes, No}\}$.
- These inputs produce a 0–100 score that maps to a **letter grade**:

A+	A	A-	B+	B	B-	C+	C	C-	D
90–100	80–89	70–79	60–69	50–59	40–49	30–39	20–29	10–19	0–9

- Each input can have a positive, negative, or no effect on the grade.
- The output you observe is the **letter grade** $y \in \{\text{A+, A, A-, B+, B, B-, C+, C, C-, D}\}$ (the 0–100 score is not displayed).
- You also see the inputs for an **11th student**, but you **do not** observe their grade **y**.

Your Task

- In each round, your task is to **guess the true grade** of the 11th student. The 10 students shown are consistent with the true input–grade relationship.
- You also see a **recommendation from a Sender**. Senders are participants from **previous study sessions**; across rounds you will see recommendations from different Senders.
  - Senders earned the **same payment as the Receiver** who saw their recommendation, so they were rewarded for helping the Receiver guess accurately.
  - Sometimes the Sender knew the **true relationship** and saw the true grade; sometimes not. Sometimes they only recommended a grade; sometimes they also gave a **written explanation**.

Your Payment

You are paid more the closer your guess is to the true grade of the 11th student, from **$14** (correct) down to **$5** (9 grades away). Full schedule below.

Your Guess \ True Grade	A+	A	A-	B+	B	B-	C+	C	C-	D
A+	$14	$13	$12	$11	$10	$9	$8	$7	$6	$5
A	$13	$14	$13	$12	$11	$10	$9	$8	$7	$6
A-	$12	$13	$14	$13	$12	$11	$10	$9	$8	$7
B+	$11	$12	$13	$14	$13	$12	$11	$10	$9	$8
B	$10	$11	$12	$13	$14	$13	$12	$11	$10	$9
B-	$9	$10	$11	$12	$13	$14	$13	$12	$11	$10
C+	$8	$9	$10	$11	$12	$13	$14	$13	$12	$11
C	$7	$8	$9	$10	$11	$12	$13	$14	$13	$12
C-	$6	$7	$8	$9	$10	$11	$12	$13	$14	$13
D	$5	$6	$7	$8	$9	$10	$11	$12	$13	$14

Bonus Rounds

- After the 13 main rounds, you complete **2 bonus rounds**, in which you can earn an extra $2.
- Total payment = main-round payment + bonus payment. Total payment is, therefore, between **$5 - $16**.