Appendix

A. Instructions and questionnaires

A.1. Instructions

Instructions to participants varied between sessions according to the different treatments in Stage 3. We present the instructions for a session with treatment r=0.25. For the other sessions, instructions were adapted accordingly and are available upon request.[1]

Instructions

General information

Thank you for participating in an experiment in which you can earn money. These earnings will be paid to you in cash at the end of the experiment. We ask you not to communicate from now on. If you have a question, then raise your hand and the instructor will come to you.

Framework of the experiment

You are 16 persons participating in this experiment. The experiment consists of 4 stages, the first one including 5 situations, the second one 10, the third one 30 and the fourth one 5. In each situation you will be randomly matched with one of the other 15 participants. You will not get to know with whom you are matched. The rules are the same for all participants. Situations are independent and in each of them, you will have to take a decision.

RULES COMMON TO ALL STAGES

Decision situation

In each situation you will be randomly matched with one of the other participants.

For each situation a number called *Z* is drawn randomly from the interval 50 to 450. This number is the same for both of you. All numbers in the interval [50, 450] have the same probability to be drawn.

When you make your decision, you will **not** know the drawn number *Z*.

However, you will be receiving two hints (numbers) on *Z*:

- You and the person with whom you are matched, both receive a common hint number *Y* for the unknown number *Z*. This common hint number is randomly selected from the interval [*Z*-10, *Z*+10]. All numbers in this interval are equally likely. This common hint number *Y* is the same for both of you.

[1] What follows is a translation (from French to English) of the instructions and the questionnaire given to the participants.

- In addition to the common hint number, each participant receives a private hint number *X* for the unknown number *Z*. The private hint numbers are also randomly selected from the interval [*Z*-10, *Z*+10]. All numbers in this interval have the same probability to be drawn. Your private hint number and the private hint number of the person whom you are matched are drawn independently from this interval, so that (in general) you will not get the same private numbers.

RULES OF THE 1ST STAGE (5 situations)

You will be asked to make a decision by choosing some number.
Your payoff positively depends on the proximity between your decision and the true value of the unknown number *Z*:

$$\textit{Payoff in ECU} = 100 - (\textit{your decision} - Z)^2 .$$

This means that your payoff only depends on how close is your decision to the true value *Z* and not on your partner's decision.
Once you have made a decision, click on the OK-button. Once all participants made their decision for the game, a situation is terminated.

RULES OF THE 2ND STAGE (10 situations)

Again, you will be asked to make a decision by choosing some number.
The rules are the same as in the first stage, but here your payoffs are given by:

$$\textit{Payoff in ECU} = 100 - (\textit{your decision} - \textit{the other participant's decision})^2 .$$

This means that your payoff only depends on how close is your action to the action of the other participant and not on the unknown number *Z*.

RULES OF THE 3RD STAGE (30 situations)

Again, you will be asked to make a decision by choosing some number.
The rules are unchanged, but here your payoffs depend positively on the one hand on the proximity between your decision and the unknown number *Z* and on the other hand on the proximity between your decision and the choice of your partner.

$$\textit{Payoff in ECU} =$$

$$400 - 3.(\textit{your decision} - Z)^2 - 1.(\textit{your decision} - \textit{the decision of the other participant})^2 .$$

This formula says that your payoff in each situation is at most 400 ECU. It is reduced for deviations of your decisions from the unknown number *Z*, and it is also reduced for deviations between your and your partner's decision. The closer is your decision to both the other participant's choice and *Z*, the higher will be your payoff.

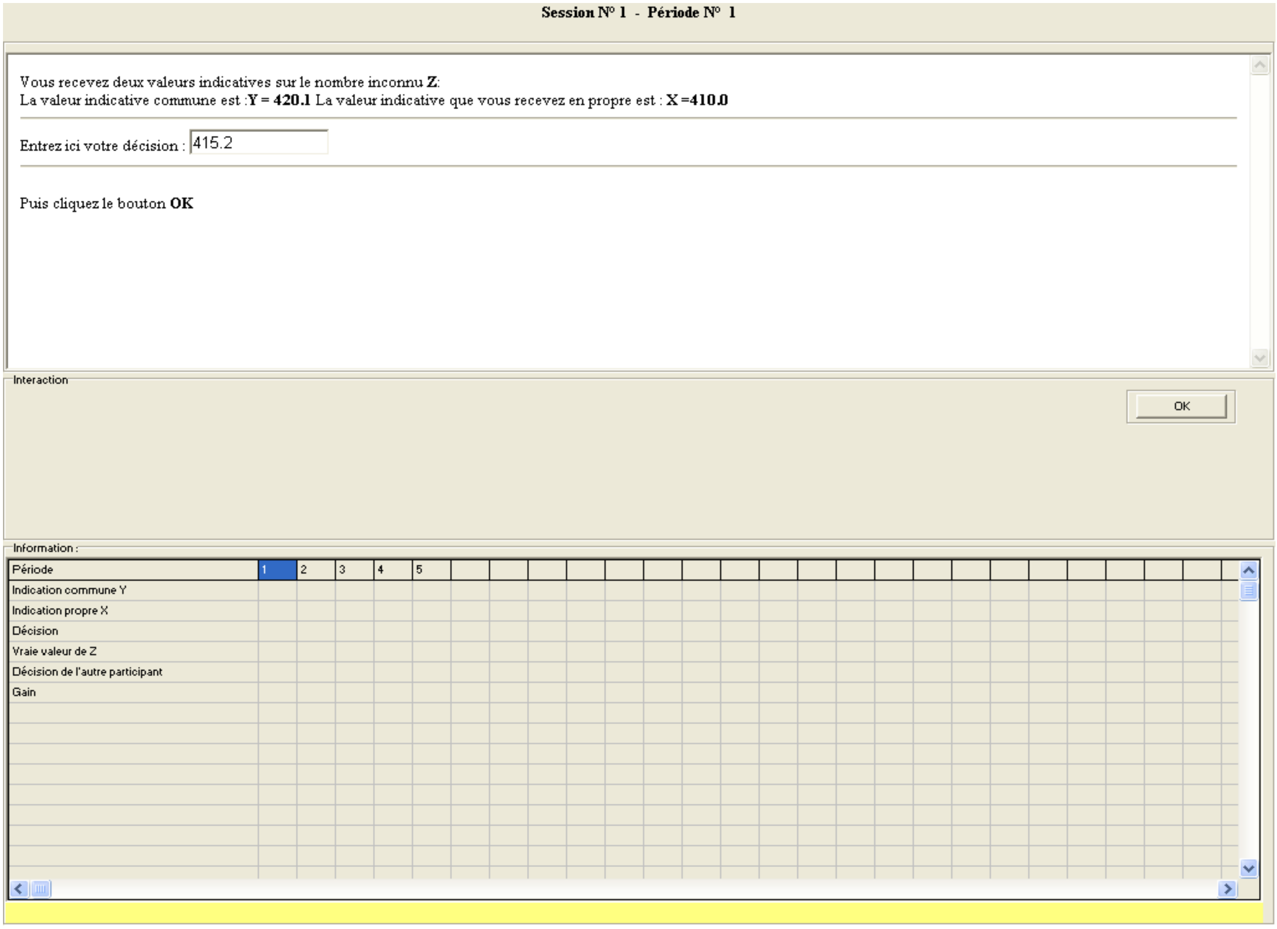
Note that your payoff depends more on the spread between your decision and *Z* than on the spread between your decision and the decision of the other participant.

Example of decision phase:

You receive two hint numbers: One hint is common for the two participants, the other is your private hint. Both hint numbers are drawn with uniform distribution from [*Z*-10, *Z*+10].

The common hint number *Y* is: 420.1.

Your private hint number *X* is: 410.0.



You enter your decision, for example here: 415.2.

Information after each situation

Each participant will be informed after each situation on

(1) the common hint *Y* and his private hint *X*,

(2) the true number *Z*, the true choice of his pair-mate,

(3) his own choice and his related payoff.

Example of information phase:

In the previous example, your own choice was 415.2.

Suppose the true value of *Z* is 419.4. Your pair-mate's decision was 413.7.

Then, your payoff will be:

Payoff = $400 - 3.(415.2 - 419.4)^2 - 1.(415.2 - 413.7)^2 = 344.83$ ECU.

RULES OF THE 4TH STAGE (5 situations)

In this stage, rules are a bit different.

Again, *Z* is unknown and you receive two hint numbers, *X* (your private hint) and *Y* (the common hint).

Now, you are asked to choose two numbers:

(1) give what you think is *Z*,

(2) give what you think the other participant think about *Z*.

Your payoff positively depends on the one hand on the proximity between your estimation on *Z* and the true value of *Z* and on the other hand between your estimation of the estimation of the other participant on *Z* and the true estimation of the other participant on *Z*.

Your payoff is given by:

$$100 - (your\ estimation\ sur\ Z - Z)^2$$
$$-(your\ estimation\ of\ the\ estimation\ of\ the\ other\ participant\ on\ Z - the\ estimation\ of\ the\ other\ participant\ on\ Z)^2$$

The closer your estimations are to true values, the higher your payoff.

You will be told about each change in stage.

Questionnaires:

At the beginning of the experiment, you will be asked to fill in an understanding questionnaire on a paper. Afterward, the experiment will begin. At the end of the experiment you will fill in a "personal" questionnaire on the computer. All information will remain secret.

Payoffs:

Also at the end of the experiment the ECUs you have obtained are converted into Euros and paid in cash. 1 ECU corresponds to 0.25 Cents.

If you have any questions, please ask them at this time.

Thanks for your participation!

A.2. Understanding and training questionnaire

Fill in

- In each situation, you interact with ________ other participant(s).

- You receive in each situation ________ hints.

- The difference between the unknown number *Z* and any hint is at most _______.

Yes or no

- At stage 3, when a participant makes a decision, does his payoff depend on the decision of his pair-mate? ____

- Do pair-mates receive the same hints? _____

- Is there a hint that is more precise than another? ______

- Do you play with the same participant during the whole length of the experiment? _____

Practice

You are at the third stage of the experiment.

You receive *Y*=135 and *X*=141.

Among the next statements, choose the right one(s):

- The true value of *Z* is between 125 and 151.
- The true value of *Z* is 135.
- The true value of *Z* is between 131 and 145.

Suppose that the true value of *Z* is 143 and that the true decision of your pare-mate is 133. What is your payoff (in ECU) if you choose 134?

Now, suppose again that the true value of *Z* is 143 and the true decision of your paire-mate is 133. What is your payoff (in ECU) if you choose 138?

A.3. Post-experimental questionnaire

1. How did you make a decision? On which criteria?

2. During the first 3 stages, have you tried to guess the value of *Z*? And the value of the decision of the other participant?
3. Do you think that one of the two indicative hints (private *versus* common) was more informative than the other on *Z*? And on the decision of the other participant?
4. Did you take into account the two indicative hints in the same manner? Or more your private hint? Or more the common hint?

B. Control sessions

B.1. Design of control sessions

We conducted 6 control sessions. Table B1 gives an overview over control sessions.

<table>
<tr><th>Sessions</th><th>Stage 1</th><th>Stage 2</th><th>Stage 3</th><th>Stage 4</th><th>Stage 5</th></tr>
<tr><td>19 to 21</td><td rowspan="2">“r=0”
(5 periods)</td><td rowspan="2">belief elicitation
(5 periods)</td><td>Treatment “r=1 with 1 common signal”
(10 periods)</td><td rowspan="2">“r=0.5”
(20 periods)</td><td rowspan="2">“r=1 with 2 signals”
(10 periods)</td></tr>
<tr><td>22 to 24</td><td>Treatment “r=1 with 1 private signal”
(10 periods)</td></tr>
</table>

Table B1 – Composition of the different control sessions

Control sessions are motivated by the results from standard sessions presented in the main text. We especially wanted to elicit higher order beliefs at an earlier stage before subjects were playing strategic games. We also wanted to test directly how public or private signals affect payoffs in a pure coordination game. In control sessions, we started with a first stage where “*r*=0” as in standard sessions. In Stage 2, we proceeded with directly eliciting beliefs (as we did in the last stage of standard sessions). Here, we used two separate payoff functions, $100-(e_i(\theta)-\theta)^2$ and $100-(e_i(e_j(\theta)-e_j(\theta))^2$ instead of one unified as in standard sessions. We did these changes, because we were concerned that results from the last stage in standard sessions were distorted by spillovers from the previous stages and because the game might be easier to understand with separate payoff functions.[2] In Stage 3, we had two distinct treatments: “*r*=1 with 1 common signal” and “*r*=1 with 1 private signal”; both treatments had the same payoff function as in Stage 5, where subjects got a private *and* a common signal. In Treatment “*r*=1 with 1 common signal” they received just a common and no private signal, in Treatment “*r*=1 with 1 private signal” they received just private signals. Here, they could choose any number in $[x_i - 20, x_i + 20]$, but the largest observed difference between choice and signal was 14.2. We used these treatments to analyze subjects’ ability to coordinate when they receive only one signal and to compare the welfare effects of adding a second

[2] In addition, Stage 1 (“*r*=0”) and Stage 2 (belief elicitation) are both tests of Bayesian rationality, though with a different framing. Proceeding with these stages one directly after another allows for a better comparison of their results.

signal in Stage 5. Stage 4 ("*r*=0.5") contained the same game as Treatment "*r*=0.5" in standard sessions and was meant to check robustness of the results. In the control sessions, we divided the 16 participants into 2 matching groups of 8, thereby gathering two independent observations per session.[3] 100 points were converted to 40 cents in sessions 19 – 24.[4]

B.2. Analyzing data from control sessions

As for standard sessions, we use the following structure in analyzing data. First, we analyze whether subjects' choices are between their two signals as predicted by the equilibrium theory. Then we give summary statistics for the weights that subjects attach to the private signal. We investigate whether observed weights on the private signal are positive, smaller than 0.5, and whether they deviate systematically from the theoretical prediction. We also test whether there is a time trend and we test comparative static predictions arising from theory.

B.2.1. Some considerations about rationality

As for standard sessions, we have a look at subjects' rationality. The next table displays the percentage of all choices within the intervals I_1 to I_3 as defined in the main text. Counting the number of choices that are closer to the common signal, closer to the private, or in the middle provides us a crude first impression of whether subjects put a larger weight on the common signal and how dispersed the distribution of relative weights is.

Value of r	*0*	*0.5*	*1*
Inside $I_1 = [\min(y, x_i), \max(y, x_i)]$	*75%*	*86%*	*94%*
Choice = y	*1%*	*4%*	*61%*
Closer to y	*28%*	*35%*	*15%*
Middle (+/-0.05)	*15%*	*18%*	*6%*
Closer to x_i	*31%*	*28%*	*12%*
Choice = x_i	*0%*	*1%*	*1%*
Outside I_1 *beyond y*	*11%*	*7%*	*3%*
Outside I_1 *beyond* x_i	*14%*	*7%*	*3%*
Inside $I_2 = [\max(Y, X) - 10; \min(y, X) + 10]$	*93%*	*97%*	*84%*
Inside $I_3 = I_1 \cup I_2$	*96%*	*99%*	*99.6%*

Table B2.1 – Crude classification of choices, control sessions 19-24

Comparing control sessions with standard sessions, the only remarkable difference is that in Stage 5 ("*r*=1") of control sessions there are many more choices equal to the public signal than in Stage 2 ("*r*=1") of standard sessions. We attribute this to the different order of stages. Experience from other stages seems to make it easier for subjects to coordinate on the public signal when coordination is the only motive for action. Result 1 still holds for control sessions. In Appendix E below we provide some more details on order effects.

[3] Subjects were randomly matched with other subjects from the same matching group only. Subjects were not informed about the size of matching groups.

[4] In sessions 22-24, 15 of 48 subjects earned negative payoffs in Treatment "*r*=1 with 1 private signal". Here, it took up to 8 periods to compensate these losses.

B.2.2. Estimated weights on the private signal

Session, group	"*r* = 0"	"*r* = 1 with 2 signals", 1st half	"*r* = 1 with 2 signals", 2nd half	"*r* = 0.5", 1st half	"*r* = 0.5", 2nd half
19, group 1	0.455	0.181	0.107	0.502	0.484
19, group 2	0.583	0.149	0.201	0.367	0.338
20, group 1	0.524	0.008	0.000	0.431	0.473
20, group 2	0.583	0.255	0.278	0.551	0.536
21, group 1	0.446	0.234	0.229	0.476	0.422
21, group 2	0.534	0.134	0.039	0.495	0.540
22, group 1	0.527	0.130	0.105	0.439	0.454
22, group 2	0.500	0.482	0.421	0.530	0.485
23, group 1	0.506	0.098	0.150	0.459	0.455
23, group 2	0.570	0.184	0.067	0.486	0.459
24, group 1	0.582	0.132	0.160	0.489	0.440
24, group 2	0.489	0.385	0.359	0.519	0.487
Average (19-24)	0.525	0.198	0.176	0.479	0.464
St.dev. (19-24)	0.048	0.129	0.128	0.050	0.053
Equilibrium weight	0.5	0	0	0.333	0.333

Table B2.2 – Group specific weights on the private signal

Table B2.2 displays the estimated group-specific weights on the private signal for control sessions. Estimates follow the same procedure as laid out in Section 4.2 for standard sessions. We tested standard sessions and control sessions separately, because they are not entirely comparable. Nonparametric tests for control sessions are reported below. We show that most results of standard sessions also hold for control sessions. We comment on the similarities and differences below.

Result 2 for control sessions also holds: for $r = 0$, subjects put an equal weight on both signals consistent with Bayesian rationality. In Stage 1 ("*r*=0"), there is no significant difference between group specific estimated weights and 0.5. Two-tailed Wilcoxon matched pairs signed rank tests yield p-values of 0.12 for control sessions. A Mann-Whitney test does not reject the hypothesis that weights in standard sessions have the same distribution as those in control sessions (p-value 0.17). Since individual decisions are independent in this stage, we also performed tests on individual weights from Regression (i). The hypothesis that individual weights are distributed around 0.5 cannot be rejected (the p-value is 0.45 for control sessions). Neither can the hypothesis be rejected that individual weights in the two groups of sessions come from the same distribution (p-value 0.65).

Result 3 also holds for control sessions: for $r = 0.5$, subjects tend to put larger weights on public than on private signals, but the difference is smaller than theoretically predicted. There is no trend towards equilibrium. One-tailed Wilcoxon matched pairs signed rank tests reject that the weight on the private signal equals the equilibrium value in favor of a higher weight ($p < 0.01$). For data from the second half of Stage 4 ("*r*=0.5") we can reject that weights equal 0.5 in favor of smaller weights ($p = 0.013$). For the first half of Stage 4 ("*r*=0.5"), however, this hypothesis cannot be rejected ($p = 0.100$). For Stage 4, estimated weights are higher than for Treatment *r*=0.5 in standard sessions (significant at 1% for the first half, but insignificant (p=7.7%) for the second half, using two-tailed Mann-Whitney tests). We attribute this to an order effect. We provide more details about order effects in Appendix E below.

Result 4 from the main text has to be slightly adjusted for control sessions. Indeed, for "r = 1, two signals", subjects assign larger weights to public than to private information ($p<1\%$) as in standard sessions. But contrary to standard sessions, in control sessions, there seems to be no trend towards improved coordination. Testing whether the weights assigned to private signals in the last 5 rounds were equal to those in the first half against the alternative of a systematic trend towards equilibrium (i.e. lower weights in the last rounds), the one-tailed Wilcoxon matched pairs test yields p-value of 12.8% for control sessions. Thus, we cannot reject the hypothesis of no trend towards equilibrium for control sessions. The latter becomes even more pronounced if we use a sign test instead (p-value 38%). Comparing behavior between the two groups of sessions, we find that in the first 5 rounds, weights in control sessions are significantly lower than in standard sessions (two-tailed Mann-Whitney, $p<1\%$). In the second half, there is no significant difference ($p=12\%$). We attribute this to an order effect explained below in Appendix E. We conclude that in the extreme case of a pure coordination game, subjects condition their choices on their private signals, which prevents full coordination. Private signals matter in a pure coordination game and may be welfare reducing. One of the reasons for control sessions was to include games with only one signal, in order to test the welfare reducing effect of private signals directly by comparing payoffs. These results are reported in Appendix B.3. below.

Finally, Result 5 holds also for control sessions: over all sessions, the weight assigned to the private signal tends to decrease in r as predicted by theory. A one-tailed Wilcoxon matched pairs signed rank finds that γ is significantly smaller in treatments with higher r, if we compare "r=0" with "r=0.5" or "r=0.5" with "r=1 with 2 signals". P-values are always below 4%.

B.3. Payoff effects of providing additional information in a pure coordination game

In a pure coordination game as with r=1, private signals should be neglected if actions can be conditioned on public signals. In equilibrium, additional private signals should not affect welfare. One purpose of control sessions was to provide a direct answer to the question of whether providing additional private signals reduces welfare. Table B3 compares average payoffs from games with "r=1".

In sessions 19-21, four groups achieved a lower payoff in with two signals than with one common signal. Hence, adding a private signal to the information structure reduced their average payoffs. Two groups (Session 20, Group 1, and Session 21, Group 2) improved their payoffs. The average over all groups is lower when both signals are provided, but the difference is not significant ($p=0.56$). This indicates that adding private information does not reduce average payoffs in pure coordination games. On the other hand, average payoffs for Stage "r=1" in standard sessions are significantly lower than in the comparable Stage "r=1 with 2 signals" and in Treatment "r=1 with 1 common signal" of Sessions 19-21 (Mann-Whitney, $p<1\%$).

We attribute these diverse findings to the order of stages: "*r*=1" (with both signals) was the second stage in standard sessions, while it was the last in control sessions. In control sessions, it is possible that the high payoff when "*r*=1 with 2 signals" is due to learning from previous stages. Comparing data from Stage "*r*=1" of standard sessions with data from "*r*=1 with 1 common signal" from Sessions 19-21 might give a better impression of how additional private information affects behavior for a given state of experience with this kind of games. However, we have to admit that we cannot provide a definite answer to the question of whether theoretically irrelevant private information impedes coordination. An experiment by Fehr et al. (2011) is better suited to answer this question and finds convincing evidence for welfare reducing effects of intrinsically irrelevant private signals in a pure coordination game.[5]

Session, Group	*"r=1, with 1 common signal"*	*"r=1, with 1 private signal"*	*"r=1, with 2 signals"*	*Session*	*"r=1" (two signals)*
19, Group 1	97.53		89.53	1	86.22
Group 2	99.60		85.08	2	78.84
20, Group 1	99.82		99.90	3	71.26
Group 2	93.06		75.46	4	73.01
21, Group 1	92.49		88.99	5	82.81
Group 2	72.25		89.98	6	76.59
Average (19-21)	92.46		88.16	7	65.47
22, Group 1		- 3.56	90.06	8	81.68
Group 2		25.19	82.09	9	70.07
23, Group 1		24.23	95.06	10	68.59
Group 2		- 1.60	94.53	11	78.22
24, Group 1		8.73	89.33	12	88.29
Group 2		7.47	78.94	13	80.16
				14	70.74
				15	68.69
				16	63.30
				17	83.33
				18	82.16
Average (22-24)		10.08	88.33	Average (1-18)	76.08

Table B3 – Average payoffs in games with"*r*=1"

From Table B3, it is obvious for Sessions 22-24 that adding a public signal increased average payoffs compared to treatment "*r*=1 with 1 private signal".[6]

[5] In a context of asymmetric information, Camerer et al. (1989) show that more information is not always better because agents are unable to ignore private information even when it is in their interest to do so.

[6] When receiving only a private signal in Treatment "*r*=1 with 1 private signal", some subjects apparently tried to find a focal point that the experimental design did not allow for. About one participant per session asked why he could not enter either 50 or 0. He or she was told that these numbers were outside the range of admissible choices. In Sessions 19 to 21 this question never occurred, probably because subjects could coordinate on the public signal. Having allowed participants to choose from a fixed range of numbers might have helped them to coordinate on a focal point, disregard the private signal, and increase their payoffs. In a related experiment, Fehr et al. (2011) show that subjects tend to neglect imprecise private signals in a pure coordination game with a fixed choice set that includes prominent numbers. However, they also find that adding private signals impedes coordination.

B.4. Cognitive hierarchy in control sessions

As explained in the main text, we define levels of reasoning by assuming that level-0 players decide randomly and level-1 players place weight $\gamma_1 = 0.5$ on the private signal, so that they ignore the strategic part of the payoff function. Higher levels of reasoning are then given by

$$\gamma_{k+1} = \frac{(1-r) + r\gamma_k}{2}.$$

Comparing weights from limited levels of reasoning displayed in Table 6 in the main text with the detected weights in Table B2.2, shows that for "*r*=0.5", average group weights are higher than the weights from level-2 reasoning with only one exception (Session 19, Group 2).

For "*r*=1 with 2 signals", most groups achieve weights that indicate less than level-2 reasoning, 5 groups go beyond Level 3, and one of them fully coordinates on the public signal (Session 20, Group 1). Here, the level-2 hypothesis cannot be rejected (p-value 0.09). Level-3 reasoning cannot be rejected either (p=0.26), and we need to go to level 4 to find a significant difference.

C. Non-Bayesian higher-order beliefs

As explained in the main text, weights on the private signal above equilibrium values can also be explained by systematic mistakes in the formation of higher-order beliefs. Here, we provide a formal model of non-Bayesian higher-order beliefs and test it with the data from the experiment.

In the experiment we elicit higher order beliefs directly and measure the weight that subjects attach to the private signal when estimating their partner's beliefs. Denote this weight by λ. Based on this measure, we test a model of boundedly rational behavior that assumes infinite levels of reasoning but a systematic error in forming higher-order beliefs. More precisely, denote subject *i*'s expectation about a variable *x* by $e_i(x)$ and suppose:

1. Subjects respond optimally to their expectations about the state and about their partner's action, i.e.,

   $$a_i = (1-r)\,e_i(\theta) + r\,e_i(a_j), \quad j \neq i.$$

   Furthermore, this behavioral assumption is common knowledge among players, which amounts to assuming infinite levels of reasoning.

2. Subjects use private and common signals correctly to forecast the state of the world, *i.e.*

   $$e_i(\theta) = E_i(\theta).$$

3. Subjects make a systematic error in forming higher-order beliefs, such that

   $$e_i(e_j(\theta)) = \lambda\, x_i + (1-\lambda)\,y,$$

   $$e_i(e_j(e_i(\theta))) = \lambda^{3/2}\, x_i + (1-\lambda^{3/2})\,y,$$

$$e_i(e_j(e_i(e_j(\theta)))) = \lambda^2\, x_i + (1-\lambda^2)\, y,$$

and so on, where λ is the weight measured in the experiment.

To understand the justification of these formulas, note that a rational first-order expectations are $E_i(\theta) = \alpha\, x_i + (1-\alpha)\, y$ with $\alpha = 0.5$. Rational higher-order expectations are then given by $E_i(E_j(\theta)) = \alpha^2\, x_i + (1-\alpha^2)\, y$, $E_i(E_j(E_i(\theta))) = \alpha^3\, x_i + (1-\alpha^3)\, y$, and so on. The estimated λ for second-order expectations replaces α^2 in the formula, so that the weight in third-order beliefs should be $\lambda^{3/2}$ and so on.

Combining these assumptions, we get

$$\begin{aligned}
a_i &= (1-r)\, e_i(\theta) + r\, e_i(a_j) \\
&= (1-r)\, E_i(\theta) + r\, e_i((1-r)\, e_j(\theta) + r\, e_j(a_i)) \\
&= (1-r)\, E_i(\theta) + (1-r)\left[r\left(\lambda x_i + (1-\lambda) y\right) + r^2\left(\lambda^{3/2} x_i + (1-\lambda^{3/2}) y\right) + r^3\left(\lambda^2 x_i + (1-\lambda^2) y\right) + r^4\left(\lambda^{5/2} x_i + (1-\lambda^{5/2}) y\right) + \ldots\right] \\
&= (1-r)\left(\frac{1}{2} + r\lambda \sum_{i=0}^{\infty} (r\sqrt{\lambda})^i\right) x_i + \left[1-(1-r)\left(\frac{1}{2} + r\lambda \sum_{i=0}^{\infty} (r\sqrt{\lambda})^i\right)\right] y \\
&= (1-r)\left(\frac{1}{2} + \frac{r\lambda}{1-r\sqrt{\lambda}}\right) x_i + \left[1-(1-r)\left(\frac{1}{2} + \frac{r\lambda}{1-r\sqrt{\lambda}}\right)\right] y. \qquad (1)
\end{aligned}$$

With λ decreasing from 0.5 to its rational value of 0.25, the weight that an agent attaches to the private signal in her decision decreases from a value below 0.5 towards the equilibrium value.

In Stage 2 of control sessions and Stage 4 of standard sessions, subjects were provided with public and private signals, but instead of being asked for an action, they had to state their beliefs (i) about the true state of the world and (ii) about their partner's stated belief about the state of the world. Thus, we directly elicit first-order and second-order beliefs. As explained in the main text, subjects should put a weight of 0.5 on x_i in estimating θ and a weight of 0.25 in estimating the other's estimation of θ.

If a subject attributes a higher [lower] weight to her private signals in predicting her partner's stated expectation, her weight on the private signals in Treatments "r=0.5", "r=0.25", and "r=0.75" should also be higher [lower] than the respective equilibrium weights. Suppose that a subject puts weight 0.312 on private signals when "r=0.75". Such behavior is consistent with level-2 reasoning, but could also be explained by a weight of λ=0.48 on private signals in higher-order beliefs according to equation (1). Direct elicitation of beliefs allows discriminating between the two theories.

Data from belief elicitation in standard sessions, summarized in Table C1, reveal that subjects attached (on average) a weight lower than 0.5 on the private signal when estimating θ and a weight higher than 0.25 when estimating their partner's estimation of θ.

For the rational second-order beliefs we assumed that first-order beliefs put weight 0.5 on either signal. In fact, data reveal that on average subjects attribute a weight of only 0.46 on their private signals. If subjects actually guessed this weight correctly, the optimal weight on the private signal in higher-order beliefs would be 0.23. Thus, the higher observed weights in second-order beliefs cannot be explained by subjects' responding to distorted first-order beliefs.

Since estimating θ does not depend on others' choices, we also test the distribution of individual weights. When we estimate weights separately for each subject, the average weight on x_i (over all subjects in standard sessions) in estimating θ is 0.471, while the average weight on x_i in forming higher-order beliefs is 0.292. We can reject the hypotheses that individual weights in estimating θ are distributed around 0.5 (p<1%) and that the weights in estimating the other player's estimation of the state are around 0.25 (p<1%). 229 out of 288 subjects attribute a weight higher than 0.25 to the private signal in this task. Individuals' higher-order expectations are not independent, but since expectations of θ are generally biased towards the public signal, any adjustment in higher-order beliefs should go in the same direction, which conflicts with our observations. Although Bayesian rationality requires that the second weight be half of the first, it is in fact 62.2% of the first. Hence, all evidence indicates that subjects underestimate the importance of public information in forming higher-order beliefs.

Sessions	*Weight in estimating θ*	*Weight in estimating the partner's estimation of θ*
1	0.394	0.228
2	0.520	0.342
3	0.481	0.376
4	0.418	0.170
5	0.424	0.210
6	0.482	0.159
7	0.493	0.335
8	0.429	0.301
9	0.532	0.379
10	0.466	0.272
11	0.500	0.375
12	0.501	0.217
13	0.439	0.271
14	0.432	0.263
15	0.520	0.420
16	0.425	0.345
17	0.437	0.277
18	0.446	0.292
Average	0.463	0.291
St.dev.	0.041	0.075

Table C1 – Group specific weights on the private signal in stated expectations of Sessions 1-18

It is surprising, though, that subjects underused private signals in forming their expectation of θ, especially in light of the results from Treatment "r=0", where subjects used (on average) a weight of 0.506. The difference in coefficients betweem Stage 1 "r=0" and stated first-order expectations in Stage 4 is significant at 0.1%. This may be caused by an order effect, such that after 40 periods in stages 2 and 3, in which the public signal was more important than the private, subjects underestimate the importance of the private signal for estimating θ during Stage 4. This, however, should also hold for forming expectations about others' expectations. Thus, without an order effect, we should see larger weights on the private signal in both belief-elicitation tasks.

Testing this hypothesis was the motivation for asking participants to form expectations in the second stage of the control sessions. Here, we elicited beliefs directly after Stage "*r*=0" and before all those stages, in which common signals are theoretically more important than private ones. Table C2 summarizes the results from Stage 2 of control sessions.

The result is striking: the bias in estimating the state is almost absent now (p=0.12), and the average weight on the private signal in estimating the other subject's estimation of the state is higher than 0.25 in all groups (p<1%). Average *individual* weights were 0.52 in estimating θ and 0.437 in second-order beliefs. Individual weights are distributed around 0.5 (p=0.37), while the difference in the weight for higher-order beliefs from 0.25 is significant at the 0.1% level.

Session, group	*Weight in estimating θ*	*Weight in estimating the partner's estimation of θ*
19, Group 1	0.524	0.475
Group 2	0.417	0.389
20, Group 1	0.463	0.358
Group 2	0.538	0.550
21, Group 1	0.543	0.459
Group 2	0.447	0.317
22, Group 1	0.459	0.388
Group 2	0.555	0.499
23, Group1	0.522	0.397
Group 2	0.603	0.399
24, Group1	0.511	0.317
Group 2	0.566	0.584
Average	0.512	0.428
St.dev	0.055	0.086

Table C2 – Group specific weights on the private signal in stated expectations of Sessions 19-24

Having ruled out order effects and best responses to others' deviations from rational first-order beliefs as possible explanations leaves us with the impression that subjects underestimate how informative the public signal is for predicting others' expectations. This can be viewed as a systematic error in forming second-order beliefs and provides an alternative explanation for results from the other stages. To our knowledge, we are the first to elicit higher-order beliefs directly and relate them to controlled information. Non-Bayesian higher-order beliefs may also be responsible for observed deviations from equilibrium in other experiments. Möbius et al. (2011) found gender differences in Bayesian updating of self assessments; we could not find a significant gender difference in stated beliefs.

Next, we turn to testing the aforementioned model of non-Bayesian higher-order beliefs. We argue that systematic errors in higher-order beliefs are too low to explain the observed deviations from equilibrium in games with an interior *r*.

For each standard session, we use the observed average weight for second-order beliefs to calculate the weight that this group should attach on the private signal in Treatments "*r*=0.5", "*r*=0.25", or "*r*=0.75" according to the model presented above. As we elicited beliefs directly after these treatments, we may expect that higher-order beliefs in their later periods are formed in about the same way as elicited beliefs. Table C3 compares the results of this calculation with the estimated weights on x_i in the second half of Treatments “*r*=0.5”, “*r*=0.25”, or “*r*=0.75”. With one exception (Session 17), the estimated weights on the private signal are higher than those that would be explained by our model of Non-Bayesian higher-order beliefs. This result stands if we use individual data instead of group averages, where the differences between estimated weights in Treatments “*r*=0.5”, “*r*=0.25”, or “*r*=0.75” and weights calculated from our model with estimated weights for higher-order beliefs as input is significant at the 0.1% level.

Session	Calculated weight on private signal using estimated error in higher-order beliefs	Estimated weight on private signal (*r*=0.5/ *r*=0.25/ *r*=0.75, 2nd half)
1	0.325	0.408
2	0.371	0.463
3	0.386	0.447
4	0.303	0.475
5	0.318	0.393
6	0.300	0.453
7	0.448	0.492
8	0.440	0.466
9	0.459	0.500
10	0.434	0.460
11	0.458	0.475
12	0.421	0.473
13	0.208	0.391
14	0.205	0.294
15	0.278	0.457
16	0.241	0.377
17	0.211	0.187
18	0.217	0.338

Table C3 – Comparing weights from a model of Non-Bayesian higher-order beliefs with observations

Thus, we conclude that errors in forming higher-order beliefs are in general not sufficiently strong to explain observed behavior in the game. There must be another form of irrationality in addition to non-Bayesian beliefs.[7] Hence, we cannot reject the hypothesis of limited levels of reasoning. Combining levels of reasoning with non-Bayesian beliefs yields higher estimated levels of reasoning.[8] However, for interior values of *r*, aggregate behavior is in most cases consistent with levels of reasoning not exceeding degree 2, even if we account for the observed systematic mistakes in forming higher-order beliefs. This is laid out in Appendix D.

[7] A third possible explanation that cannot be dealt with in this paper was pointed out by a discussant: subjects might pay less attention to the actions of others, because they have less information about them. Strategic uncertainty turns beliefs about others’ behavior into an ambiguous guess as opposed to estimating the fundamental state for which probabilistic information is available.

[8] We are grateful to Gabriel Desgranges for asking us how non-Bayesian beliefs affect the result on levels of reasoning.

D. Combining non-Bayesian higher-order beliefs with limited levels of reasoning

What is the level of reasoning, if we account for systematic mistakes in forming higher-order beliefs? Combining the model of Appendix C with limited levels of reasoning yields the following weights on private signals denoted by $\hat{\gamma}_k$:

Level 1: $a_i^1 = e_i(\theta) = 0.5(x_i + y)$ => $\hat{\gamma}_1 = 0.5$.

Level 2: $a_i^2 = (1-r)e_i(\theta) + re_i(a_j^1) = (1-r)E_i(\theta) + re_i(e_j(\theta))$

$$= (1-r)E_i(\theta) + r[\lambda_i x_i + (1-\lambda_i)y]$$

$$= [0.5(1-r) + r\lambda_i]x_i + [0.5(1+r) - r\lambda_i]y \quad \Rightarrow \quad \hat{\gamma}_2 = 0.5(1-r) + r\lambda.$$

Level 3: $a_i^3 = (1-r)e_i(\theta) + re_i(a_j^2) = (1-r)E_i(\theta) + re_i\left[(1-r)e_j(\theta) + re_j(a_i^1)\right]$

$$= (1-r)E_i(\theta) + (1-r)re_i(e_j(\theta)) + r^2 e_i(e_j(e_i(\theta)))$$

$$= (1-r)E_i(\theta) + (1-r)r[\lambda_i x_i + (1-\lambda_i)y] + r^2\left[\lambda_i^{3/2} x_i + (1-\lambda_i^{3/2})y\right]$$

$$= \left[(1-r)[0.5 + r\lambda_i] + r^2\lambda_i^{3/2}\right]x_i + \left(1 - \left[(1-r)[0.5 + r\lambda_i] + r^2\lambda_i^{3/2}\right]\right)y$$

$$\Rightarrow \hat{\gamma}_3 = (1-r)[0.5 + r\lambda_i] + r^2\lambda_i^{3/2}.$$

Level 4: $a_i^4 = (1-r)e_i(\theta) + re_i(a_j^3) = (1-r)E_i(\theta) + re_i\left[(1-r)e_j(\theta) + re_j(a_i^2)\right]$

$$= (1-r)E_i(\theta) + (1-r)re_i(e_j(\theta)) + r^2 e_i(e_j(\left[1-r)e_i(\theta) + re_i(a_j^1)\right]))$$

$$= (1-r)E_i(\theta) + (1-r)\left[r[\lambda_i x_i + (1-\lambda_i)y] + r^2\left[\lambda_i^{3/2} x_i + (1-\lambda_i^{3/2})y\right]\right] + r^3\left[\lambda_i^2 x_i + (1-\lambda_i^2)\right]$$

$$= \left[(1-r)\left[0.5 + r\lambda_i + r^2\lambda_i^{3/2}\right] + r^3\lambda_i^2\right]x_i + \left(1 - \left[(1-r)\left[0.5 + r\lambda_i + r^2\lambda_i^{3/2}\right] + r^3\lambda_i^2\right]\right)y$$

$$\Rightarrow \hat{\gamma}_4 = (1-r)\left[0.5 + r\lambda_i + r^2\lambda_i^{3/2}\right] + r^3\lambda_i^2.$$

Level $k \geq 3$: $\hat{\lambda}_k = (1-r)\left[0.5 + r\lambda\sum_{j=0}^{k-3}(r\sqrt{\lambda})^j\right] + r^{k-1}\lambda_i^{k/2}$.

Table D1 compares the estimated weights in the second half of treatments with $0 < r < 1$ with those that subjects should put on the private signal according to the model combining limited levels of reasoning with non-Bayesian higher-order beliefs. In Treatments “*r*=0.5” and “*r*=0.25” (Sessions 1-12), the estimated weights are all higher than those arising from non-Bayesian beliefs and level-2 reasoning. When “*r*=0.75” (Sessions 13-18), however, in 3 out of 6 cases estimated weights are between those of level 2 and level 3, in 2 cases estimated weights are between those of level 1 and 2, and for Session 17, the estimated weight on the private signal is even smaller than the equilibrium weight.

Session, group	Estimated weight on private signal (r=0.5/ r=0.25/ r=0.75, 2nd half)	Calculated weights in the model combining estimated errors in higher-order beliefs with limited levels of reasoning				
		Level 1	Level 2	Level 3	Level 4	Equilibrium
1	0.408	0.5	0.364	0.334	0.327	0.325
2	0.463	0.5	0.421	0.386	0.375	0.371
3	0.447	0.5	0.438	0.402	0.391	0.386
4	0.475	0.5	0.335	0.310	0.305	0.303
5	0.393	0.5	0.355	0.326	0.320	0.318
6	0.453	0.5	0.330	0.306	0.301	0.300
7	0.492	0.5	0.459	0.450	0.449	0.448
8	0.466	0.5	0.450	0.442	0.441	0.440
9	0.500	0.5	0.470	0.461	0.459	0.459
10	0.460	0.5	0.443	0.435	0.434	0.434
11	0.475	0.5	0.469	0.460	0.458	0.458
12	0.473	0.5	0.429	0.422	0.421	0.421
13	0.391	0.5	0.328	0.255	0.227	0.208
14	0.294	0.5	0.322	0.250	0.222	0.205
15	0.457	0.5	0.440	0.357	0.316	0.278
16	0.377	0.5	0.384	0.304	0.268	0.241
17	0.187	0.5	0.333	0.259	0.230	0.211
18	0.338	0.5	0.344	0.269	0.238	0.217

Table D1 – Comparing weights from a model of Non-Bayesian higher-order beliefs and limited levels of reasoning with observations. Underlined numbers indicate cases where data are consistent with an application of higher levels of reasoning.

E. Order effects

The data indicate order effects. During the course of the experiment, subjects seem to learn that public signals are more important than private ones for estimating the likely action of other participants. In particular:

1. For "r=0.5", the estimated weights for the private signal are higher in Sessions 19-24 than in Sessions 1-6. The difference is significant (p=1%) for the fist half of the stage where "r=0.5", but insignificant (p=7.7%) for the second half, using two-tailed Mann-Whitney tests. The difference could result from the different stages that subjects went through before they reached "r=0.5".

2. When "r=1" is conducted in an early stage (as in Sessions 1-18), the weight on public information significantly increases in the second half of the treatment (Result 4), which is not the case in Sessions 19-24 with "r=1" in the last stage. Furthermore, when "r=1" is conducted in as last stage, subjects assign a larger weight to public signals from the start compared to Stage "r=1" in the standard sessions (first half: p<1%, second half: p=0.12, two-tailed Mann-Whitney). In consequence of the larger weight on public signals, average payoffs are higher for "r=1 with 2 signals" in control sessions than in standard sessions (p<1%). These results indicate that learning has not settled within the 10 periods of this treatment. This is in line with

Fehr et al. (2011), who also find that the convergence process to an equilibrium in pure coordination games with extrinsic private and public signals takes surprisingly long.

3. Belief elicitation: When asked for their expectation of the fundamental state θ directly after "*r*=0" (Sessions 19-24), subjects assign equal weights to both signals (as in Stage "*r*=0"). When asked after all other stages in Sessions 1-18, subjects assign a larger weight to the public signal. The difference between stating expectations at the end of the experiment in standard sessions and at Stage 2 of the experiment in control sessions is significant with p=1% for group data and 0.5% for individual data (two-tailed Mann-Whitney). In Sessions 1-18, the difference between the weights in stated expectations in the last 5 periods and that put in Treatment "*r*=0" is significant (p=0.1%, two-tailed Wilcoxon matched pairs) for group data as well as for individual data.
4. In the formation of higher-order beliefs, the weight on the public signal is significantly larger when beliefs are elicited after all other stages (Sessions 1-18) as compared to when they are elicited directly after "*r*=0" (Sessions 19-24). The difference is significant for both, group and individual data (p<1%, two-tailed Mann-Whitney).
5. Treatments "*r*=1 with 1 common signal" and "*r*=1 with 1 private signal" do not seem to have different effects on subsequent behavior. Two-tailed Mann-Whitney tests cannot reject the hypotheses that weights in subsequent stages of Sessions 19-21 come from the same distribution as in Sessions 22-24 (p-values are above 0.5).

The first four points mentioned here indicate that individual weights are adjusted gradually when a new stage starts. Subjects seem to start a stage with weights closer to the final weights of the previous stage, which is not surprising. Our tests indicate, however, that 30 periods are sufficient to undo these order effects. Since our analysis focuses on treatments with $0 < r < 1$ that all had 30 periods, we have no reasons to believe that tests on data from the second half of these treatments are affected by order effects.

References for the appendix

Camerer, C. F., G. Loewenstein, and M. Weber (1989), "The Curse of Knowledge in Economic Settings: An Experimental Analysis," *Journal of Political Economy*, 97, 1232-54.

Fehr, D., F. Heinemann, and A. Llorente-Saguer (2011), "The Power of Sunspots", *SFB 649 Discussion Paper* 2011-070.

Möbius, M. M., M. Niederle, P. Niehaus, and T. S. Rosenblat (2011), "Managing Self-Confidence: Theory and Experimental Evidence," *NBER Working Paper* 17014.