The Nature of Social Learning: Experimental Evidence

Stefan P. Penczynski*
University of Mannheim

June 2016

In the wide economic literature on social learning, many models of behavior – rational and non-rational – have been proposed. This paper analyzes experimental data that is able to distinguish between most of them on the individual level. It relies on rich, existing data from Çelen and Kariv (2004) as well as new experimental data that includes written accounts of reasoning from incentivized intra-team communication. Three datasets provide consistent evidence that naïve inference in form of the best response to truthful play is the most common approach to social learning. In terms of a proposed level-k model that unifies various forms of inferential reasoning, the empirical type distributions feature heterogeneity similar to other level-k applications.

Keywords: Social learning, levels of reasoning, naïve inference

JEL Classification: C91, D82, D83, D84

*Department of Economics, University of Mannheim, L7, 3-5, 68131 Mannheim, Germany, stefan.penczynski@uni-mannheim.de, Tel. +49 621 181 3656, Fax. +49 621 181 1893. I thank Ayala Arad, Konrad B. Burchardi, Dirk Engelmann, David Gill, Rosemarie Nagel, Anders Poulsen, David Rojo Arjona, Robert Sugden and seminar participants at ESA Chicago 2011, University of East Anglia, SABE Granada 2012, HeiMaX Heidelberg 2012, GfeW Karlsruhe 2012, M-BEES Maastricht 2012, and University of Mannheim for helpful comments. I also thank Shachar Kariv and Boğaçhan Çelen for the provision of the data from their 2004 study. Katharina Momsen, Allison Nuñez, Dennis Shiraev, and Julia Walecki provided excellent research assistance.

Most of the current economic literature on social learning goes back to the seminal model introduced by Bikhchandani, Hirshleifer and Welch (1992). It captures essential points of diverse situations such as investments in financial markets, technology adoption, sequential voting, mating choice, takeover decisions, etc.: In a sequence of Bayesian decision-makers that want to act according to the true but uncertain state of the world, players hold private information and observe their predecessors' decisions. A striking conformity and a wastefully low revelation of private information are predicted to eventually emerge in form of an information cascade.

The experimental investigation of social learning was initiated by Anderson and Holt (1997, AH henceforth), whose basic setup is still used in most recent experiments: Subjects observe a binary private draw that reveals the state of the world correctly with a probability of $2/3$. Before subjects choose a binary action with the aim of matching the state, they can observe the actions of all predecessors.

Due to the coarse binary structure, this setting provides only limited information about the reasoning underlying the observed decisions. Eyster and Rabin (2010, p. 236) criticize that "the existing experimental literature is generally not well-designed to differentiate among likely hypotheses about the nature of observational learning." For example, they introduce a model of naïve inference which corresponds to a best response to truthful play (Best response trailing naïve inference, BRTNI) and notice that the predictions of play in the AH setup coincide with those of Bayesian-Nash play. Furthermore, even quite distinct behavioral models such as Quantal Response Equilibrium (QRE, McKelvey and Palfrey, 1995) and cognitive hierarchy (Camerer, Ho and Chong, 2004) make similar predictions for action data and lead to similarly good fits of the data as shown by Goeree, Palfrey, Rogers and McKelvey (2007).

This paper sets out to differentiate between hypotheses about the nature of social learning with the help of three kinds of experimental data. First, I investigate behavior in the AH setting with the help of intra-team communication, using an experimental design that was introduced by Burchardi and Penczynski (2014). The communication protocol provides incentivized written accounts of individual reasoning which allow me to distinguish various types of inference. Second, I analyze existing data from a social learning framework with continuous strategies and signals established by Çelen and Kariv (2004, ÇK henceforth). A "fingerprint" of 15 decisions from each subject is able to differentiate various types of behavior in a way methodologically similar to studies in the strategic thinking literature (for example, Costa-Gomes, Crawford and Broseta, 2001). Finally, using intra-team communication within this framework allows me to cross-validate results from message contents and choice profiles and check for omitted types.

In order to fix ideas about types of inference, I specify a level-k model of social learning that encompasses a variety of types proposed in the literature.[1] Like the level of reasoning model of strategic thinking (Nagel, 1995; Stahl and Wilson, 1995; Costa-Gomes and Crawford, 2006; Camerer et al., 2004), the model organizes types by the number of iterated best responses to uninformative level-0 play. A level-1 player believes that all predecessors' actions stem from such level-0 play; her best response is therefore to follow the private signal. Believing all others to play according to their signal, the level-2 player views all previous actions as similarly informative as her own private signal. Level-3 players realize that some level-2 actions do not reflect private signals and assess them as uninformative. And so on for level-4, level-5, etc.

In all instances of the data analysis, I find evidence in favor of the modeled heterogeneity and inferential naïvety. Subjects of level-1, level-2 and level-3 are detected throughout. The modal type in every analysis turns out to be level-2, which is equivalent to Eyster and Rabin's (2010) "BRTNI" or Hung and Plott's (2001) "naïve Bayesians".

[1] The idea of a level-k model of social learning is not new to this paper. Kübler and Weizsäcker (2004) model a related heterogeneity in an error-rate model. Goeree et al. (2007), Dominitz and Hung (2009), Eyster and Rabin (2010) and Bohren (2015) discuss cognitive hierarchy or level-k as possible alternatives to their models.

The paper contributes to a large experimental literature that, starting with the seminal paper by AH, set out to investigate behavior in social learning. Studies in that literature generally find substantial deviations from Bayesian inference but rarely focus on the nature of those. Most studies used either QRE frameworks which modeled deviations as random noise (for example, Çelen and Kariv, 2004; Choi, Gale and Kariv, 2005; Drehmann, Oechssler and Roider, 2005) or relied on data from the AH setup which is less informative about the nature of deviations (see, for example, the papers analyzed in Weizsäcker, 2010). Few studies collected complementary data. For example, Ziegelmeyer, Koessler, Bracht and Winter (2010) elicit beliefs about signals and states of the world. Kübler and Weizsäcker (2004) give the option to buy at a cost further signals, Çelen and Hyndman (2012) allow for costly link formation. Both methods allow to infer subjects' beliefs. The intra-team communication in this experiment gives a very direct insight into subjects' reasoning and is therefore well-suited to investigate the nature of social learning.

The paper also contributes to the large literature on strategic thinking (Crawford, Costa-Gomes and Iriberri, 2013) and shows that the level-k concept is useful to describe naïve inference in social learning. Level-k models have been implemented in other private information settings like common value auctions (Crawford and Iriberri, 2007) and zero-sum betting (Brocas, Carrillo, Wang and Camerer, 2014). Both applications feature strategic interaction as well as inference of private information. The investigation of social learning abstracts from strategic interaction and shows that the inference alone is equally well described with this framework.

The methodology of generating incentivized written accounts of reasoning allows for illuminating results that are complementing insights from action data and neuroeconomic data. The topic of "theory of mind" that treats the capacity of human subjects to think about the state of mind of another human subject has received a lot of attention in neuroeconomics (for example, Hampton, Bossaerts and O'Doherty, 2008; Coricelli and Nagel, 2009). When observed through the lens of incentivized verbalization, the various stages – or levels – of complexity that these theories of mind attain are identified very concretely in the present study.

The paper is structured as follows. Section 1 introduces the general model and the experimental design. Section 2 presents the results of the experiments with intra-team communication in the AH framework. Section 3 presents the analysis of ÇK's original data as well as new data with intra-team communication. Section 4 concludes.

1. General model and experimental design

1.1. The general model

Countably infinite players identified by their position in the sequence $t \in \mathbb{N}$ take actions $a_t \in \{A, B\}$ sequentially. The payoff of player t is determined by the match with the unknown state of the world $\omega \in \{A, B\}$, that is,

$$\pi_t(a_t) = \begin{cases} 1, & \text{if } a_t = \omega, \\ 0, & \text{if } a_t \neq \omega. \end{cases}$$

Both states are ex ante equally likely. A player t receives a private signal $s_t \in S$ that induces a private belief $Pr(\omega = A|s_t) \in (0, 1)$. Further, each player observes the history of actions up to round $t-1$, $H_{t-1} = \{a_1, a_2, \ldots, a_{t-1}\}$. Each observed action a_i in H_{t-1} induces beliefs $\psi_i \in \mathcal{S}$, a probability measure over the signal space S. Inferring likely signals from observed actions is reflected as mapping $I : \times_{i=1}^{t-1}\{A, B\}_i \mapsto \times_{i=1}^{t-1} \mathcal{S}_i$.

Importantly, in a level of reasoning model, the inference depends on the population beliefs of player t, which, for simplicity, are assumed to be homogeneous and degenerate on level $k-1$. The strategy of player t of level k is denoted $\sigma_t^k(s_t, I^{k-1}(H_{t-1}))$ and maps the private signal s_t and the inference from the observed history $I^{k-1}(H_{t-1})$ into an action $a_t \in \{A, B\}$ in a way that the expected payoff under the available information is maximized. Not regarding or not being able to infer a signal is reflected with the empty signal or inference $\varnothing$ for which $Pr(\omega = A|\varnothing) = \frac{1}{2}$.

As usually assumed, the level-0 player randomizes uniformly over the action space and thus plays irrespectively of his own signal and the observed history: $\sigma_t^0(\varnothing, \varnothing) \sim U(\{A, B\})$. Intuitively, such a level-0 player lacks either an understanding of the game or the motivation to use any information for his action.

The level-1 player assumes others to be level-0 players and best responds accordingly.[2] The recovery of the private signals of preceding level-0 players is not possible because no preceding action is connected to the private signal. Solely the own private signal is informative to level-1 players and enters their strategy, $\sigma_t^1(s_t, I^0(H_{t-1})) = \sigma_t^1(s_t, \varnothing)$. This player understands how her private signal is informative but either expects others to play randomly or does not understand how others' decisions can be like hers: informative.

While level-1 players disregard the history of what is perceived as uninformative actions, level-2 strategies rely on it heavily since all predecessors' actions are believed to be based exclusively on the private signals. The level-2 strategy simply aggregates the own signal and what seem to be the others' signals given only their action. This player does not see how others might be like him: influenced by their history.

Level-3 players understand that level-2 players are influenced strongly in their choice by their predecessors' choices. They thus differentiate individual predecessors by their specific history when inferring from their action. Level-3 is the lowest type to do so. Knowing to which extent previous choices matter for level-3, level-4 players will take this into account for their inference, and so on will higher level players. Depending on the social learning framework, there can be a point in the hierarchy from which on levels cannot be distinguished by their actions.

[2]Crawford and Iriberri (2007) distinguish two level-0 types, a truthful and a random one. In my setup, since one is a best response to the other, they both induce a similar set of types. The exhibition is easier with a single set. Any difference between higher types would be merely semantic.

1.2. Similar models in the literature

The level-k types in this model make similar action predictions as various types of behavior that have been proposed in the literature. Table 1 provides a short list of those types and points out conceptual differences. It shows special noisy cases of the QRE model as used, for example, in Goeree et al. (2007) or in the related model by Kübler and Weizsäcker (2004) which incorporates limited depth of reasoning. Furthermore, Eyster and Rabin's (2005) cursedness takes the same manifestation as level-1 in social learning although it is formulated in an equilibrium context. If players believe that such types exist, concepts like BRTNI (Eyster and Rabin, 2010) naturally arise. The level-3 player framework infers – due to the modelled population belief – signals when actions in cascades are impossible to result from level-2 play. In the AH setup this predicts the same actions and beliefs as a Bayesian that infers information from non-equilibrium cascade deviations, put forward by AH (p. 850).

Regarding network structure, the similarity between level-1 types and rational players at socially uninformed nodes in general networks is noteworthy. Both can take an important role as sources of new private information, as results in Acemoglu et al. (2011) point out for the latter case.

The level-k model thus gives an interesting range of types that the informative data in this study can potentially distinguish. Other models of reasoning are not nested or easily discriminated, so that this study will be silent on their relevance for social learning. For example, Guarino and Jehiel (2013) use the analogy-based equilibrium concept (Jehiel, 2005) in social learning to the effect that players know how the frequencies of actions relate to the state of the world, but are not able to relate it to the specific history or the private signal of others. Similarly, the cognitive hierarchy model by Camerer et al. (2004) differs from level-k models most substantially in the non-degenerate population belief. Both models differ from the proposed model in subtle ways that are likely to go beyond the discriminatory power of the data at hand.

	Study	Comment/Difference
Level-0 player		
Crazy type	Smith and Sørensen (2000)	Plays a fixed action
Fully noisy player	e.g. Goeree et al. (2007)	$\lambda = 0$*
Level-1 player		
Private-information revealer	Hung and Plott (2001)	
"Follow your own signal"-type	Kübler and Weizsäcker (2004)	$\lambda_1 \to \infty$, $\lambda_2 = 0$*
Fully cursed player	Eyster and Rabin (2005)	Equilibrium type, fails to see link between predecessors' types and actions
Socially uninformed type	Bohren (2015)	Observes private signal only
Level-2 player		
Naïve Bayesian	Hung and Plott (2001)	
"Counting heuristic"-type	Kübler and Weizsäcker (2004)	$\lambda_1 = \lambda_2 \to \infty$, $\lambda_3 = 0$*
Best Response Trailing Naïve Inference (BRTNI)	Eyster and Rabin (2010)	Best response to fully cursed player
Level-3 player		
Bayesian	Anderson and Holt (1997)	Equilibrium type, but infers from non-equilibrium cascade deviations in AH setup

* λ is the precision in a logistic choice function, λ_i relates to the ith iterated best response.

Table 1: Behavioral types related to level-k types.

1.3. Population belief

For ease of exposition, the model is so far formulated with a degenerate population belief.[3] While non-degenerate beliefs are straightforward to incorporate, the sequential nature of the social learning decisions introduces a further challenge. Because individual players differ by their location and the information they hold, successors potentially hold differentiated beliefs about them. The mere history can contain information that allows a player to update her population belief. For example, a level-3 player observes an action that cannot come from a level-2 player given her history.

To account for this sequential structure, I propose a model with a nearly degenerate population belief. Before observing his history, a level-k player believes with near certainty that all predecessors are level $k-1$, attributing very small probabilities to the levels further below. In particular, I assume that $p_i^{k,k-1} = 1 - \sum_{d=1}^{k-1} \varepsilon^d$ and $p_i^{k,l} = \varepsilon^{(k-1)-l}$ for $l = 0, \ldots, k-2$, for a small and positive constant ε. Upon observing the

[3]Data from beauty contest games suggest that this is a reasonable assumption (Nagel, 1995; Costa-Gomes and Crawford, 2006; Burchardi and Penczynski, 2014).

history, the player uses Bayes' rule to update the population belief on each predecessor. In the example, the prior beliefs are $p_i^{3,2} = 1 - \varepsilon - \varepsilon^2$, $p_i^{3,1} = \varepsilon$, and $p_i^{3,0} = \varepsilon^2$. If the action is equally likely to come from level-0 and level-1, the updated population belief is nearly degenerate on level-1: $p_i^{3,0} = \frac{\varepsilon}{\varepsilon+1}$ and $p_i^{3,1} = \frac{1}{\varepsilon+1}$.

This specification is chosen for two reasons. First, if an observed action and history rule out that the player is of level-$k-1$, the update with Bayes' rule for this individual leads again to a nearly degenerate population belief on level $k-2$. Second, with this formulation ties are broken in favor of the inference under a population belief of level-$k-2$ or lower. For example, level-2 players break ties that would arise under $p_i^{2,1} = 1$ in favor of their private information due to the remaining ε probability of random level-0 play.

1.4. Team communication

One measure that this study takes to overcome the problem of limited insight into the nature of social learning is the gathering of complementary information about participants' reasoning process. For that purpose, I conducted two experiments with an intra-team communication protocol that yields incentivized written accounts of subjects' reasoning (Burchardi and Penczynski, 2014).[4]

The communication protocol incentivizes the individuals' messages within the team as follows. Participants are randomly assigned into teams of two players. The two members are connected through the modified chat module of the experiment software.[5] Once the situation is observed, each team member can state a so-called "suggested decision" and justify it in a written message. As soon as this is done for the first three situations, the suggestions and messages get exchanged simultaneously. In a next step, both team members individually state their "final decision" for the three first decisions. It is known to them that for each situation one of the two team members' final deci-

[4]Instructions are reprinted in sections C and E of the online appendix.
[5]The experiments were programmed and conducted with the software z-Tree (Fischbacher, 2007).

sions will be chosen randomly by the computer to count as the "team's action". This construction provides incentives to state the full reasoning underlying the suggested decision in a clear and convincing way.

In contrast to the original design in Burchardi and Penczynski (2014), subjects first write suggested decisions for *three* situations before both the suggested decisions and the messages are simultaneously exchanged. This ensures that the first three messages are written prior to any communication with the team partner and reflect only the individual's reasoning. The same procedure is repeated for the last three decisions.[6]

The message is entered freely without explicit space or time limitations. In the AH experiment, apart from the suggested decision, another structured part of the communication consists of quantifying the confidence in the proposal. Subjects can put numbers between 50 and 100, indicating the subjective probability that their suggested decision coincides with the true state of the world. As part of the incentivized communication stage, this is similar to eliciting individual beliefs.

The messages are classified independently by two research assistants (RA). For each individual message they indicate the level of reasoning that the message corresponds to most closely. For this task, the RAs are introduced to the level-k model and receive detailed instructions about characteristics of the individual types.[7]

The following features of reasoning are derived from the model and should be present for the message to be classified as a certain level. Random level-0 play results from misunderstanding the nature of the game or from putting arguments that are orthogonal to any reasonable inference from private signals and public actions. Level-1 play features an open disregard of others' actions or a strong emphasis on the unambiguity of the own signal in contrast to others' decisions. Level-2 play requires that others' actions are taken at face value as a signal and predecessors are not dif-

[6]Having messages exchanged after three rather than one suggested decision has the advantage that all three messages are not "contaminated" by the team partner's arguments. In the original setup with an immediate exchange, this would only be true for the very first message.

[7]Instructions for the RAs are reprinted in sections D and F of the online appendix.

ferentiated. In contrast to this, level-3 play distinguishes individual predecessors and evaluates the information content of the action for each of them.

Both RAs first provide independent sets of classifications. After this, both are anonymously informed about all classifications of the other RA and have the possibility to revise their own classification. This iteration serves to reconsider diverging classifications and to screen errors or misperceptions. Importantly, the information is merged by using only those classifications that coincide between the two RAs. The RAs agree in a large majority of classifications (AH: 567 (541) out of 636, 89.2% (85.1%); ÇK: 218 (194) out of 252, 86.5% (77.0%); pre-revision in brackets). In AH, 17 out of 1272 individual RA level classifications imply a different decision than observed in the suggested decision. 10 of those come from 5 decisions in which the two RAs agree, suggesting a discrepancy between message and suggested decision on the subject side.[8]

1.5. Hypotheses

On the basis of the described model, I can formulate hypotheses that can be tested thanks to the communication data and ÇK's action data.

Hypothesis 1 *Classified in terms of level-k types, reasoning in social learning situations is heterogeneous.*

Hypothesis 2 *The level-k distribution is hump-shaped and features level-1 or level-2 reasoning as most common types.*

This hypothesis is motivated by commonly observed distributions in the level-k literature (Costa-Gomes and Crawford, 2006; Arad and Rubinstein, 2012) as well as the main types discussed in the social learning literature (see table 1).

[8]Burchardi and Penczynski (2014) provide further evidence for the robustness and replicability of this kind of classification.

Hypothesis 3 *The communicated reasoning is not different from the theoretical type's reasoning.*

2. Anderson and Holt (1997)

2.1. The model

The framework used by AH has a simple binary information structure with states of the world $\omega \in \{A, B\}$, private signals $s_t \in \{A, B\}$, and actions $a_t \in \{A, B\}$. The private signal is informative with $Pr(s = \omega) = \frac{2}{3}$.

The level-0 player randomizes over the action space, $\sigma_t^0(\varnothing, \varnothing) \sim U(\{A, B\})$. As in the general case, the level-1 player best-responds by simply following his signal, $\sigma_t^1(s_t, I(H_{t-1})) = \sigma_t^1(s_t, \varnothing) = s_t$. Best responding to this, the level-2 player views previous decisions as equally informative as her own signal, $I^1(H_{t-1}) = H_{t-1}$, and aggregates the information by counting the evidence – be it previous actions or the private signal – for a given state of the world. Consequently, the level-3 player can infer the private signal from the resulting actions only if a different signal had caused the level-2 player to choose a different action.

Note that in this environment level-2 and level-3 strategies yield the same actions, they only differ in their beliefs about the state of the world. In turn, level-3 actions and beliefs are identical to those of Bayesians for a given history in equilibrium. Consequently, players of levels higher than 3 only differ in the population belief and not in their actions or beliefs about the state of the world. This shows why in this framework action data alone is not well-suited to distinguish certain forms of behavior, as noted by Eyster and Rabin (2010).

2.2. Experimental procedures

The AH experiment was conducted in 7 experimental sessions at the LEEDR Laboratory of Cornell University. In teams of two, 106 mostly undergraduate students were

		Urn choice					
		1	2	3	4	5	6
	Information*	$BBAb$	$BAABAb$	$BAAa$	$BBABa$	$ABAAAb$	$BAAAb$
	Suggested decision**	0.953	0.689	0.057	0.538	0.415	0.547
	Confidence	69.76	60.75	69.14	64.27	67.17	64.68
Probability	Level-0	0.5	0.5	0.5	0.5	0.5	0.5
of choice B	Level-1	1	1	0	0	1	1
	Level-2/-3/Bay.***	1	1	0	1	0	0

Notes: * A and B denote observed actions, a and b denote private signals favoring A or B, respectively. ** Suggested decisions are reported in fractions of choice B. *** For urn choices 1 and 4, the Bayesian probability of choice is not restricted to 1 in equilibrium since the histories are off the equilibrium path.

Table 2: Suggested decisions and confidences as well as choice probabilities for different theoretical types in the 6 urn choices ($N = 106$)

taking decisions in 6 social learning situations that arose in the original experiment of AH. I chose decision situations with a substantial history of public information. Different types take different decisions in some but not all of them, since in this framework, those situations are somewhat particular.

A session took on average 60 minutes and led to average payments of 11.30 USD per subject. In the experiment, the binary setting described in section 2.1 has equally likely states $\omega \in \{A, B\}$ reflected by two urns A and B. The signals $s \in \{\text{White}, \text{Black}\}$ are informative draws from the urns, where urn A contains 2 white and 1 black ball and urn B contains 1 white and 2 black balls. Possible actions are $a \in \{A, B\}$.

2.3. Results from team communication data

Table 2 gives an overview of the 6 implemented urn choices. The first line gives the information (history and signal) available to subjects. The next two lines give means of the suggested decisions and confidences observed in the experiment. The suggested decisions take values A or B and the confidences take values between 50 and 100. The majority of urn choices 1-3 are in accordance with the predictions given in the last three lines. For choice 4-6, subjects are split between A and B.

The messages from the intra-team communication give direct information about in-

		Highest level (∅ 1.66)					
		0	1	2	3	NA	Total
	0	8	3	3	1	0	15
Lowest level	1		12	6	0	0	18
(∅ 1.38)	2			40	3	0	43
	3				1	0	1
	NA					29	29
	Total	8	15	49	5	29	106

Table 3: Level classification in first three decisions

dividuals' reasoning. In the upcoming analysis, I focus on the first three decisions that were taken prior to any contact with the team partner.[9] I pool the information from the first three decisions and consider all coinciding level classifications other than NA. Table 3 summarizes the data by giving the subjects' lowest and highest level over these three decisions.[10] 29 subjects (27%) are not classified, mostly because they did not write any message on any decision. 79% of the remaining 77 subjects are attributed a single level.

Focusing on the marginal distributions or the numbers of subjects on the diagonal with coinciding lowest and highest level, the table reflects a pronounced heterogeneity with all levels 0-3 arising at least once, thus supporting hypothesis 1. By a large margin, the mode behavior is level-2 with 40 subjects classified exclusively as level-2, a result which supports hypothesis 2.

Messages. In the following, a closer look at the messages investigates hypothesis 3 and the question whether subjects' comments differ from the reasoning postulated in

[9]The classification for the last three decisions is reported in table 16 on page 39. A Wilcoxon signed-rank test does not reject the hypothesis that the level lower bounds or the level upper bounds are the same between the first three decisions (table 3) and the last three decisions (table 16, $p_{lower} = 0.515$, $p_{upper} = 0.470$). To control for order effects, the order of choices was reversed in the last three sessions. If the choices are not numbered, the analysis presents the temporal order of the decisions, i. e. the "first three decisions" will always refer to the decisions taken without prior conversation with the team partner.

[10]The data excludes any outcome of a non-classification (NA). If all three decisions are not classified (NA), the subject appears under NA. For more details, table 17 on page 40 reports this data by decision. Tables 22 and 23 in the online appendix report the data of both RAs on an individual level.

the model.

Level-0. The model is based on uninformative level-0 reasoning, which is modeled by random play. The exemplary messages in table 9 on page 32 show that subjects misunderstand the game or self-report to have no understanding of the game. Fittingly, more than 70% of those messages come with the lowest possible confidence of 50.

Interestingly, only 2 out of the 24 decisions from level-0 players are not in line with the private signal. Based on those, therefore, 6 subjects would be classified as level-1 rather than level-0 players in the model. The literature suggests a truthful level-0 type (Crawford and Iriberri, 2007), which could be an alternative explanation that fits the decisions, the communication as well as the confidence.

Level-1. The messages that are categorized as level-1 reasoning illustrate an increment of understanding beyond level-0. Table 10 shows that subjects mostly explain why the private draw is informative and often report a confidence of 66 or 67 based on this single private signal. Subjects follow their signal without exception. Rather than explicitly expecting random behavior in their predecessors, subjects express confusion or ignorance as to how to make use of the observed history.

Level-2. Tables 11 and 12 show how subjects classified as level-2 take into account the history of others' actions in addition to their private signal. Characteristically, this perception of other players is, however, not very differentiated. The observed history is often summarized in statistics reporting, for example, that "3 out of 5 predecessors selected A", without taking into account the predecessors' position or choice situation. As a result, subjects can become very confident in their own choice. The information from the history is then mostly combined with the odds of the private draw being correct. As modelled, level-2 players do craft a model of others' reasoning: the simple one of truthful play.[11]

Level-3. Finally, tables 13 and 14 present communication that was classified as reflecting level-3 reasoning. As expected from the model, the messages show clearly how players differentiate between predecessors, make use of the knowledge they have about individual histories, are aware that other might just have followed their history and try to back out private signals. The model of their predecessors' thinking is clearly more elaborate than those of lower levels. In the few few observations, the updating of the population belief is compatible with the nearly degenerate population belief (see, for example, subject 15 at BBAb).

3. Çelen and Kariv (2004)

3.1. The model

The setting implemented in ÇK has discrete actions $a_t \in \{A, B\}$ but continuous, uniformly distributed, signals $s_t \sim U[-10, 10]$ and a state of the world defined as

$$\omega = \begin{cases} A, & \text{if } \sum_{t=1}^{T} s_t \geqslant 0, \\ B, & \text{otherwise.} \end{cases}$$

Usually, a strategy would map the observed history and the signal into an action. Since the expected payoffs are monotonic in the private signal, the optimal strategies are summarized in a threshold $\theta \in [-10, 10]$. In the experiment, subjects choose a threshold before seeing their own signal. After the realization of the signal, the computer derives actions from the thresholds as follows,

$$\sigma_t(s_t, \theta) = \begin{cases} A & \text{if } s_t \geqslant \theta, \\ B & \text{if } s_t < \theta. \end{cases} \quad (1)$$

[11] The model by Guarino and Jehiel (2013) predicts similar behavior with the difference that the public information is used directly as evidence for the state of the world, not regarding the connections to predecessors' private signals. Although it is not clear what fraction of A and B actions players expect in the different states of the world, it seems more plausible from the messages that the actions are simply viewed as identical to the signal rather than evaluated according to a certain belief as to how often they occur in each of the states of the world.

As opposed to the general setup, strategies will therefore be given by the threshold θ. Consequently, the observed action only indicates that the signal is above or below the unobserved threshold. The inference from the action thus requires above all a belief about the threshold $\hat{\theta}$. In conjunction with the observed action, a possible range of the signal realization can be inferred and leads to an expected value of the signal.

A rational Bayesian player uses previous actions as well as beliefs about previous thresholds to craft her optimal cutoff strategy,

$$\theta_t^B(H_{t-1}, \{\hat{\theta}_i^B\}_{i=1}^{t-1}) = -E\left[\sum_{i=1}^{t-1} s_i \middle| H_{t-1}, \{\hat{\theta}_i^B\}_{i=1}^{t-1}\right].$$

Bayesian inference assumes rational expectations with respect to the thresholds used by the predecessors (see equations 1 and 2 in ÇK).

In the level-k model, a level-0 player randomizes over the strategy space,

$$\theta^0 \sim U[-10, 10].$$

Although this is a plausible level-0 strategy, it turns out to be an atypical level-0 *belief*: because the computer processes the threshold and the signal automatically, the resulting action is as informative as a threshold of 0. I therefore model the level-0 *belief* as randomizing uniformly over the two uninformative actions -10 and 10,

$$\hat{\theta}_t^0 = \begin{cases} -10, & \text{with Pr} = 0.5, \\ 10, & \text{with Pr} = 0.5. \end{cases}$$

This level-0 belief reflects a uniform random choice of actions $a_t \in \{A, B\}$.[12] The best response for a level-1 player is to follow the private signal and play the cutoff strategy

$$\theta_t^1(H_{t-1}, \{\hat{\theta}_i^0\}_{i=1}^{t-1}) = 0.$$

For a level-2 reasoner, the observation of one action resulting from such a level-1 strategy leads to an expected signal of 5 after action A and -5 after action B. It follows that $E\left[\sum_{i=1}^{t-1} s_i \middle| H_{t-1}, \{\hat{\theta}_i^1 = 0\}_{i=1}^{t-1}\right] = 5 \cdot \#A - 5 \cdot \#B$, where $\#a$ gives the

[12]The main role of the level-0 is commonly to provide a starting point for iterated beliefs. This construction ensures in this particular, computer-assisted design that these beliefs start out with an assumption of uninformative play. At the same time, it is important to allow for the existence of level-0 players that randomize over $[-10, 10]$.

number of previously played A or B actions. Thus, level-2 players reach the limits of the strategy space very quickly and play -10 or 10 once one action was observed in two more occasions than the other one. Here, the level-k model predicts information cascades in which own signals are entirely disregarded ($\theta \in \{-10, 10\}$). Only five different actions are predicted to be observed from a level-2 player, $\theta_t^2(H_{t-1}, \{\hat{\theta}_i^1\}_{i=1}^{t-1}) \in \{-10, -5, 0, 5, 10\}$. The best response to level-2 play fully discounts actions that resulted from an uninformative threshold of -10 or 10 and uses only the remaining actions to infer signals, and so on for higher levels.

Due to the continuous strategy and signal space in this framework, higher level thresholds can be differentiated. They form different beliefs about the informativeness of observed actions and therefore choose different thresholds. This framework is thus very well suited to experimentally investigate the level-k and other types. Already ÇK's experiment is informative about the type distribution.

The strategies can be recursively expressed as

$$\theta_t^k(H_{t-1}, \{\hat{\theta}_i^{k-1}\}_{i=1}^{t-1}) = -E\left[\sum_i^{t-1} s_i \middle| H_{t-1}, \{\hat{\theta}_i^{k-1}\}_{i=1}^{t-1}\right].$$

3.2. Analyzing choice data

ÇK implemented their social learning framework in an experiment at New York University with 40 undergraduate subjects that played 15 independent rounds in sequences of 8 players. After each completed sequence, participants were informed whether their action coincided with the state of the world and were paid \$2 if it did, and nothing otherwise.

Thanks to the rich signal and action structure, I can use the 15 choices of each subject to investigate her individual type. Similar to Costa-Gomes et al. (2001) and Costa-Gomes and Crawford (2006), I compare the chosen thresholds $\{\theta_r\}_{r=1}^{15}$ to the thresholds $\{\theta_r^k\}_{r=1}^{15}$ that a theoretical level-k or rational type is predicted to choose in the exact same 15 situations. I distinguish between the 4 types: level-1, level-2, level-3, and Bayesian, and assume the level types to hold nearly degenerate population beliefs as specified in section 1.1.

I use the sum of squared differences, $SSD^k = \sum_{r=1}^{15}(\theta_r - \theta_r^k)^2$, $k = 1, 2, 3, B$ as the distance metric between fingerprints of subjects and theoretical types.[13] In order to test whether observed behavior is close to a theoretical type by chance, I simulate each type's distribution of SSD^k resulting from uniform random thresholds. The obtained distributions are type-specific since, for example, random thresholds are closer in quadratic distance to $\theta^1 = 0$ than to thresholds that frequently reach the limits of the action space. I will use the resulting p-values of a given fingerprint for each type not only in its original meaning in hypothesis testing, but with slight abuse also as a type-independent measure of closeness between observed fingerprints and theoretical types.

I classify fingerprints into types in two alternative ways. First, I classify subjects into the type that gives the lowest significant SSD^k (p-value $\leqslant 0.05$). Second, I classify subject i that is significantly close to, say, two types, i. e. with two types' p-values below 0.05, as partly one type and partly the other type. The share s_i^k attributed to each type is chosen anti-proportionally to its p^k-value.[14]

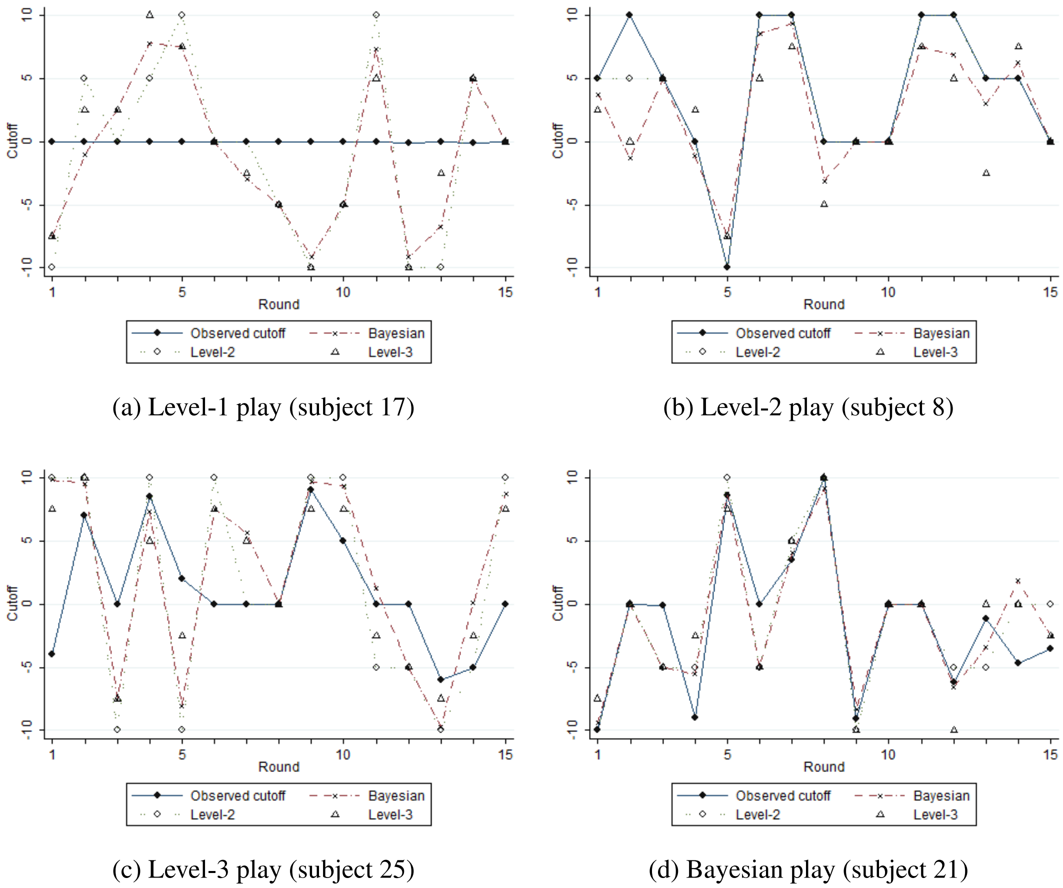
3.3. Results from Çelen and Kariv (2004) data

The original data from ÇK enables a first validation of the previous results in a different setting. Figure 1 illustrates the data for individual subjects over 15 rounds arranged into

[13] I use the SSD^k because calculating the more standard expected payoff for a cutoff in a given situation is very complex due to the payoff structure and the uniformly distributed signals. The SSD^k is a useful approximation since the probability of playing the correct action is quadratic in the cutoff because the cutoff linearly influences i) the probability of playing a certain action and ii) the probability of playing the correct action given the action played as long as the considered signal realizations for all players are in $(-10, 10)$.

[14] More precisely, if $p_i^k \leqslant 0.05$, the share is

$$s_i^k = \frac{\frac{\Sigma}{p_i^k}}{\sum_{\{l:p_i^l \leqslant 0.05\}} \frac{\Sigma}{p_i^l}}, \text{ with } \Sigma = \sum_{\{l:p_i^l \leqslant 0.05\}} p_i^l.$$



(a) Level-1 play (subject 17)
(b) Level-2 play (subject 8)
(c) Level-3 play (subject 25)
(d) Bayesian play (subject 21)

Figure 1: Subjects' and types' "fingerprints" in the data of ÇK

"fingerprints" and compared to theoretical types.

Type k	Frequency	Fraction (%)	$\hat{p}^k$
1	6	15.0	0.176
2	20	50.0	0.588
3	5	12.5	0.147
B	3	7.5	0.088
NA	6	15.0	–
Total	40	100.00	1

(a) Classification by lowest significant SSD^k

Type	Frequency	Fraction (%)
1	6.42	16.0
2	18.61	46.5
3	4.07	10.2
B	4.90	12.3
NA	6.00	15.0
Total	40	100.00

(b) Classification by p-values below 0.05

Table 4: Type overview for 40 subjects in ÇK

The type distribution resulting from the first kind of classification is shown in table 4a.[15] The players that do not differ significantly from random play for any type are

[15]Table 21 on page 1 of the online appendix reports the SSD^k for each subject and type.

reported as unclassified (NA). A pronounced heterogeneity of reasoning across the 40 subjects can be observed with each type being present at least a few times, supporting hypothesis 1. The level-k distribution is hump-shaped and, by a large margin, the most common behavior turns out to be level-2 play, again supporting hypothesis 2. Table 4b shows that the main results remain unchanged when taking into account how well types can be differentiated.[16]

Like subject 8 in figure 1b, 7 players choose thresholds $\theta \in \{-10, -5, 0, 5, 10\}$ in all 15 situations. Players classified as level-2 account for 103 of the 136 instances (75.7%) of fully ignoring the own signal ($\theta \in \{-10, 10\}$). The model thus proposes a simple behavioral explanation for the information cascading that motivated ÇK's study.

Goeree et al. (2007) show that their extensive AH-type data is equally well explained by QRE as by the iterated best-response model of cognitive hierarchy. ÇK set up a QRE model that features both random and rational behavior, with the latter taking into account previous decision errors in a Bayesian way. Their data allows me to compare the explanatory power of this model with the level-k model as presented in this paper. For that purpose, I use the estimated parameter values that they present in table 3 and recursively calculate the predicted cutoffs $\hat{\theta}$ as described in equation 9 (ÇK, p. 492–493). Analogously, I use the estimates of the level distribution $\hat{p}^k$ in table 4a in order to calculate the cutoffs under the level-k model. Again, the cutoffs are recursively calculated as a mixture of the types' cutoffs θ^k with the weights $\hat{p}^k$, with $k = 1, 2, 3, B$, $\theta^{\text{Mix}} = \sum_k \theta^k \cdot \hat{p}^k$.

The distance between the models' fitted cutoffs and the chosen cutoffs is judged again by the SSD. The average SSD per decision is 34.6 in ÇK's model and 27.3 in the level-k model. The hypothesis that the two models explain behavior equally well is rejected, the level-k cutoffs lead to significantly lower SSDs (Wilcoxon signed-rank

[16]The experiment was not explicitly designed to investigate behavior in one-shot situations. After each of the 15 sequences, subjects received feedback on the state of the world and their payoff. The online appendix B.2 shows that subjects seem to slowly learn over time.

test, $p < 0.001$). However, since the level-k estimates are on the individual level, the level-k model has a larger degree of freedom than ÇK's model. Indeed, in the out-of-sample prediction for the data in the next section, the average SSD per decision is lower for ÇK's (21.5) than for the level-k predicted cutoffs (29.7).

3.4. Experimental procedures

I conducted the experiment with the team communication design in 4 experimental sessions at the mLab of the University of Mannheim. In teams of two, the 42 mostly undergraduate students were taking decisions in 6 social learning situations that arose in the original experiment of ÇK. Again, I chose decision situations that occurred later in the sequence and in which different types choose different thresholds. A session took on average 60 minutes and led to average payments of 8.30 EUR per subject. In both experiments, subjects are told at the beginning that they will experience situations that arose in a previous experiment which was played by individuals.

3.5. Results from team communication data

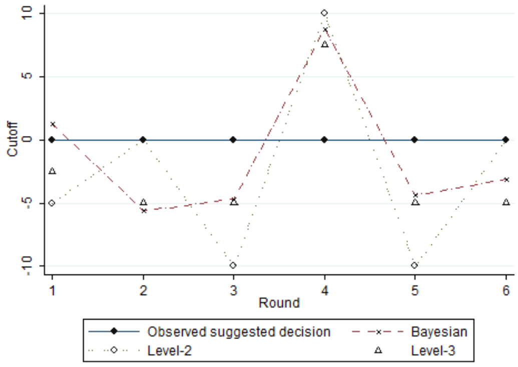
For the team communication experiments in the ÇK framework, table 5 gives an overview of the six implemented scenarios which were chosen in order to distinguish types $k = 1, 2, 3, B$ as cleanly as possible. The first line gives the information about the history available to subjects. The following lines give mean and standard deviation of the choices in $[-10, 10]$ as well as predicted play by level-k and Bayesian players.

Like in section 3.3, the fingerprints of the six suggested decisions per subject can be used to identify their type as illustrated in figure 3. The analysis of the SSD^k yields the results shown in table 6.[17] Similar to the results in sections 2.3 and 3.3, pronounced type heterogeneity and a mode behavior of level-2 can be observed.

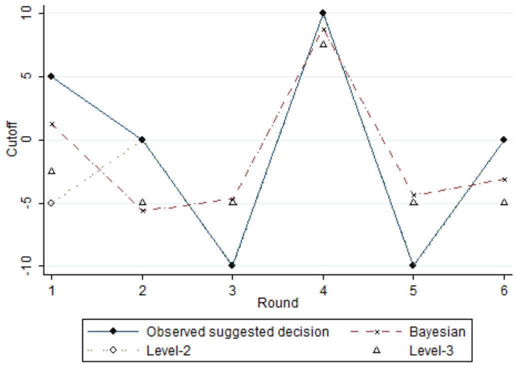
[17] All six decisions are included since the analysis of the first three decisions is less powerful to reject random play and yields significant results only for slightly more than 50% of the players. Results for the first three and second three decisions are reported in tables 18 and 19, the type distributions are not significantly different from each other (Wilcoxon rank-sum, $p = 0.703$).

		Round					
		1	2	3	4	5	6
	History	AAB	$BBAA$	$AAABA$	BBB	$AABA$	$BABA$
	Sugg. decision (mean)	0.17	0.18	-3.52	4.29	-2.50	-0.48
	Sugg. decision (s.d.)	4.20	3.95	5.26	5.43	4.84	3.02
	Level-1	0	0	0	0	0	0
Predicted	Level-2	-5	0	-10	10	-10	0
choice	Level-3	-2.5	-5	-5	7.5	-5	-5
	Bayesian	1.25	-5.625	-4.6875	8.75	-4.375	-3.125

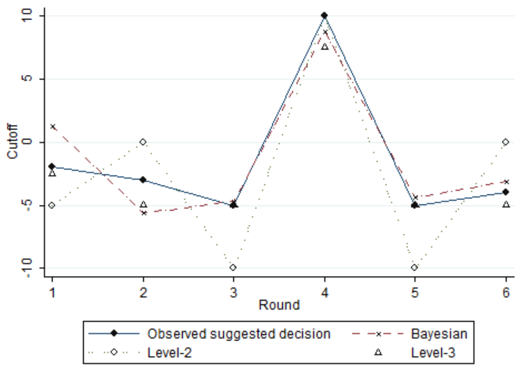
Table 5: Empirical and predicted decisions in the 6 rounds ($N = 42$)



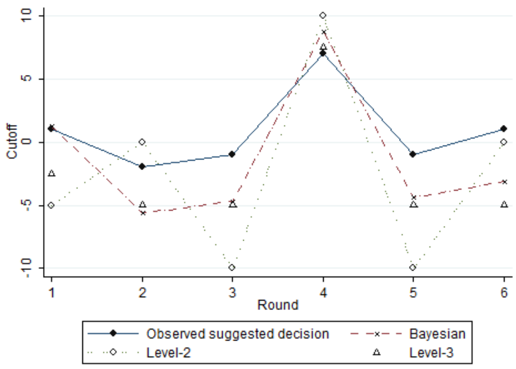
(a) Level-1 play (cutoff $= 0$, subject 36).



(b) Level-2 play (subject 5).



(c) Level-3 play (subject 26).



(d) Bayesian play (subject 12).

Figure 2
Subjects' and types' "fingerprints" in the team experiments à la ÇK

Messages. Qualitatively, the messages in this experiment are very similar to the ones reported in the AH setup.[18] The level classification yields the results shown in

[18] All messages are in German and are available from the author. Table 8 shows translated examples of very sophisticated ones.

Type	Frequency	Fraction (%)
1	10	23.8
2	15	35.7
3	3	7.1
B	4	9.5
NA	10	23.8
Total	42	100.00

(a) Classification by lowest significant SSD

Type	Frequency	Fraction (%)
1	10.40	24.8
2	13.22	31.5
3	4.41	10.5
B	3.98	9.5
NA	10.00	23.8
Total	42	100.00

(b) Classification by p-value below 0.05

Table 6: Type overview for 42 subjects

table 7 for the first three rounds, table 20 on page 41 shows similar results for the second three decisions. The reasoning is found to be heterogeneous and the mode behavior is level-2 to a similar extent as before. At this point, it can be summarized that – based on all analyses – hypotheses 1 and 2 can be evaluated positively.

Result 1 *Classified in terms of level-k types, reasoning in social learning situations is heterogeneous.*

Result 2 *The level-k distribution is hump-shaped and features level-2 reasoning as most common types.*

		Highest level ($\varnothing$ 2.03)					
		0	1	2	3	NA	Total
	0	1	0	1	0	0	2
Lowest level	1		3	0	0	0	3
($\varnothing$ 1.86)	2			18	3	0	21
	3				3	0	3
	NA					13	13
	Total	1	3	19	6	13	42

Table 7: Level classification in first three decisions, ÇK framework

Out of the 28.02 subjects that have been associated with a type 1, 2, or 3 in table 6b, 17.48 (62.4%) fall into the level intervals resulting from the classifications.[19] Given

[19]Table 24 on page 4 of the online appendix reports individual classifications based on messages and fingerprints.

that slightly more than 70% are classified in table 7, this value indicates a high level of agreement between the two methods. One interesting question relating to hypothesis 3 is at this point whether any message should be categorized as Bayesian reasoning?[20]

Both fingerprint analyses in the ÇK framework suggest the existence of Bayesian types. Two out of all six players with a level-3 classification potentially exhibit Bayesian reasoning. Their translated messages in the first three decisions are reported in table 8.

Subject 26 indeed clearly states in the first message the assumption that predecessors played rationally. He is calculating correctly the expected signals following the first two A's in the history, but makes a rough guess for the B in the third position. Since the sequences in decisions 2 and 3 are even longer, the subject simply takes into account that players positioned later in the sequence have a strong signal when they play against the majority. Although the message clearly reflects his assumption of rational preceding play, the fingerprint analysis identifies him as 80% level-3 and 20% Bayesian due to his imprecisions (figure 2c).

Subject 9 starts reasoning about her predecessors with the implicit assumption of rational play but forms expectations faultily. A signal between 0 and 10 results in an expected threshold of 2. In the second message, the reasoning is not differentiating predecessors at all and appears to be level-2. The very hesitant movements away from 0 make her fingerprint be classified as 53% level-1, with the remaining weight distributed across level-2, level-3 and Bayesian. Overall, there is some evidence from the communication for the intention of a Bayesian approach to the game, but the implementation is quite distinct.

The two analyses based on the intra-team communication allow me to evaluate hypothesis 3 as follows.

Result 3 *The communicated reasoning is generally not different from the theoretical*

[20]The question for omitted types can be investigated by a specification test that checks whether some subjects' fingerprints are closer to each other ("pseudotypes") than to a theoretical type, see appendix A.2. No missing type is found, only subjects which confuse signs (cluster D) and put too low thresholds throughout (cluster E).

type's reasoning. The truthful actions of players classified as level-0 make the case for a truthful level-0 type. Bayesian reasoning can be observed in fewer cases than expected from the fingerprint analysis.

4. Concluding remarks

This paper applies the level of reasoning concept of iterated beliefs to social learning and investigates experimentally the nature of social learning. Informative action data from the ÇK framework as well as incentivized written accounts of reasoning provide evidence of the heterogeneous types of reasoning known from strategic interaction. The predominant mode of reasoning is level-2 reasoning which is similar to Eyster and Rabin's (2010) BRTNI type that overinfers from signals.

The nature of inference observed in this study gives a simple level-2 explanation for the information cascades observed in Çelen and Kariv (2004). Furthermore, the heterogeneity observed provides the context for a fruitful interplay between truthful level-0 and level-1 players that make private information available on the one side and level-2+ players that use such information on the other side. Bohren (2015) and Acemoglu, Dahleh, Lobel and Ozdaglar (2011) describe theoretically that a heterogeneous population with private information revealers and naïve or rational Bayesians that make use of this information is able to accumulate enough information to certainly match the state of the world asymptotically. Such an interplay can explain why social learning among humans features self-correcting cascades as observed by Goeree et al. (2007).

Most likely, the level-k model is a very useful model of both strategic reasoning and inference because it captures the cognitive processes involved in thinking about the mental state of others. The communication data in this study shows how subjects first struggle to understand the situation for themselves (level-0) and then play without holding a definite model of others' reasoning (level-1). The simplest theory of mind is that others play truthfully (level-2). Also, the easiest way to obtain a new model

Subject ID	(H_t)	Message	Sugg. Decision	Classification
26	AAB	Hello. I propose that we assume that all participants before decided rationally. The sequence suggests that at action 1 $y > 0$. Expected value would so far be 5. Was this considered in action 2, the decision for all numbers greater than -5 would be A too. Expected value (5+2,5)/2=3,75. Was this considered in action 3, b was chosen only for strongly negative numbers. This is why I would propose b as well, that is to say a negative number.	-2	$\{3\}$
	BBAA	Decider 3 and 4 had more information, expected value consequently positive. Would therefore tend to A. At each $x > -3$ jump to A.	-3	$\{3\}$
	AAABA	Last participant has again most information. Apparently participant 4 had a strongly negative number. Still, I would be to chose A at negative numbers up to -5.	-5	$\{3\}$
9	AAB	Purely mathematically assumed: ;-) I am fourth and have information A,A,B. The first should have set 0 as marginal point to have a 50 50 chance. The second can thus be sure that the first had a positive number. He thus sets his point a bit higher at 2. With two A decisions the third can assume that both previous numbers were positive, so he can courageously bet on A with -2. This might be too courageous, he has maybe a negative number. Long speech, short meaning A with still high risk -2	-2	$\{2, 3\}$
	BBAA	I am fifth with B,B,A,A as information. This gives a 50:50 chance so I would guess a 0.	0	$\{2\}$
	AAABA	6. with information A,A,A,B,A assume: 1. chooses 0 → positive number, 2. chooses 0 as well → positive number, 3. becomes more courageous chooses -2 → possibly small negative number but in sum still everything positive, 4. still more courageous -3, now we would be back in the area around zero, so 5 chooses 0 again → has a positive number. Means we should be more courageous :D hehe I think this makes little sense	-3	$\{3\}$

Notes: "Classification" gives the set of possible levels k not ruled out by the lower and upper bounds.

Table 8: Examples of messages potentially reflecting Bayesian reasoning

of others' minds is to attribute ways of thinking to them that oneself just has, leading to the modeled degenerate population belief. This way level-3 players believe that some players might just follow the majority and hence do not act informatively. In uncovering these perceptions, the communication data in this study complements neuroeconomic data on brain activity which has as well been generated to understand how humans think about others' minds (Hampton et al., 2008; Coricelli and Nagel, 2009).

In general, these results highlight that for social learning in financial markets, takeover decisions, sequential voting, technology adoption, etc. to be successful, the decision-maker has to know how observed decisions have been made previously. Not only the quantity and quality of available information is relevant, but also the sophistication of the preceding decision-makers that linked information and decision.

References

Acemoglu, Daron, Munther A. Dahleh, Ilan Lobel, and Asuman Ozdaglar, "Bayesian Learning in Social Networks," *The Review of Economic Studies*, 2011, *78* (4), 1201–1236.

Anderson, Lisa R. and Charles A. Holt, "Information Cascades in the Laboratory," *American Economic Review*, December 1997, *87* (5), 847–62.

Arad, Ayala and Ariel Rubinstein, "The 11-20 Money Request Game: Evaluating the Upper Bound of k-Level Reasoning," *American Economic Review*, 2012, *102* (7).

Bikhchandani, Sushil, David Hirshleifer, and Ivo Welch, "A Theory of Fads, Fashion, Custom, and Cultural Change in Informational Cascades," *Journal of Political Economy*, October 1992, *100* (5), 992–1026.

Bohren, J. Aislinn, "Informational Herding with Model Misspecification," *Journal of Economic Theory*, 2015. forthcoming.

Brocas, Isabelle, Juan D. Carrillo, Stephanie W. Wang, and Colin F. Camerer, "Imperfect Choice or Imperfect Attention? Understanding Strategic Thinking in Private Information Games," *The Review of Economic Studies*, 2014.

Burchardi, Konrad B. and Stefan P. Penczynski, "Out of your mind: Eliciting individual reasoning in one shot games," *Games and Economic Behavior*, 2014, *84* (1), 39 – 57.

Camerer, Colin F., Teck-Hua Ho, and Juin-Kuan Chong, "A Cognitive Hierarchy Model of Games," *The Quarterly Journal of Economics*, August 2004, *119* (3), 861–898.

Çelen, Boğaçhan and Kyle Hyndman, "Social learning through endogenous information acquisition: an experiment," *Management Science*, 2012, *58* (8), 1525–1548.

__ **and Shachar Kariv**, "Distinguishing Informational Cascades from Herd Behavior in the Laboratory," *American Economic Review*, June 2004, *94* (3), 484–498.

Choi, Syngjoo, Douglas Gale, and Shachar Kariv, "Behavioral aspects of learning in social networks: an experimental study," *Advances in Applied Microeconomics*, 2005, *13*, 25–61.

Coricelli, Giorgio and Rosemarie Nagel, "Neural correlates of depth of strategic reasoning in medial prefrontal cortex," *Proceedings of the National Academy of Sciences*, 2009, *106* (23), 9163–9168.

Costa-Gomes, Miguel A. and Vincent P. Crawford, "Cognition and Behavior in Two-Person Guessing Games: An Experimental Study," *American Economic Review*, December 2006, *96* (5), 1737–1768.

__ , __ , and Bruno Broseta, "Cognition and Behavior in Normal-Form Games: An Experimental Study," *Econometrica*, September 2001, *69* (5), 1193–1235.

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

__ , Miguel A. Costa-Gomes, and Nagore Iriberri, "Structural Models of Nonequilibrium Strategic Thinking: Theory, Evidence, and Applications," *Journal of Economic Literature*, September 2013, *51* (1), 5–62.

Dominitz, Jeff and Angela A. Hung, "Homogeneous Actions and Hetergeneous Beliefs: Experimental Evidence on the Formation of Information Cascades," Econometric Society 2004 North American Winter Meetings 64, Econometric Society August 2004.

__ and __ , "Empirical models of discrete choice and belief updating in observational learning experiments," *Journal of Economic Behavior & Organization*, February 2009, *69* (2), 94–109.

Drehmann, Mathias, Jorg Oechssler, and Andreas Roider, "Herding and Contrarian Behavior in Financial Markets: An Internet Experiment," *American Economic Review*, December 2005, *95* (5), 1403–1426.

Eyster, Erik and Matthew Rabin, "Cursed Equilibrium," *Econometrica*, September 2005, *73* (5), 1623–1672.

__ and __ , "Naïve Herding in Rich-Information Settings," *American Economic Journal: Microeconomics*, November 2010, *2* (4), 221–43.

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

Goeree, Jacob K., Thomas R. Palfrey, Brian W. Rogers, and Richard D. McKelvey, "Self-Correcting Information Cascades," *Review of Economic Studies*, 07 2007, *74* (3), 733–762.

Guarino, Antonio and Philippe Jehiel, "Social learning with coarse inference," *American Economic Journal: Microeconomics*, 2013.

Hampton, Alan N., Peter Bossaerts, and John P. O'Doherty, "Neural correlates of mentalizing-related computations during strategic interactions in humans.," *Proceedings of the National Academy of Sciences of the United States of America*, May 2008, *105* (18), 6741–6746.

Hung, Angela A. and Charles R. Plott, "Information Cascades: Replication and an Extension to Majority Rule and Conformity-Rewarding Institutions," *American Economic Review*, December 2001, *91* (5), 1508–1520.

Jehiel, Philippe, "Analogy-based expectation equilibrium," *Journal of Economic*

Theory, 2005, *123* (2), 81 – 104.
Kübler, Dorothea and Georg Weizsäcker, "Limited depth of reasoning and failure of cascade formation in the laboratory," *Review of economic studies*, 2004, *71*, 424–441.
McKelvey, Richard D. and Thomas R. Palfrey, "Quantal Response Equilibria for Normal Form Games," *Games and Economic Behavior*, July 1995, *10* (1), 6–38.
Nagel, Rosemarie, "Unraveling in Guessing Games: An Experimental Study," *American Economic Review*, December 1995, *85* (5), 1313–1326.
Smith, Lones and Peter Sørensen, "Pathological Outcomes of Observational Learning," *Econometrica*, March 2000, *68* (2), 371–398.
Stahl, Dale O. and Paul W. Wilson, "On Players' Models of Other Players: Theory and Experimental Evidence," *Games and Economic Behavior*, July 1995, *10* (1), 218–254.
Weizsäcker, Georg, "Do We Follow Others When We Should? A Simple Test of Rational Expectations," *American Economic Review*, December 2010, *100* (5), 2340–60.
Ziegelmeyer, Anthony, Frédéric Koessler, Juergen Bracht, and Eyal Winter, "Fragility of information cascades: an experimental study using elicited beliefs," *Experimental Economics*, 2010, *13* (2), 121–145.

A. Appendix

A.1. Messages by classified level (AH)

Subject ID	(H_t, s_t)	Message	Sugg. Decision	Confidence	Classification
27	BBAb	I don't know this is all probability so this is a guess	B	50	–
	BAABAb	I have no clue.	B	50	$\{0\}$
	BAAa	I have no clue.	A	50	$\{0\}$
51	BBAb	randomly choose my selection. i will choose B, u select A and computer will take decison randomly on behalf of us	B	50	$\{0\}$
	BAABAb	randomly choose my selection. i will choose B, u select A and computer will take decison randomly on behalf of us	B	50	$\{0\}$
	BAAa	randomly choose my selection. i will choose B, u select A and computer will take decison randomly on behalf of us	B	50	$\{0\}$
79	BAAAb	idk, just a guess	B	50	$\{0\}$
	ABAAAb	just a guess	B	50	$\{0\}$
	BBABa	just a guess	A	50	$\{0\}$

Notes: "Classification" gives the set of possible levels k not ruled out by the lower and upper bounds. A missing classification is indicated with –.

Table 9: Examples of messages classified as level-0

Subject ID	(H_t, s_t)	Message	Sugg. Decision	Confidence	Classification
58	BBAb	since the private draw is black, urn b is more likely to be the true urn.	B	50	$\{1\}$
71	BAAAb	The ball picked was black, right? So, there's a greater chance that it came from urn B than urn A	B	67	$\{1\}$
	ABAAAb	The ball picked was black. There are two black balls in urn B and only one black ball in urn A. Therefore, there's a 2/3 chance the ball came from urn B and a 1/3 chance it came from urn A.	B	66	$\{1\}$
	BBABa	Urn A contains two white balls and urn B contains one. Therefore, there's a 2/3 chance the ball came from urn A.	A	67	$\{1\}$
72	BAAAb	It would be foolish not to choose the urn with the greater proportion of black balls.	B	67	$\{1\}$
81	BAAAb	it doesnt matter what the other teams found there is still a higher probability that a black ball is chosen from the urb B therefore i would choose b	B	66	$\{1\}$
	ABAAAb	there is still a higher chance that the black ball is chosen (2/3 versus 1/3) therefore i would still choose B it doesnt matter what the other teams chose	B	66	$\{1\}$
	BBABa	it doesnt matter the order of sequence of the other teams i would still choose A since the probability of choosing a white ball is twice that of a black ball (2/3 vs. 1/3) therefore, i would still choose A	A	66	$\{1\}$
84	BAAAb	I don't see how the previous draws affect anything. We got black, it's 2/3 chance from B, 1/3 chance from A unless I'm completely confused.	B	67	$\{1\}$
	ABAAAb	2/3 chance that black ball is from B, 1/3 chance from A	B	67	$\{1\}$
85	BAAAb	there is a 66.7% chance that the ball is black. i'm not sure how the previous draws would affect our guess	B	66	$\{1\}$
	ABAAAb	I think it would be the best to stick with urn B since the probability is in favor of that.	B	66	$\{1\}$
	BBABa	I went with A because the chances it came from A are more likely that B.	A	66	$\{1\}$
97	BAAAb	If the other teams are mostly going with A, and we have a black ball, I think our best choice is B.	B	–	$\{1\}$
	ABAAAb	I think we have a better chance selecting urn B.	B	50	–
	BBABa	I think since we are drawing a white ball, our best chance is to select from urn A.	A	100	$\{1\}$
98	BAAAb	Urn B has the most black balls	B	60	–
	ABAAAb	Even though th teams before mostly picked A, I don't think that should have a bearing on the choice here; B still has the higher statistical probability of having a black ball	B	60	$\{1\}$

Notes: "Classification" gives the set of possible levels k not ruled out by the lower and upper bounds. A missing classification is indicated with –.

Table 10: Examples of messages classified as level-1

Subject ID	(H_t, s_t)	Message	Sugg. Decision	Confidence	Classification
1	BBAb	It is more likely that a black ball came from Urn B, and 2/3 of the other teams have made this same choice.	B	75	$\{2\}$
	BAABAb	A black ball is more likely to have come from Urn B. Not enough teams have gone before us to be affected by the fact that 3/5 of them selected A	B	60	$\{2\}$
	BAAa	A white ball is more likely to be from Urn A, and 2/3 of the previous teams have selected this urn as well	A	75	$\{2\}$
8	BBAb	Since 2 out of 3 of the previous teams chose urn B, and our draw corresponds to that choice, I'd say that urn b is a good bet.	B	70	$\{2\}$
	BAABAb	this is a tough one, but based on our draw, B is a good choice. Also, 2 out of 5 other teams chose B	B	55	$\{2\}$
	BAAa	2 out of 3 previous teams chose A and our draw corresponds to A as well.	A	70	$\{2\}$
56	BAAa	I'm choosing this based on the other teams and our draw. Also, I think I forgot to press enter for the last message... So for the last one, I made that decision based on the other teams draws not ours.	A	75	$\{2\}$
80	BAAAb	3 out of 4 previous teams are picking A so that is more likely they probably decide to pick A because see a white ball drawn	A	90	$\{2\}$
	ABAAAb	there are 5 teams before us and 4 of them choose A, so i am confident that it is A most of them must see a white ball from A	A	98	$\{2\}$
	BBABa	4 teams before us and 3 of them picked B so I would pick B	B	85	$\{2\}$
82	BAAAb	Most people must have seen a white draw for there to be predominatley A being chosen.	A	75	$\{2\}$
	ABAAAb	It seems that most people saw a white ball in their private draw.	A	60	$\{2\}$
	BBABa	Hard to say, still though most people seem to have seen a black ball.	B	55	$\{2\}$
86	BAAAb	Only because 3 teams chose A	A	–	$\{2\}$
	ABAAAb	4 out of 5 say A	A	60	$\{2\}$
88	BAAAb	3 white balls in a row from the black urn is only a (1/3)times(1/3) times (1/3) chance - therefore, very low. I suspect that people chose urn A if they got a white ball	A	75	$\{2\}$
	ABAAAb	I assume teams chose urn A if they drew white. The chance of getting 4 whites from Urn B is very low therefore I believe it is urn A. Even though we drew a black, it's still greater probability that it is A	A	75	$\{2\}$
	BBABa	It appears from previous teams choices that 3 of the 4 balls were black. Therefore, even though we drew a white ball, it's more likely that our urn is B	B	70	$\{2\}$

Notes: "Classification" gives the set of possible levels k not ruled out by the lower and upper bounds.

Table 11: Examples of messages classified as level-2 (continued next table)

Subject ID	(H_t, s_t)	Message	Sugg. Decision	Confidence	Classification
100	BAAAb	since 3 of the teams before us chose A, we might assume that they got a white ball. we may have just gotten a black ball by chance from urn A	A	80	$\{2\}$
	ABAAAb	since most teams before us have chosen A, we may assume that most of them got the white ball from their drawing. using probability, it's more likely that we got a black by with 1/3 chance than the previous teams getting a white ball from urn B with much lower probability	A	80	$\{2\}$
	BBABa	we can assume that team 1 got the black ball, team 2 probably also got the black ball, and team 3 probably got the white ball. probability wise, it's more likely that its from urn B	B	80	$\{2\}$
101	BAAAb	Okay so the teams before us chose A more frequently, which means they private draw must have been white, with one black. We got a black one, but since its a 50% chance that both urns could be the true urn, we can attribute ours to simple probability Thus, we have a 75% chance of it being A More frequency-better chance it comes from one urn than the other	A	75	$\{2\}$
	ABAAAb	Frequency of a is greater than B, thus, it's a good chance its urn A because although we picked black 4 whites to two blacks shows a 2 to 1 ratio meaning that for everty 3 balls, 2 are white and one is black this reflects the probabilty of urn A	A	90	$\{2\}$
	BBABa	The first team must have seen a black ball and chose B second team probably got black as well so chose B third team probably got white and decided to go with A 4th team got black again and went with B Thus if we assume all the choices before are rational then the choices are more black than white thus B MUST be the rationally right answer as in if you don't pick B, it would be quite irrational	B	80	$\{2\}$
106	BAAAb	Given that we were shown a black ball, there is a 2/3 chance that it is in B. However, the other teams chose A, which means that they were given a white ball, meaning that there is a greater chance of A	A	90	$\{2\}$

Notes: "Classification" gives the set of possible levels k not ruled out by the lower and upper bounds.

Table 12: Examples of messages classified as level-2

Subject ID	(H_t, s_t)	Message	Sugg. Decision	Confidence	Classification
15	BBAb	The first team definitely saw a black ball, they went in blind and chose B The second team must have also saw a black ball, and corroborated with what team 1 saw, the 3rd team saw a white ball, which contradicted what the previous team saw It is possible that the 2nd team saw white, but just went along with team 1 so, I would say, out of 4 draws, there's certainly 1 black, 1 maybe black, and 1 white our draw is a black one I am pretty confident that we should vote on urn B	B	100	$\{3\}$
	BAABAb	team 1 saw black team 2 saw white, for sure team 3 saw white for sure, since it could be influenced by either team 1 or 2, but chose to go with white team 4 saw black team 5 saw white team 5 has 2 guaranteed black before it, and 2 guaranteed white before it, but chose white, so probably got white our draw is black, there's 3 white, 2 black before us I'd say its 50 50, let's go with black, urn B since we're sure about what we saw	B	70	$\{2, 3\}$
	BAAa	same logic, lets go with A since majority came out as white	A	100	$\{2\}$
46	BBAb	Teams 1 and 2 probably saw a black ball, and team 3 a white ball 3 blacks and 1 white is much more likely for urn b	B	75	$\{2\}$
	BAABAb	Team 1 probably drew a black ball, as they have no information Team 2 must have drawn white, if they chose A despite team 1's decision to choose black Team 3 probably drew white as well, as the previous teams decisions are split so they decide on their own team 4 must have drawn black, as them drawing white would be totally agains ttheir choice to choose B It's 50-50	A	50	$\{2, 3\}$
	BAAa	draws are almost purely based on the individual ball draw, as the disparity of information in the teams previous theirs indicates 3 white balls to 1 black ball, basically	A	75	–
74	ABAAAb	It seems the other teams saw more white balls though it is possible the third and the fourth team actuall saw black balls but they had to make a random choince.	A	50	$\{3\}$
	BBABa	team 1 and team 4 drew black balls and team 3 and team 5 drew white. therefore it is a 50/50 chance.	A	50	$\{0\}$

Notes: "Classification" gives the set of possible levels k not ruled out by the lower and upper bounds. A missing classification is indicated with –.

Table 13: Examples of messages classified as level-3 (continued next table)

Subject ID	(H_t, s_t)	Message	Sugg. Decision	Confidence	Classification
93	BAAAb	Well, from our observations it seems that we would be choosing Urn B. However, if we look at previous guesses, a desire may be made to choose Urn A from belieivng that other people must be drawing white balls. We have to consider the fact that other teams are also being influenced by previous decisions. For example, team 4 may have drawn a black ball and wanted to choose Urn B, but didn't in order to match the previous people. Therefore, I still think we should choose Urn B.	B	50	$\{3\}$
	ABAAAb	Even though we drew a black ball, feel that for 80% of people to have chosen A, this may indicate that white is the majority. However, I'm not sure at all, due to previous people also being influenced by previous decisions. We know that the first decision is not based off of previous decisions, therefore we can assume that the first person drew a white ball. The second person contrasted the first person so we can assume they went by their ball and chosen B. The third person was going off of a balanced group where half of the people had chosen whtie and half of the people had chosen black, therefore their decision is valid too. After that it was based on how people analyzed it somewhat. I believe the fourth person would still have stayed with what they got so I think A is better.	A	50	$\{3\}$
	BBABa	We know that the first person would stick to what ball they drew, therefore it can be decided to be B. The second person would have no reason to follow the first person since its just one statistic so they are probably accurate too. The third person needed a reason to deviate so they were definitely A. The last person probably would have stayed with what they got so I think it was B, regardless of our drawing.	B	70	$\{2\}$

Notes: "Classification" gives the set of possible levels k not ruled out by the lower and upper bounds.

Table 14: Examples of messages classified as level-3

A.2. Specification Test

Costa-Gomes and Crawford (2006) check for unknown types by testing whether some subjects' fingerprints are closer to each other ("pseudotypes") than to a theoretical type. Such "clusters" of multiple subjects could suggest a common, but not predicted way of thinking. As opposed to the ÇK data analyzed in section 3.3, my experiment confronts all subjects with the same six situations, enabling such a specification test.

Costa-Gomes and Crawford (2006) define a cluster as "a group of two or more subjects such that: (a) each subject's original estimated type has smaller likelihood than the pseudotypes of all other subjects in the group; and (b) all subjects in the group make 'sufficiently similar' guesses" (pp. 1761-1762). I correspondingly define a cluster as a group of two or more subjects such that: (a) the p-value of the originally estimated theoretical type is higher than the p-values with respect to the pseudotypes of all other subjects in the group; and (b) all subjects in the group make guesses that are significantly closer to each other than random play ($p < 0.05$).

The analysis yields six clusters as shown in table 15. Two clusters feature regularities that might explain some of the discrepancy between messages and decisions as mere errors. Cluster D features four players who persistently put the wrong sign on their thresholds, thus exhibiting probably the most simple and extreme case of decision error (see figure 3a). Reversing the sign of their thresholds leads to a partial agreement with the messages that categorized some as level-1 or level-2. Relatedly, cluster E features four players who share a pronounced tendency towards negative numbers which is probably due to a faulty calculation not featured by other players (see figure 3b).

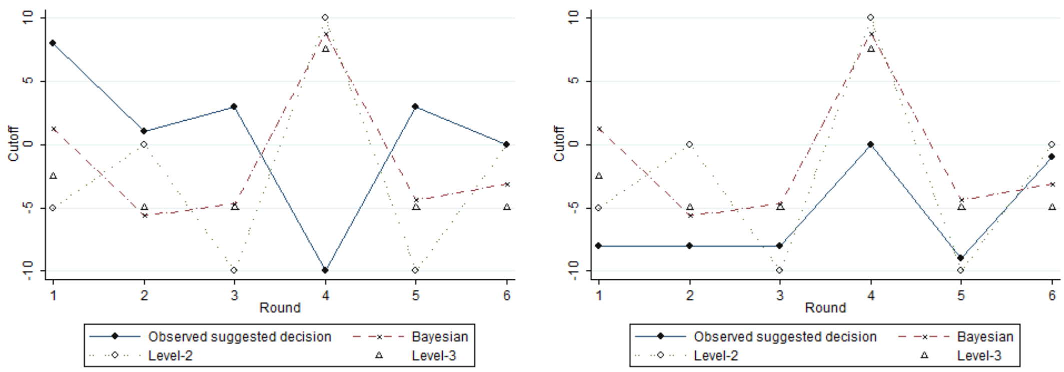
Featuring less persistent regularities, clusters B, C, and F consist of players who are significantly closer than random play only to level-2 play. Players in cluster B and C start off with higher thresholds than level-2 in round 1, in cluster B they also choose a smaller fourth threshold. Players in cluster F start off with lower thresholds in round 1. Overall, all those clusters seem to form due to more or less persistent regularities, which probably do not result from an entirely different kind of thinking.

Although the decisions were chosen to discriminate as much as possible between level-2, level-3 and Bayesian types, the specification test yields one large set A of overlapping and non-nested clusters of 19 players. Most of these players are classified by the fingerprints and messages as level-2, level-3 or Bayesian. While the fingerprint data is theoretically able to discriminate between these types, decision noise partially prevents this in an analysis with only 6 decisions. In that instance the messages are more informative.

Cluster	Subjects	Characteristic
A	2, 6, 7, 9, 12, 13, 14, 16, 21, 22, 23, 27, 29, 33, 34, 37, 38, 39, 42	Level-2, Level-3 or Bayesian
B	4, 17	Level-2, deviation round 1 (pos.), round 4 (neg.)
C	5, 28, 31	Level-2, deviation round 1 (pos.)
D	11, 19, 35, 40	Switched sign
E	15, 18, 32, 41	Neg. deviation throughout
F	20, 30	Level-2, deviation round 1 (neg.)

Notes: A is not a proper cluster, it is the union of overlapping, non-nested clusters.

Table 15: Clusters in the specification test



(a) Mistaken sign, subject 35.

(b) Negative deviation thoughout, subject 15.

Figure 3
Subjects' and types' "fingerprints" in the team experiments à la ÇK

A.3. Additional tables

A.3.1. AH framework with intra-team communication

		Highest level (⌀ 1.76)					
		0	1	2	3	NA	Total
	0	5	0	6	0	0	11
Lowest level	1		13	5	0	0	18
(⌀ 1.50)	2			42	3	0	45
	3				2	0	2
	NA					30	30
	Total	5	13	53	5	30	106

Table 16: Level classification in second three decisions, AH framework

	Urn choice					
	1	2	3	4	5	6
Information	$BBAb$	$BAABAb$	$BAAa$	$BBABa$	$ABAAAb$	$BAAAb$
Level-0	5	7	7	8	5	8
Level-1	5	10	3	12	16	17
Level-2	40	29	44	39	40	36
Level-3	3	2	3	1	4	4
NA	53	58	49	46	41	41
Total	106	106	106	106	106	106

Table 17: Level classification by urn choice, AH framework

A.3.2. ÇK framework with intra-team communication

Type	Frequency	Fraction (%)
(0)	17	40.5
1	12	28.6
2	10	23.8
3	2	4.8
B	1	2.4
Total	42	100.00

(a) Classification by lowest significant SSD

Type	Frequency	Fraction (%)
(0)	17.00	40.5
1	12.00	28.6
2	10.00	23.8
3	1.89	4.5
B	1.11	2.6
Total	42	100.00

(b) Classification by p-value below 0.05

Table 18: Type overview for 42 subjects in first three decisions, ÇK framework

Type	Frequency	Fraction (%)
(0)	18	42.9
1	9	21.4
2	8	19.0
3	0	0.0
B	7	16.7
Total	42	100.00

(a) Classification by lowest significant SSD

Type	Frequency	Fraction (%)
(0)	18.00	42.9
1	9.00	21.4
2	8.55	20.4
3	0.33	0.8
B	6.11	14.6
Total	42	100.00

(b) Classification by p-value below 0.05

Table 19: Type overview for 42 subjects in second three decisions, ÇK framework

		Highest level ($\varnothing$ 2.11)					
		0	1	2	3	NA	Total
	0	1	0	1	2	0	4
Lowest level	1		2	2	0	0	4
($\varnothing$ 1.68)	2			15	2	0	17
	3				3	0	3
	NA					14	14
	Total	1	2	18	7	14	42

Table 20: Level classification in second three decisions, ÇK framework

B. Online Appendix

B.1. Individual data

Subject ID	Level 1 SSD	Level 1 p	Level 2 SSD	Level 2 p	Level 3 SSD	Level 3 p	Bayesian SSD	Bayesian p
1	549.13	0.671	1475.83	0.801	1143.33	0.750	1255.61	0.697
2	522.00	0.585	1022.00	0.183	870.75	0.199	850.50	0.202
3	714.61	0.964	946.61	0.095	777.61	0.224	823.28	0.140
4	331.00	0.068	131.00	0.000	134.75	0.000	99.08	0.000
5	1500.00	1.000	375.00	0.000	656.25	0.107	610.19	0.040
6	289.86	0.029	1020.36	0.237	796.35	0.306	721.46	0.178
7	725.00	0.970	200.00	0.000	368.75	0.004	441.56	0.012
8	700.00	0.954	25.00	0.000	268.75	0.003	170.86	0.000
9	255.00	0.012	320.00	0.000	203.75	0.000	298.23	0.001
10	627.00	0.862	97.00	0.000	245.75	0.001	177.84	0.000
11	687.17	0.943	275.17	0.000	156.42	0.000	168.43	0.000
12	423.00	0.260	818.00	0.027	426.75	0.009	595.65	0.016
13	661.25	0.914	51.25	0.000	117.50	0.000	97.09	0.000
14	400.01	0.197	450.01	0.002	363.01	0.003	494.64	0.006
15	838.61	0.997	394.61	0.001	310.11	0.002	318.56	0.001
16	883.00	0.999	158.00	0.000	289.25	0.002	305.77	0.001
17	0.02	0.000	723.65	0.059	536.24	0.034	529.73	0.033
18	100.00	0.000	475.00	0.014	243.75	0.001	269.21	0.002
19	186.39	0.001	548.09	0.023	555.54	0.041	575.62	0.055
20	224.69	0.005	485.09	0.004	316.89	0.001	328.71	0.002
21	524.25	0.593	120.25	0.000	140.75	0.000	112.44	0.000
22	391.00	0.175	656.00	0.035	528.50	0.042	587.05	0.042
23	421.00	0.254	141.00	0.000	137.25	0.000	124.54	0.000
24	242.25	0.008	52.25	0.000	162.25	0.000	115.79	0.000
25	308.25	0.043	743.25	0.012	415.75	0.007	607.40	0.013
26	1386.52	1.000	783.02	0.062	929.27	0.552	936.15	0.324
27	1171.15	1.000	1182.65	0.163	845.15	0.150	947.88	0.159
28	444.25	0.324	194.25	0.000	209.25	0.000	214.01	0.000
29	1028.33	1.000	258.33	0.000	490.58	0.006	389.41	0.001
30	514.66	0.561	602.66	0.019	455.91	0.019	456.18	0.010
31	893.26	0.999	537.26	0.003	538.26	0.030	522.84	0.008
32	619.25	0.847	939.25	0.043	684.25	0.073	834.31	0.064
33	5.00	0.000	885.00	0.093	652.50	0.054	740.49	0.071
34	396.98	0.190	1331.88	0.579	1110.53	0.705	1001.84	0.453
35	899.00	0.999	389.00	0.001	521.50	0.045	494.42	0.016
36	500.00	0.511	175.00	0.000	243.75	0.001	168.24	0.000
37	197.13	0.002	659.63	0.011	268.38	0.001	475.27	0.005
38	103.00	0.000	768.00	0.156	545.50	0.127	723.30	0.202
39	1500.00	1.000	575.00	0.013	881.25	0.444	669.35	0.066
40	807.11	0.995	505.11	0.023	727.31	0.309	665.61	0.174

Table 21: Type analysis in data of Çelen and Kariv (2004)

		Decision 1 (BBAb)					Decision 2 (BAABAb)					Decision 3 (BAAa)					Decision 4 (BBABa)					Decision 5 (ABAAAb)					Decision 6 (BAAAb)				
Session	Subject ID	Sugg. dec.	Confidence	Fingerprint	RA 1	RA 2	Sugg. dec.	Confidence	Fingerprint	RA 1	RA 2	Sugg. dec.	Confidence	Fingerprint	RA 1	RA 2	Sugg. dec.	Confidence	Fingerprint	RA 1	RA 2	Sugg. dec.	Confidence	Fingerprint	RA 1	RA 2	Sugg. dec.	Confidence	Fingerprint	RA 1	RA 2
1	1	B	75	{0, 1, 2, 3}	2	2	B	60	{0, 1, 2, 3}	2	2	A	75	{0, 1, 2, 3}	2	2	A	60	{0, 1}	2	1	B	75	{0, 1}	1	1	B	60	{0, 1}	1	1
1	2	B	0	{0, 1, 2, 3}	2	2	A	50	{0, 2, 3}	.	0	A	50	{0, 1, 2, 3}	.	.	B	50	{0, 2, 3}	2	2	A	50	{0, 2, 3}	.	.	A	0	{0, 2, 3}	.	.
1	3	B	75	{0, 1, 2, 3}	2	2	B	75	{0, 1, 2, 3}	2	2	A	75	{0, 1, 2, 3}	2	2	B	80	{0, 2, 3}	2	2	A	60	{0, 2, 3}	0	2	A	50	{0, 2, 3}	.	.
1	4	B	70	{0, 1, 2, 3}	.	.	B	70	{0, 1, 2, 3}	.	.	A	70	{0, 1, 2, 3}	2	1	B	50	{0, 2, 3}	.	.	B	70	{0, 1}	1	1	A	55	{0, 2, 3}	2	2
1	5	B	80	{0, 1, 2, 3}	2	1	B	80	{0, 1, 2, 3}	1	1	A	80	{0, 1, 2, 3}	2	1	A	80	{0, 1}	1	1	B	80	{0, 1}	1	1	B	80	{0, 1}	1	1
1	6	B	75	{0, 1, 2, 3}	0	0	B	75	{0, 1, 2, 3}	.	1	A	75	{0, 1, 2, 3}	2	2	B	65	{0, 2, 3}	2	2	A	65	{0, 2, 3}	2	2	A	65	{0, 2, 3}	2	2
1	7	B	95	{0, 1, 2, 3}	2	2	A	50	{0, 2, 3}	0	2	A	90	{0, 1, 2, 3}	2	2	B	75	{0, 2, 3}	2	2	A	70	{0, 2, 3}	2	2	A	70	{0, 2, 3}	2	2
1	8	B	70	{0, 1, 2, 3}	2	2	B	55	{0, 1, 2, 3}	2	2	A	70	{0, 1, 2, 3}	2	2	B	50	{0, 2, 3}	2	2	A	55	{0, 2, 3}	2	2	A	60	{0, 2, 3}	2	2
1	9	B	67	{0, 1, 2, 3}	2	1	B	67	{0, 1, 2, 3}	.	1	A	67	{0, 1, 2, 3}	.	.	A	67	{0, 1}	1	1	B	67	{0, 1}	1	1	B	67	{0, 1}	1	1
1	10	B	0	{0, 1, 2, 3}	.	.	A	0	{0, 2, 3}	0	2	A	0	{0, 1, 2, 3}	.	.	B	0	{0, 2, 3}	2	2	A	0	{0, 2, 3}	2	2	A	0	{0, 2, 3}	2	2
1	11	B	70	{0, 1, 2, 3}	2	1	B	50	{0, 1, 2, 3}	.	0	A	60	{0, 1, 2, 3}	2	2	B	60	{0, 2, 3}	2	2	A	60	{0, 2, 3}	2	2	A	60	{0, 2, 3}	2	2
1	12	B	50	{0, 1, 2, 3}	.	.	A	50	{0, 2, 3}	.	.	B	60	{0}	.	.	A	50	{0, 1}	.	0	B	60	{0, 1}	1	1	B	50	{0, 1}	.	.
2	13	B	0	{0, 1, 2, 3}	2	2	B	0	{0, 1, 2, 3}	2	1	A	100	{0, 1, 2, 3}	2	2	B	55	{0, 2, 3}	2	2	A	100	{0, 2, 3}	2	2	A	60	{0, 2, 3}	2	2
2	14	B	70	{0, 1, 2, 3}	2	1	B	70	{0, 1, 2, 3}	2	1	A	70	{0, 1, 2, 3}	2	1	A	70	{0, 1}	.	.	B	60	{0, 1}	.	.	B	60	{0, 1}	.	.
2	15	B	100	{0, 1, 2, 3}	3	3	B	70	{0, 1, 2, 3}	2	3	A	100	{0, 1, 2, 3}	2	2	B	60	{0, 2, 3}	2	2	A	100	{0, 2, 3}	2	2	A	70	{0, 2, 3}	2	2
2	16	B	0	{0, 1, 2, 3}	2	1	B	65	{0, 1, 2, 3}	0	0	A	60	{0, 1, 2, 3}	0	0	A	67	{0, 1}	1	0	B	67	{0, 1}	.	.	B	67	{0, 1}	0	.
2	17	B	0	{0, 1, 2, 3}	.	.	A	50	{0, 2, 3}	.	.	A	50	{0, 1, 2, 3}	0	0	B	0	{0, 2, 3}	.	.	B	50	{0, 1}	.	0	B	50	{0, 1}	.	.
2	18	B	75	{0, 1, 2, 3}	2	2	B	50	{0, 1, 2, 3}	.	0	A	75	{0, 1, 2, 3}	.	.	B	50	{0, 2, 3}	2	2	A	50	{0, 2, 3}	2	2	B	50	{0, 1}	3	3
2	19	B	80	{0, 1, 2, 3}	2	2	B	50	{0, 1, 2, 3}	2	2	A	80	{0, 1, 2, 3}	2	2	B	60	{0, 2, 3}	2	2	A	75	{0, 2, 3}	2	2	A	75	{0, 2, 3}	2	2
2	20	B	85	{0, 1, 2, 3}	2	2	B	50	{0, 1, 2, 3}	.	2	A	80	{0, 1, 2, 3}	2	2	B	55	{0, 2, 3}	2	2	A	60	{0, 2, 3}	2	2	A	60	{0, 2, 3}	2	2
2	21	B	69	{0, 1, 2, 3}	.	.	A	69	{0, 2, 3}	.	0	A	69	{0, 1, 2, 3}	.	.	B	69	{0, 2, 3}	.	.	A	99	{0, 2, 3}	.	.	A	0	{0, 2, 3}	.	.
2	22	B	80	{0, 1, 2, 3}	.	.	A	50	{0, 2, 3}	.	.	A	60	{0, 1, 2, 3}	.	.	B	70	{0, 2, 3}	.	.	A	60	{0, 2, 3}	.	.	A	60	{0, 2, 3}	.	.
2	23	B	67	{0, 1, 2, 3}	.	.	B	67	{0, 1, 2, 3}	.	.	A	67	{0, 1, 2, 3}	.	.	A	75	{0, 1}	.	.	B	70	{0, 1}	.	.	B	67	{0, 1}	.	.
2	24	B	70	{0, 1, 2, 3}	2	2	A	70	{0, 2, 3}	2	0	A	70	{0, 1, 2, 3}	2	2	B	60	{0, 2, 3}	2	2	A	80	{0, 2, 3}	2	2	A	70	{0, 2, 3}	2	2
2	25	B	75	{0, 1, 2, 3}	2	2	A	65	{0, 2, 3}	2	2	A	80	{0, 1, 2, 3}	2	2	B	65	{0, 2, 3}	2	2	A	85	{0, 2, 3}	2	2	A	75	{0, 2, 3}	2	2
2	26	B	66	{0, 1, 2, 3}	1	1	A	50	{0, 2, 3}	2	2	A	66	{0, 1, 2, 3}	2	2	B	60	{0, 2, 3}	2	2	A	60	{0, 2, 3}	2	2	A	60	{0, 2, 3}	2	2
2	27	B	50	{0, 1, 2, 3}	0	.	B	50	{0, 1, 2, 3}	0	0	A	50	{0, 1, 2, 3}	0	0	A	50	{0, 1}	0	0	B	50	{0, 1}	.	.	B	50	{0, 1}	0	0
2	28	B	50	{0, 1, 2, 3}	2	2	B	60	{0, 1, 2, 3}	.	.	A	50	{0, 1, 2, 3}	.	.	A	70	{0, 1}	.	.	B	70	{0, 1}	.	.	B	70	{0, 1}	.	.
2	29	B	75	{0, 1, 2, 3}	2	2	A	60	{0, 2, 3}	2	2	A	75	{0, 1, 2, 3}	2	2	B	70	{0, 2, 3}	2	2	A	70	{0, 2, 3}	2	2	A	50	{0, 2, 3}	0	0
2	30	B	70	{0, 1, 2, 3}	.	.	B	60	{0, 1, 2, 3}	.	.	A	65	{0, 1, 2, 3}	.	.	A	60	{0, 1}	.	.	A	70	{0, 2, 3}	2	2	A	60	{0, 2, 3}	2	2
3	31	B	70	{0, 1, 2, 3}	.	.	B	70	{0, 1, 2, 3}	.	.	B	0	{0}	.	.	B	80	{0, 2, 3}	2	2	A	90	{0, 2, 3}	2	2	A	80	{0, 2, 3}	.	.
3	32	B	0	{0, 1, 2, 3}	.	.	B	66	{0, 1, 2, 3}	.	.	A	66	{0, 1, 2, 3}	.	.	A	66	{0, 1}	.	.	B	67	{0, 1}	.	.	B	66	{0, 1}	.	.
3	33	B	50	{0, 1, 2, 3}	0	0	B	80	{0, 1, 2, 3}	.	.	A	55	{0, 1, 2, 3}	0	0	A	65	{0, 1}	.	.	A	88	{0, 2, 3}	.	.	A	75	{0, 2, 3}	.	.
3	34	B	50	{0, 1, 2, 3}	.	.	A	60	{0, 2, 3}	.	.	A	50	{0, 1, 2, 3}	.	.	A	50	{0, 1}	.	.	B	50	{0, 1}	.	.	B	50	{0, 1}	.	.
3	35	B	67	{0, 1, 2, 3}	.	1	B	67	{0, 1, 2, 3}	.	.	A	67	{0, 1, 2, 3}	.	.	A	67	{0, 1}	.	.	B	67	{0, 1}	.	.	B	67	{0, 1}	2	.
3	36	B	90	{0, 1, 2, 3}	2	2	B	50	{0, 1, 2, 3}	.	2	A	90	{0, 1, 2, 3}	2	2	B	51	{0, 2, 3}	2	2	A	60	{0, 2, 3}	2	2	A	55	{0, 2, 3}	2	2
3	37	B	50	{0, 1, 2, 3}	.	.	A	50	{0, 2, 3}	.	.	A	50	{0, 1, 2, 3}	.	.	B	50	{0, 2, 3}	2	2	A	50	{0, 2, 3}	2	2	A	50	{0, 2, 3}	.	2
3	38	B	66	{0, 1, 2, 3}	2	2	B	50	{0, 1, 2, 3}	2	2	A	75	{0, 1, 2, 3}	2	2	B	60	{0, 2, 3}	2	2	A	70	{0, 2, 3}	2	2	A	70	{0, 2, 3}	2	2
3	39	B	0	{0, 1, 2, 3}	.	.	A	0	{0, 2, 3}	.	2	A	80	{0, 1, 2, 3}	2	2	A	50	{0, 1}	.	.	A	50	{0, 2, 3}	.	.	B	50	{0, 1}	0	0
3	40	B	75	{0, 1, 2, 3}	2	2	B	50	{0, 1, 2, 3}	2	2	A	90	{0, 1, 2, 3}	2	1	B	75	{0, 2, 3}	2	2	A	90	{0, 2, 3}	2	2	A	60	{0, 2, 3}	2	2
3	41	B	75	{0, 1, 2, 3}	2	2	A	60	{0, 2, 3}	2	2	A	60	{0, 1, 2, 3}	2	1	B	60	{0, 2, 3}	2	2	A	65	{0, 2, 3}	2	2	B	50	{0, 1}	0	0
3	42	B	75	{0, 1, 2, 3}	0	0	B	75	{0, 1, 2, 3}	.	.	A	75	{0, 1, 2, 3}	.	.	A	75	{0, 1}	1	1	B	75	{0, 1}	2	1	B	75	{0, 1}	.	.
3	43	B	70	{0, 1, 2, 3}	2	2	B	60	{0, 1, 2, 3}	2	2	A	75	{0, 1, 2, 3}	2	2	B	0	{0, 2, 3}	2	2	A	70	{0, 2, 3}	2	2	B	50	{0, 1}	3	2
3	44	B	75	{0, 1, 2, 3}	2	2	B	75	{0, 1, 2, 3}	1	1	A	75	{0, 1, 2, 3}	2	1	A	50	{0, 1}	1	2	B	50	{0, 1}	1	2	B	50	{0, 1}	1	1
3	45	B	65	{0, 1, 2, 3}	.	.	A	55	{0, 2, 3}	2	2	A	70	{0, 1, 2, 3}	2	2	B	70	{0, 2, 3}	2	2	A	70	{0, 2, 3}	2	2	A	70	{0, 2, 3}	3	3
3	46	B	75	{0, 1, 2, 3}	2	2	A	50	{0, 2, 3}	2	3	A	75	{0, 1, 2, 3}	.	.	B	50	{0, 2, 3}	2	3	A	50	{0, 2, 3}	2	.	A	50	{0, 2, 3}	.	.
3	47	B	50	{0, 1, 2, 3}	2	0	B	50	{0, 1, 2, 3}	.	.	B	100	{0}	.	.	A	50	{0, 1}	.	1	A	50	{0, 2, 3}	.	.	B	0	{0, 1}	.	.
3	48	B	75	{0, 1, 2, 3}	2	2	B	75	{0, 1, 2, 3}	.	.	A	75	{0, 1, 2, 3}	2	2	A	75	{0, 1}	0	0	B	75	{0, 1}	0	0	B	75	{0, 1}	1	0
4	49	B	50	{0, 1, 2, 3}	.	.	B	100	{0, 1, 2, 3}	.	.	A	60	{0, 1, 2, 3}	.	.	A	50	{0, 1}	.	.	A	50	{0, 2, 3}	.	.	B	75	{0, 1}	.	.
4	50	B	50	{0, 1, 2, 3}	.	.	B	50	{0, 1, 2, 3}	1	1	A	50	{0, 1, 2, 3}	2	1	A	75	{0, 1}	.	1	B	75	{0, 1}	1	1	B	75	{0, 1}	1	1
4	51	B	50	{0, 1, 2, 3}	0	0	B	50	{0, 1, 2, 3}	0	0	B	50	{0}	0	0	B	0	{0, 2, 3}	0	0	B	50	{0, 1}	0	0	B	50	{0, 1}	0	0
4	52	B	0	{0, 1, 2, 3}	2	2	A	65	{0, 2, 3}	.	2	A	80	{0, 1, 2, 3}	2	2	B	80	{0, 2, 3}	2	2	A	100	{0, 2, 3}	2	2	A	100	{0, 2, 3}	.	2
4	53	B	75	{0, 1, 2, 3}	2	2	A	60	{0, 2, 3}	0	0	A	75	{0, 1, 2, 3}	2	2	A	50	{0, 1}	1	1	A	50	{0, 2, 3}	2	2	B	50	{0, 1}	1	1
4	54	B	80	{0, 1, 2, 3}	1	1	B	60	{0, 1, 2, 3}	1	1	A	70	{0, 1, 2, 3}	2	2	A	66	{0, 1}	1	1	B	60	{0, 1}	1	1	B	70	{0, 1}	1	1
4	55	B	75	{0, 1, 2, 3}	2	2	B	75	{0, 1, 2, 3}	2	2	A	100	{0, 1, 2, 3}	2	2	A	75	{0, 1}	1	1	A	75	{0, 2, 3}	2	2	A	75	{0, 2, 3}	2	2
4	56	B	75	{0, 1, 2, 3}	.	.	A	50	{0, 2, 3}	.	.	A	75	{0, 1, 2, 3}	2	2	B	50	{0, 2, 3}	2	2	A	50	{0, 2, 3}	2	2	A	50	{0, 2, 3}	2	2
4	57	B	75	{0, 1, 2, 3}	.	1	B	60	{0, 1, 2, 3}	1	1	A	85	{0, 1, 2, 3}	2	2	B	60	{0, 2, 3}	2	2	A	60	{0, 2, 3}	2	2	A	50	{0, 2, 3}	2	2
4	58	B	50	{0, 1, 2, 3}	1	1	B	50	{0, 1, 2, 3}	.	.	A	50	{0, 1, 2, 3}	.	.	B	50	{0, 2, 3}	.	.	A	50	{0, 2, 3}	.	.	A	0	{0, 2, 3}	.	.
4	59	B	75	{0, 1, 2, 3}	2	2	A	50	{0, 2, 3}	2	2	A	85	{0, 1, 2, 3}	2	2	B	60	{0, 2, 3}	2	2	A	67	{0, 2, 3}	2	2	A	60	{0, 2, 3}	2	2
4	60	B	50	{0, 1, 2, 3}	2	1	B	50	{0, 1, 2, 3}	1	1	A	50	{0, 1, 2, 3}	.	.	A	0	{0, 1}	.	.	B	0	{0, 1}	.	.	B	0	{0, 1}	1	1
4	61	B	75	{0, 1, 2, 3}	.	.	A	50	{0, 2, 3}	.	.	A	75	{0, 1, 2, 3}	.	.	B	60	{0, 2, 3}	.	.	A	75	{0, 2, 3}	.	.	A	55	{0, 2, 3}	.	.
4	62	B	66	{0, 1, 2, 3}	2	2	A	67	{0, 2, 3}	2	2	A	78	{0, 1, 2, 3}	2	2	B	53	{0, 2, 3}	.	.	A	50	{0, 2, 3}	.	.	A	50	{0, 2, 3}	2	2
4	63	B	67	{0, 1, 2, 3}	2	2	B	50	{0, 1, 2, 3}	2	2	A	50	{0, 1, 2, 3}	2	2	A	75	{0, 1}	0	0	A	50	{0, 2, 3}	2	2	B	60	{0, 1}	1	1
4	64	A	50	{0}	.	0	B	100	{0, 1, 2, 3}	1	1	A	100	{0, 1, 2, 3}	1	1	A	100	{0, 1}	1	1	B	100	{0, 1}	0	1	B	100	{0, 1}	1	1

Table 22: Individual data in the team communication experiment à la Anderson and Holt (1997), sessions 1-4

		Decision 1 (BBAb)					Decision 2 (BAABAb)					Decision 3 (BAAa)					Decision 4 (BBABa)					Decision 5 (ABAAAb)					Decision 6 (BAAAb)				
Session	Subject ID	Sugg. dec.	Confidence	Fingerprint	RA 1	RA 2	Sugg. dec.	Confidence	Fingerprint	RA 1	RA 2	Sugg. dec.	Confidence	Fingerprint	RA 1	RA 2	Sugg. dec.	Confidence	Fingerprint	RA 1	RA 2	Sugg. dec.	Confidence	Fingerprint	RA 1	RA 2	Sugg. dec.	Confidence	Fingerprint	RA 1	RA 2
5	65	B	75	{0, 1}	.	.	B	67	{0, 1}	.	.	A	67	{0, 1}	.	.	A	67	{0, 1, 2, 3}	2	.	B	67	{0, 1, 2, 3}	.	.	B	67	{0, 1, 2, 3}	.	.
5	66	A	65	{0, 2, 3}	2	2	A	80	{0, 2, 3}	2	2	B	60	{0, 2, 3}	2	2	A	100	{0, 1, 2, 3}	.	.	B	100	{0, 1, 2, 3}	2	2	B	100	{0, 1, 2, 3}	2	2
5	67	B	70	{0, 1}	.	.	B	70	{0, 1}	.	.	A	70	{0, 1}	.	.	A	60	{0, 1, 2, 3}	.	.	B	50	{0, 1, 2, 3}	.	.	B	60	{0, 1, 2, 3}	.	.
5	68	B	75	{0, 1}	0	0	B	75	{0, 1}	0	0	A	75	{0, 1}	1	1	A	75	{0, 1, 2, 3}	2	1	B	75	{0, 1, 2, 3}	2	1	B	75	{0, 1, 2, 3}	.	.
5	69	B	75	{0, 1}	.	.	B	75	{0, 1}	.	.	A	75	{0, 1}	.	.	A	50	{0, 1, 2, 3}	.	.	B	63	{0, 1, 2, 3}	1	1	B	63	{0, 1, 2, 3}	1	1
5	70	B	60	{0, 1}	1	1	B	65	{0, 1}	1	1	A	75	{0, 1}	0	0	A	60	{0, 1, 2, 3}	2	1	B	55	{0, 1, 2, 3}	2	2	A	60	{0}	2	0
5	71	B	67	{0, 1}	1	1	B	66	{0, 1}	1	1	A	67	{0, 1}	1	1	A	50	{0, 1, 2, 3}	0	0	B	67	{0, 1, 2, 3}	.	.	B	80	{0, 1, 2, 3}	2	2
5	72	B	67	{0, 1}	1	1	B	67	{0, 1}	.	.	A	67	{0, 1}	.	.	A	67	{0, 1, 2, 3}	.	.	B	67	{0, 1, 2, 3}	.	.	B	67	{0, 1, 2, 3}	.	.
5	73	A	60	{0, 2, 3}	2	2	A	60	{0, 2, 3}	2	2	B	60	{0, 2, 3}	2	2	A	60	{0, 1, 2, 3}	2	2	B	50	{0, 1, 2, 3}	2	2	B	60	{0, 1, 2, 3}	2	2
5	74	A	50	{0, 2, 3}	.	.	A	50	{0, 2, 3}	3	3	A	50	{0, 1}	0	0	A	70	{0, 1, 2, 3}	3	3	A	0	{0, 2, 3}	2	2	B	70	{0, 1, 2, 3}	3	3
5	75	B	50	{0, 1}	.	.	B	60	{0, 1}	.	.	B	50	{0, 2, 3}	.	.	A	50	{0, 1, 2, 3}	.	.	B	50	{0, 1, 2, 3}	.	.	B	60	{0, 1, 2, 3}	.	.
5	76	B	70	{0, 1}	3	3	B	70	{0, 1}	3	3	B	50	{0, 2, 3}	3	3	A	66	{0, 1, 2, 3}	3	3	B	66	{0, 1, 2, 3}	3	3	B	66	{0, 1, 2, 3}	.	.
5	77	B	80	{0, 1}	0	0	B	85	{0, 1}	.	.	A	80	{0, 1}	.	.	A	65	{0, 1, 2, 3}	2	2	A	65	{0, 2, 3}	.	.	B	65	{0, 1, 2, 3}	2	2
5	78	B	50	{0, 1}	1	1	A	50	{0, 2, 3}	2	2	B	50	{0, 2, 3}	.	.	A	75	{0, 1, 2, 3}	2	2	B	50	{0, 1, 2, 3}	2	2	B	75	{0, 1, 2, 3}	2	2
6	79	B	50	{0, 1}	0	0	B	50	{0, 1}	0	0	A	50	{0, 1}	0	0	A	50	{0, 1, 2, 3}	0	0	B	50	{0, 1, 2, 3}	0	0	B	50	{0, 1, 2, 3}	0	0
6	80	A	90	{0, 2, 3}	2	2	A	98	{0, 2, 3}	2	2	B	85	{0, 2, 3}	2	2	A	70	{0, 1, 2, 3}	2	2	B	50	{0, 1, 2, 3}	1	1	B	95	{0, 1, 2, 3}	2	2
6	81	B	66	{0, 1}	1	1	B	66	{0, 1}	1	1	A	66	{0, 1}	1	1	A	50	{0, 1, 2, 3}	2	2	A	50	{0, 2, 3}	.	.	B	50	{0, 1, 2, 3}	2	2
6	82	A	75	{0, 2, 3}	2	2	A	60	{0, 2, 3}	2	2	B	55	{0, 2, 3}	2	2	A	75	{0, 1, 2, 3}	.	.	B	50	{0, 1, 2, 3}	2	2	B	75	{0, 1, 2, 3}	.	.
6	83	B	50	{0, 1}	.	.	B	50	{0, 1}	.	.	B	50	{0, 2, 3}	.	.	A	50	{0, 1, 2, 3}	.	.	B	50	{0, 1, 2, 3}	.	.	B	50	{0, 1, 2, 3}	.	.
6	84	B	67	{0, 1}	1	1	B	67	{0, 1}	1	1	A	67	{0, 1}	.	.	A	75	{0, 1, 2, 3}	2	2	B	67	{0, 1, 2, 3}	2	2	B	70	{0, 1, 2, 3}	2	1
6	85	B	66	{0, 1}	1	1	B	66	{0, 1}	1	1	A	66	{0, 1}	1	1	A	66	{0, 1, 2, 3}	1	1	B	66	{0, 1, 2, 3}	1	1	B	66	{0, 1, 2, 3}	1	1
6	86	A	0	{0, 2, 3}	2	2	A	60	{0, 2, 3}	2	2	A	50	{0, 1}	.	.	A	50	{0, 1, 2, 3}	.	.	A	50	{0, 2, 3}	.	.	B	50	{0, 1, 2, 3}	.	.
6	87	B	50	{0, 1}	.	.	A	50	{0, 2, 3}	.	.	B	50	{0, 2, 3}	.	.	A	70	{0, 1, 2, 3}	.	.	B	50	{0, 1, 2, 3}	.	.	B	70	{0, 1, 2, 3}	.	.
6	88	A	75	{0, 2, 3}	2	2	A	75	{0, 2, 3}	2	2	B	70	{0, 2, 3}	2	2	A	70	{0, 1, 2, 3}	2	2	B	50	{0, 1, 2, 3}	0	0	B	75	{0, 1, 2, 3}	.	1
6	89	B	67	{0, 1}	.	.	B	67	{0, 1}	.	.	A	50	{0, 1}	.	.	B	50	{0}	.	.	B	50	{0, 1, 2, 3}	.	.	A	60	{0}	.	.
6	90	A	50	{0, 2, 3}	2	2	A	50	{0, 2, 3}	0	0	B	50	{0, 2, 3}	2	2	A	70	{0, 1, 2, 3}	2	2	B	50	{0, 1, 2, 3}	0	0	B	60	{0, 1, 2, 3}	2	2
6	91	B	60	{0, 1}	.	.	B	60	{0, 1}	3	3	B	70	{0, 2, 3}	2	2	A	80	{0, 1, 2, 3}	3	3	B	70	{0, 1, 2, 3}	3	3	B	90	{0, 1, 2, 3}	3	3
6	92	B	50	{0, 1}	.	0	A	50	{0, 2, 3}	.	.	B	50	{0, 2, 3}	.	.	A	70	{0, 1, 2, 3}	.	.	A	65	{0, 2, 3}	2	2	B	75	{0, 1, 2, 3}	.	.
7	93	B	50	{0, 1}	3	3	A	50	{0, 2, 3}	3	3	B	70	{0, 2, 3}	2	2	A	50	{0, 1, 2, 3}	2	2	B	50	{0, 1, 2, 3}	2	2	B	80	{0, 1, 2, 3}	2	2
7	94	B	80	{0, 1}	1	1	A	75	{0, 2, 3}	2	2	B	50	{0, 2, 3}	.	.	A	90	{0, 1, 2, 3}	2	2	A	80	{0, 2, 3}	2	2	B	95	{0, 1, 2, 3}	2	2
7	95	B	75	{0, 1}	.	1	B	75	{0, 1}	.	.	A	75	{0, 1}	.	.	B	50	{0}	.	.	A	50	{0, 2, 3}	.	.	A	50	{0}	.	.
7	96	B	100	{0, 1}	1	1	B	75	{0, 1}	1	1	A	75	{0, 1}	0	0	A	55	{0, 1, 2, 3}	.	.	B	50	{0, 1, 2, 3}	.	.	B	57	{0, 1, 2, 3}	.	.
7	97	B	0	{0, 1}	1	1	B	50	{0, 1}	.	1	A	100	{0, 1}	1	1	A	50	{0, 1, 2, 3}	1	1	B	100	{0, 1, 2, 3}	2	1	A	80	{0}	.	.
7	98	B	60	{0, 1}	.	1	B	60	{0, 1}	1	1	A	60	{0, 1}	.	.	A	60	{0, 1, 2, 3}	2	2	A	50	{0, 2, 3}	2	2	B	50	{0, 1, 2, 3}	.	.
7	99	A	80	{0, 2, 3}	2	2	A	80	{0, 2, 3}	2	2	B	70	{0, 2, 3}	2	2	A	80	{0, 1, 2, 3}	2	2	B	50	{0, 1, 2, 3}	2	2	B	80	{0, 1, 2, 3}	2	2
7	100	A	80	{0, 2, 3}	2	2	A	80	{0, 2, 3}	2	2	B	80	{0, 2, 3}	2	2	A	85	{0, 1, 2, 3}	2	2	A	50	{0, 2, 3}	2	3	B	85	{0, 1, 2, 3}	2	2
7	101	A	75	{0, 2, 3}	2	2	A	90	{0, 2, 3}	2	2	B	80	{0, 2, 3}	2	2	A	95	{0, 1, 2, 3}	2	2	B	50	{0, 1, 2, 3}	2	2	B	80	{0, 1, 2, 3}	2	2
7	102	B	75	{0, 1}	.	.	B	75	{0, 1}	.	.	A	90	{0, 1}	.	.	A	0	{0, 1, 2, 3}	.	.	B	65	{0, 1, 2, 3}	.	.	B	100	{0, 1, 2, 3}	.	.
7	103	B	50	{0, 1}	.	.	A	80	{0, 2, 3}	.	.	A	80	{0, 1}	.	.	A	50	{0, 1, 2, 3}	.	.	A	75	{0, 2, 3}	.	.	B	75	{0, 1, 2, 3}	.	.
7	104	B	80	{0, 1}	.	.	A	80	{0, 2, 3}	.	.	A	75	{0, 1}	.	.	A	70	{0, 1, 2, 3}	.	.	B	85	{0, 1, 2, 3}	.	.	B	80	{0, 1, 2, 3}	.	.
7	105	A	80	{0, 2, 3}	2	2	A	80	{0, 2, 3}	.	.	B	80	{0, 2, 3}	.	.	A	80	{0, 1, 2, 3}	.	.	B	80	{0, 1, 2, 3}	.	.	B	80	{0, 1, 2, 3}	.	.
7	106	A	90	{0, 2, 3}	2	2	A	90	{0, 2, 3}	.	.	B	80	{0, 2, 3}	.	.	A	95	{0, 1, 2, 3}	2	2	B	50	{0, 1, 2, 3}	.	0	B	90	{0, 1, 2, 3}	2	1

Table 23: Individual data in the team communication experiment à la Anderson and Holt (1997), sessions 5-7

Subject ID	Decision 1 (AAB) Sugg. dec.	RA 1	RA 2	Decision 2 (BBAA) Sugg. dec.	RA 1	RA 2	Decision 3 (AAABA) Sugg. dec.	RA 1	RA 2	Decision 4 (BBB) Sugg. dec.	RA 1	RA 2	Decision 5 (AABA) Sugg. dec.	RA 1	RA 2	Decision 6 (BABA) Sugg. dec.	RA 1	RA 2	Fingerprint Analysis, decisions 1-6 (Shares s_i^k, level classification in bold) Level-1	Level-2	Level-3	Bayesian	Cluster
1	2	.	.	5	.	.	0	.	.	-1	.	.	0	1	1	-1	.	.	**1.00**	0.00	0.00	0.00	
2	8	.	.	2	.	.	2	.	.	4.32	.	.	7.31	.	.	-1.11	.	.	0.00	0.00	0.00	0.00	A
3	0	1	1	0	.	.	0	.	.	0	.	.	0	.	.	0	.	.	**1.00**	0.00	0.00	0.00	
4	4	0	2	0	2	1	-8	2	2	2	2	0	-10	2	2	0	2	2	0.00	**1.00**	0.00	0.00	B
5	5	.	.	0	.	.	-10	.	.	10	2	2	-10	.	.	0	.	.	0.00	**1.00**	0.00	0.00	C
6	0	.	.	0	.	.	-5	2	2	7	.	.	0	.	.	3	.	.	0.27	**0.32**	0.00	0.41	A
7	-2	.	.	-2	3	3	-4	.	.	7	.	.	3	3	3	0	.	.	0.27	0.00	**0.32**	0.41	A
8	-4	2	2	-3	.	.	-10	.	.	10	.	.	0	.	.	-10	.	.	0.00	**0.00**	1.00	0.00	
9	-2	2	3	0	2	2	-3	3	3	5	2	2	-2	2	2	0	2	2	0.53	**0.12**	**0.18**	0.17	A
10	5	0	0	0	.	.	0	.	.	0	0	0	1.11	.	.	-2	.	.	1.00	0.00	0.00	0.00	
11	0	0	1	0	1	1	0	1	1	-10	.	.	0	.	.	0	.	.	**0.00**	0.00	0.00	0.00	D
12	1	2	1	-2	3	3	-1	.	.	7	2	2	-1	3	3	1	0	0	0.33	**0.00**	**0.16**	0.51	A
13	-5	2	2	0	.	.	-8	.	.	4	2	2	-4	.	.	0	.	.	0.00	**0.84**	0.16	0.00	A
14	4	0	0	4	0	0	2	2	2	9	2	2	2	.	0	4	0	0	**0.00**	**0.00**	0.00	0.00	A
15	-8	.	0	-8	.	.	-8	.	.	0	2	2	-9	2	0	-1	.	.	0.00	**1.00**	0.00	0.00	E
16	-2	2	0	0	2	2	-10	2	2	8	2	0	-1	2	2	0	2	2	0.00	**0.67**	0.16	0.17	A
17	-1	.	.	0	2	2	-8	2	2	0	2	1	-8	2	2	0	2	2	0.00	**1.00**	0.00	0.00	B
18	0	.	.	-10	.	.	-5	.	.	0	.	.	-6	.	.	-6	.	.	0.00	0.00	0.59	0.41	E
19	3	.	.	5	.	.	6	.	.	-2	.	.	3	.	.	2	.	.	0.00	0.00	0.00	0.00	D
20	-10	.	.	0	2	2	-10	2	2	8	2	2	-10	2	2	-1	2	0	0.00	**1.00**	0.00	0.00	F
21	4	2	2	3	.	.	2	2	2	9	2	2	3	2	2	7	.	.	0.00	**0.00**	0.00	0.00	A
22	-5	2	2	0	2	2	-8	2	2	9	2	2	-2	2	2	-2	3	3	0.00	**0.70**	**0.21**	0.09	A
23	1	2	2	0	2	2	-4	2	2	8	2	2	-1	2	2	0	2	2	0.10	**0.14**	0.19	0.57	A
24	-1	.	.	1	.	.	-1	.	.	-3	.	.	1	.	.	-1	.	.	1.00	0.00	0.00	0.00	
25	7	0	2	7	0	2	7	.	.	10	.	.	0	.	.	-9	.	.	0.00	0.00	0.00	0.00	
26	-2	3	3	-3	3	3	-5	3	3	10	.	.	-5	3	3	-4	2	3	0.00	0.02	**0.80**	0.19	
27	0	1	1	0	1	1	-9	.	.	8	2	2	-0.5	2	1	0	.	.	**0.00**	**0.44**	0.21	0.35	A
28	4	.	.	6	.	.	-10	2	2	8	2	2	-10	2	2	5	.	0	0.00	**1.00**	0.00	0.00	C
29	3	2	0	2	2	2	-1	2	2	3.5	3	3	-1	2	2	2.5	0	0	1.00	**0.00**	**0.00**	0.00	A
30	-8	2	2	0	2	2	-10	2	2	10	2	2	-8	2	2	0	2	2	0.00	**1.00**	0.00	0.00	F
31	0	.	.	0	2	1	-10	2	2	10	2	2	-10	2	2	0	1	1	**0.00**	**1.00**	0.00	0.00	C
32	0	.	.	-10	.	.	-10	3	2	5	.	.	-10	.	.	0	.	.	0.00	0.59	0.00	0.41	E
33	3.1	.	.	0	.	.	1	.	.	4.25	.	.	3.1	.	.	0	.	.	1.00	0.00	0.00	0.00	A
34	0	.	.	8	.	.	7	.	.	10	.	.	-8	.	.	0	.	.	0.00	0.00	0.00	0.00	A
35	8	0	2	1	2	0	3	2	2	-10	2	2	3	0	.	0	1	1	**0.00**	**0.00**	0.00	0.00	D
36	0	0	1	0	.	1	0	.	.	0	1	1	0	.	.	0	.	.	**1.00**	0.00	0.00	0.00	
37	-2	3	3	0	.	2	-4	2	2	2	2	.	-1.5	2	0	1	.	.	0.91	**0.03**	**0.03**	0.03	A
38	2	2	0	4	.	.	0	2	2	8	2	2	-4	2	2	0	2	2	0.00	**0.00**	0.00	0.00	A
39	0	2	2	-1	3	3	-2	.	0	2	2	3	-1.5	3	3	-0.5	.	.	0.99	**0.00**	**0.00**	0.00	A
40	4	.	.	6	.	.	3	2	0	-7	2	2	5	.	.	-3	.	.	0.00	**0.00**	0.00	0.00	D
41	-8	2	2	-8	.	.	-10	.	.	6	.	.	-10	.	.	-4	.	.	0.00	**0.76**	0.24	0.00	E
42	-1	2	2	0.5	2	0	-7	2	2	9	2	2	-3	3	3	0	.	.	0.00	**0.61**	**0.15**	0.24	A

Table 24: Individual data in the team communication experiment à la ÇK

B.2. Learning in Çelen and Kariv (2004)

In order to investigate learning over the course of ÇK's experiment, table 25 shows the fingerprint analysis by the first 8 and the last 7 rounds. None of the two classifications yields significantly different results (Chi-squared test, $p = 0.114$, $p_{mix} = 0.638$).

The data from fewer rounds is less distinctive, the classification thus features more subjects that are not significantly close to any type. This fraction is stable around a quarter of the subjects. Regarding the identified types, both the classification by the lowest significant SSD^k (left) and by the p-values below 0.05 (right) indicate that the fraction of level-1 players reduces over time while the fraction of level-2 and level-3 players increases. For Bayesians, the two classifications yield different tendencies, which might be due to an insufficient differentiation between level-3 and Bayesian types. Overall, the numbers indicate some learning over time and a tendency towards higher level inference. There is, however, no strong convergence to level-3 or rational Bayesian play.

Type	Frequency	Fraction (%)
1	10	25.0
2	13	32.5
3	0	0.0
B	7	17.5
NA	10	25.0
Total	40	100.00

(a) Classification by lowest significant SSD, first 8 rounds

Type	Frequency	Fraction (%)
1	8.50	21.3
2	12.22	30.6
3	2.87	7.2
B	5.41	13.5
NA	11.00	27.5
Total	40	100.00

(b) Classification by p-value below 0.05, first 8 rounds

Type	Frequency	Fraction (%)
1	6	15.0
2	18	45.0
3	4	10.0
B	3	7.5
NA	9	22.5
Total	40	100.00

(c) Classification by lowest significant SSD, last 7 rounds

Type	Frequency	Fraction (%)
1	4.04	10.1
2	14.58	36.5
3	5.12	12.8
B	6.26	15.6
NA	10.00	25.0
Total	40	100.00

(d) Classification by p-value below 0.05, last 7 rounds

Table 25: Type overview for 40 subjects in ÇK, early and late rounds

C. Experiment instructions (AH framework)

Welcome to the experiment!

Introduction

You are about to participate in an experiment in team decision making. The experiment is funded by Cornell University. Please follow the instructions carefully.

In addition to the participation fee of $5, you may earn a considerable additional amount of money. Your decisions determine the additional amount. You will be instructed in detail how your earnings depend on your decisions. All that you earn is yours to keep, and will be paid to you in private, in cash, after today's session.

It is important to us that you remain silent and do not look at other people's screens. If you have any questions or need assistance of any kind, please raise your hand, and an experimenter will come to you. If you talk, shout out loud, etc., you will be asked to leave. Thank you.

The experiment consists of a test round and 2 parts. Part I is designed as a warm-up with 3 trivia questions. Part II consists of 2 rounds of 3 decisions each. In all 6 decisions your task will be identical and success is identically rewarded.

Since this is a team experiment, you will be randomly matched with another participant in this room, to form a team that plays as one entity. Your teammate will change every round, so please do not assume content of previous communication to be known by your new partner. The way you interact as a team to take decisions will be the same throughout all rounds.

Now, let me explain how your ***Team's Action*** is determined. In fact, both your teammate and you will enter a ***Final Decision*** individually and the computer will choose randomly which one of your two final decisions counts as your team's action. The probability that your teammate's final decision is chosen is equal to the probability that your final decision will be chosen (i. e. your chances are 50:50). However, you have the possibility to influence your partner's final decision in the following way: Before you enter your final decision, you can propose to your partner a ***Suggested Decision*** and send him one and only one text ***Message***. *Note that this message is your only chance to convince your partner of the reasoning behind your suggested decision. Therefore, use the message to explain your suggested decision to your teammate.* After you finish entering your suggested decision and your message, these will be shown to your teammate. Simultaneously, you will receive your partner's suggested decision and message. Both of you will then make your final decision. As outlined above, once you both enter your final decision, the computer chooses randomly one of your final decisions as your team's action.

If you have any questions at this point, please raise your hand. In order for you to get familiar with the messaging system, you will now try it out in a **Test Period**. Please turn the page for further instructions.

Test period

A participant in this room is now randomly chosen to be your teammate. The **Test Period** has two rounds, with communication in each round. Since this is only a test, your earnings will not depend on anything that happens now. In both test rounds you will need to send and receive pieces of information. The information consists of the answer to a question and one given phrase. After the successful exchange, you will enter the number again. This way, the communication structure is identical to the one in the experiment rounds.

The messenger allows ***Messages*** of any size. However, you have to enter the message line by line since the input space is only one line. Within this line you can delete by using the usual "Backspace" button of your keyboard. By pressing "Enter" on the keyboard, you add the written sentence to the message. Please note that <u>only</u> added sentences will be sent and seen by your partner. *The words in the blue input line will **not** be sent*. You can always delete previously added sentences by clicking the “Clear Input” button. The number of lines you send is not limited. You can therefore send messages of any length. You finally send the message to your partner by clicking the “Send Message” button.

If you have any questions at this point, please raise your hand. When you are ready, please click the “Ready” button to start the **Test Period**.

Experiment - Part II

You are about to start ***Part II*** of the experiment.[21] This part consists of 2 periods of 3 decisions each.[22] In each period you will team up with a different, randomly chosen participant. Therefore you will make 3 urn choices with a given partner.

The 6 scenarios for your urn choices are taken from another experiment. They are randomly chosen and simply allow me to confront you with more scenarios than if we played it out ourselves. I will now explain the setup to you in the same fashion the participants of the original experiment were instructed. The original experiment was not done by teams, but I will explain it with teams for simplicity.

In each urn choice your task is the following:

Your team is asked to **predict** from which **randomly chosen urn** a ball was drawn. It is equally likely that urn A or urn B will be the true urn. That is, there is a 50 percent chance that Urn A is the true urn, and a 50 percent chance that Urn B is the true urn. **Urn A** contains **2 white** balls, and **1 black** ball. **Urn B** contains **1 white** ball and **2 black** balls. Therefore, there is a $\frac{2}{3}$ chance that a white ball comes from urn A and a $\frac{2}{3}$ chance that a black ball comes from urn B.

To help determine which urn is the right one in a given scenario, you will see one ball, drawn at random, from the urn. **Only** your team will see the outcome of this private draw. And your team will only see this **one draw**, you will get the same information as your teammate. After each draw, the ball is returned to the container before making the next draw. Therefore, each team will have **one private draw**, with the **ball being replaced** after each draw. This way each team draws from an urn that contains 3 balls in total.

When it is your turn to decide, bold text in the top of your screen will give you information about your draw. If the ball the computer has randomly drawn for you is white, your window will read, “**Your team’s private draw from the urn is WHITE.**” Your window will read, “**Your team’s private draw from the urn is BLACK.**”, if the ball the computer has randomly drawn for you is black.

In each scenario, **decisions are made sequentially**, i.e. one team after the other. The order in which teams decide in a given scenario is determined randomly. Once the first team has agreed on its action based on its private draw, the second team will be asked to make its decision. The members of team 2 will see their private draw, and **also** which urn was chosen by the first team. Until the end of the round, all following teams will be **informed about the chosen urns of all earlier teams**, but not about the other teams’ draw from the urn.

In each round, you see the three scenarios and will write down your ***Suggested Decision*** and ***Message*** for each of them. After that you see your partner’s ***Suggested Decision*** and ***Message*** for each of the three scenarios and will make your ***Final Decisions***. Finally, you will be informed about your ***Team’s Action*** and the true urn. The order of events is illustrated in the table below. For each of the 6 scenarios, if your team’s action and the chosen urn coincide, you will individually earn $1.00 (your teammate will get $1.00 as well).

As described earlier, you will send your teammate a ***Suggested Decision*** and a ***Message***. Remember to explain in the message your reasoning behind your suggested decision. *(And*

[21] The experiment consists of two parts, part I being independent of part II.

[22] Table 26 was shown to give an overview.

note again that the words in the blue input line will ***not*** *be sent. Press "Enter" to add them to the message.)* After this information is exchanged, both of you enter your ***Final Decision***, from which the computer randomly chooses the ***Team's Action***.

As part of the communication, you can quantify your ***Confidence*** for each choice you make. You can put numbers between 50% and 100%, where 50% implies that you think both urns are equally likely and higher numbers reflect your higher confidence in your choice, up to certainty (100%).

Let me summarize the main points: (1) In each scenario, it is equally likely that urn A or urn B is the true urn. (2) There is a $\frac{2}{3}$ chance that a white ball came from urn A and a $\frac{2}{3}$ chance that a black ball came from urn B. (3) In the knowledge of the previous teams' actions and your draw, choose either urn A or urn B. (4) Like you, other teams saw one draw from the true urn and their predecessors' urn choice. (5) If your team chooses the true urn, you will earn $\$1.00$.

If you have any questions at this point, please raise your hand. When you click the "Ready" button, you will start the first round of the experiment.

	Round	Urn choice	Action
↓		1	
↓		2	Suggested Decision
↓		3	
↓		1	
↓	1	2	Final Decision
↓		3	
↓		1	
↓		2	Get result
↓		3	
↓		4	
↓		5	Suggested Decision
↓		6	
↓		4	
↓	2	5	Final Decision
↓		6	
↓		4	
↓		5	Get result
↓		6	

Table 26: Order of events in Part II.

D. Classification instructions (AH framework)

In the following I will describe the classification process for the analysis of an experiment. Subjects play a game of incomplete information. Their reasoning will be classified along the lines of a model of level of reasoning, which will be laid out in the first section. It is set up in analogy to the level-k model in complete information settings as introduced by Nagel (1995) and Camerer et al. (2004).

Follow the instructions of this booklet. Read them entirely to get an overview and then start the classification. The game is the social learning experiment as implemented by Anderson and Holt (1997), which I assume you are fully familiar with. The subjects were put in 6 different situations that occurred in the original study by Anderson and Holt (1997) and decided in a team about their action, which would be remunerated according to the true urn in Anderson and Holt (1997).

The model will introduce four different ways of reasoning in the context of this game. The aim of the classification is to take a look at the 6 messages per player and connect him/her with one or more types. Details will be explained below. Please limit yourself to making inferences only from what can clearly be derived from the message stated, i.e. do not try to think about what the player *might have thought*.

Reasoning types in the model

In the context of social learning, reasoning types differ in the ways they process private information (in form of draws from an urn) and public information (in form of actions of predecessors). A natural model of level of reasoning starts out with a random level-0 type and builds a hierarchy based on the number of iterated best responses. In this game, this hierarchy turns out as follows:

Level-0 Randomizing between A and B, irrespective of own private signal.

Level-1 If others' signal is not informative, like level-0 random play, then the best response is to get informed by the own signal. Therefore, level-1 players are following their own private signal.

Level-2 Since level-1 play is fully informative about the private signals, the best response in light of this is to follow the majority if it is ahead by more than one signal difference. A level-2 player will thus only follow his signal if the previous actions were split equally between the two urns or one urn was chosen just one time more often.

Level-3 Since level-2 play is as informative as Bayesian play, level-3 play, a best-response, is like original Bayesian play. This implies that the play and beliefs of level-3 players are identical to play and beliefs in the Bayesian equilibrium if everybody is level-3.

Starting from this model, I will now explain some expected message contents for the individual levels of reasoning.

Level-0 **Random play** should result from messages that clearly do not understand the nature of the game or put reasons for play that are orthogonal to any reasonable inference using private signal and previous actions.

Level-1 Level-1 is characterized by **disregarding others' signal**. In the message, this can be an open disregard of others' actions or an emphasis on the unambiguity of the own signal.

Level-2 A level-2 players with a degenerate population belief (as introduced in the model above) implies that all **others' actions are taken at face value** like a signal. It follows that the ratio of As vs. Bs in the previous actions might weighted against the 2/3 chance of the own signal to be correct. In any case, a level-2 player **never engages in a differentiation of individual predecessors**. If everybody is assumed to play their signal, this implies that others are regarded as a homogeneous crowd that simply differs in the signals they received not in their position in the sequence. Alternatively to a degenerate population belief, a level-2 player might think that some players played random, like level-0. Then, the signals are not taken at full face value and probably the own signal is valued more than the observed actions.

Level-3 heterogeneity Under a degenerate population belief, only level-3 players **distinguish individual predecessors** in the extent to which their actions reflect their private signals. It is therefore a characteristic of level-3 players to differentiate informative and uninformative observed actions depending on the position of predecessors in the sequence. One example is that a level-3 player will note in a history of AAAA that the later players might just have followed the majority, he therefore does not infer those players signal with certaint. Put starkly, observing BBBA implies that a level-3 player rules out the fourth player to be level-2, an observation only level-3 will make.

The data

You will see different situations in which the players decided to choose urn A or urn B.

Each situation is explained in brackets by the information available to the subject. The capital letters indicate previous players' actions. The small letters indicate the private signal received by the player at hand. In addition to the sent message, you see the suggested decision, which is 1 for urn A and 2 for urn B. The communication was structured in the sense that it gave the players the possibility to indicate the confidence in the own urn choice. This number is given to you as well.

The classification

I would like you to classify messages into **one of the 4 levels** described above. Indicate the closest type under 'level' in the according space.

Indicate whether the population belief is **degenerate** on the next lower level or **non-degenerate**. If it is non-degenerate, denote which types (in terms of levels) are assumed to be present.

In addition to this classification, I ask you to indicate **any difference** between the observed reasoning and the type given. This might include differences in the belief about when others start imitating others' actions, the weight the public information receives compared to the private information, the way the population updating is done by a level-3 player, etc. Also, if you think the potential level could be more than one, indicate your considerations here.

E. Experiment instructions (ÇK framework)

The experiment instructions in German combine the communication instructions (see Introduction and Test period above) with instructions from Çelen and Kariv (2004) translated to German and adapted to the team setting. A translation is available upon request.

F. Classification instructions (ÇK framework)

In the following I will describe the classification process for the analysis of an experiment. Subjects play a game of incomplete information. Their reasoning will be classified along the lines of a model of level of reasoning, which will be laid out in the first section. It is set up in analogy to the level-k model in complete information settings as introduced by Nagel (1995) and Camerer et al. (2004).

Follow the instructions of this booklet. Read them entirely to get an overview and then start the classification. The game is the social learning experiment as implemented by Çelen and Kariv (2004), which I assume you are fully familiar with. The subjects in my experiment were put in 6 different situations that occurred in the original study by Çelen and Kariv (2004) and decided in a team about their action, which would be remunerated according to the true signal in Çelen and Kariv (2004).

The model will introduce five different ways of reasoning in the context of this game. The aim of the classification is to take a look at the 6 messages per player and connect him/her with one or more types. Details will be explained below. Please limit yourself to making inferences only from what can clearly be derived from the message stated, i.e. do not try to think about what the player *might have thought*.

Reasoning types in the model

In the context of social learning, reasoning types differ in the ways they process private information (in form of draws from an urn) and public information (in form of actions of predecessors). Çelen and Kariv (2004) analyse Bayesian behavior, which is fully rational and believes others to be fully rational. Please consult the paper to get a thorough understanding of this kind of reasoning in this context.

A natural model of level of reasoning starts out with a random level-0 type and builds a hierarchy based on the number of iterated best responses. In this game, this hierarchy turns out as follows:

Level-0 Playing a random threshold on the action space.

Level-1 If others' signals are not informative, like level-0 random play, then the best response is to get informed by the own signal. Therefore, level-1 players are following their own private signal and set a threshold of 0.

Level-2 With a threshold of 0, the observed action of a predecessor indicates whether the signal was in $[-10, 0)$ or $[0, 10]$. With this information, the best response is to add 5 or subtract 5 for each previous action A or B in the expected value of the sum of the signals. A level-2 player will thus expect a sum of signals of 0 if the history has equally many A's as B's and, e.g., 5 if one more B's than A's have been observed and 10 if two more B's than A's have been observed. The threshold will be set as the negative of the expected sum.

Level-3 Thresholds of -10 or 10 lead to uninformative play since the action is decided irrespective of the signal. Level-3 players hence only use information from level-2 players that are not cascading, i.e. play -5, 0, or 5.

Level-4 Level-4 players only use information from level-3 players that are not cascading. Although level-4 play is distinct from level-3, we will not pursue this or higher levels in our classification.

Starting from this model, I will now explain some expected message contents for the individual levels of reasoning.

Level-0 **Random play** should result from messages that clearly do not understand the nature of the game or put reasons for play that are orthogonal to any reasonable inference using private signal and previous actions.

Level-1 Level-1 is characterized by **disregarding others' action**. In the message, this can be an open disregard of others' actions or an emphasis on the *un*ambiguity of the own signal.

Level-2 A level-2 player with a degenerate population belief (as introduced in the model above) implies that all **others' actions are equally informative** and indicative of the sign of the signal and thus each one changes the expected sum of signals by $+5$ or -5. It follows that the relative number of As vs. Bs in the previous actions determine the threshold. In any case, a level-2 player **never engages in a differentiation of individual predecessors**. If everybody is assumed to play their signal, this implies that others are regarded as a homogeneous crowd that simply differs in the signals they received not in their position in the sequence.

Level-3 heterogeneity Level-3 players are the first in the hierarchy to **distinguish individual predecessors** in the extent to which their actions are informative about their private signals. In particular, level-2 actions might be completely uninformative if they result from a threshold of 10 or -10. At the same time, actions that result from a signal below a threshold of -5 or above a threshold of 5 induce a strong change in the expected sum of signals. Therefore, sum actions might have a stronger influence on the level-3's threshold than others. It is therefore a characteristic of level-3 players to differentiate more or less informative observed actions depending on the position of predecessors in the sequence and the likely threshold they had set. One example is that a level-3 player will note in a history of AB that the second player's signal must have been very low since a threshold of -5 was still undercut. This informativeness changes to uninformativeness when the threshold is expected to be 10 or -10. Then, the action is uninformative.

Bayesian More than level-3 players, Bayesian players realize that *every* predecessor has most likely set a different threshold. This way, particular later actions in the sequence can have a strong influence on the expected sum of signals and thus the set threshold.

The data

You will see different situations in which the players observed histories of A's and B's and set a Suggested Decision between -10 and 10.

Each situation is described by the information available to the subject. In addition to the sent message, you see the suggested decision.

The classification

I would like you to classify messages into **one of the 5 levels** described above. Indicate the closest type under 'level' in the according space.

Indicate whether the population belief is **degenerate** on the next lower level or **non-degenerate**. If it is non-degenerate, denote which types (in terms of levels) are assumed to be present.

In addition to this classification, I ask you to indicate **any difference** between the observed reasoning and the type given. This might include differences in the belief about what threshold is assumed to be behind others actions, the way the population updating is done by a level-3 player, etc. Also, if you think the potential level could be more than one, indicate your considerations here.

Please also note that some messages contain considerations regarding the threshold to set etc. which are orthogonal to the inferential considerations. Please make sure to only classify according to the strategic content.