Coordination games with asymmetric payoffs:
An experimental study with intra-group communication[*]

Jonas van Elten[†]
Stefan P. Penczynski[‡]

October 2018

Abstract

Two complementary modes of reasoning in coordination games are prominently discussed in the literature: level-k thinking and team reasoning. In order to differentiate between the two modes of reasoning, we experimentally investigate payoff asymmetric coordination games using an intra-group communication design that incentivizes subjects to explain the reasoning behind their decisions. We find that the reasoning process is significantly different between games. In payoff symmetric games, team reasoning plays an important role for coordination. In payoff asymmetric games, level-k reasoning results in frequent miscoordination. Our study clearly illustrates how small differences between strategic situations have a strong influence on reasoning.

Keywords: Coordination games, asymmetric payoffs, levels of reasoning, team reasoning.

JEL Classification: C72, C92

[†]Department of Economics, University of Mainz, Jakob Welder-Weg 9, 55128 Mainz, Germany, jonas.vanelten@uni-mainz.de, Tel. +49 6131 39 29904, Fax. +49 6131 39 29908.

[‡]School of Economics, University of East Anglia, Norwich NR4 7TJ, United Kingdom, S.Penczynski@uea.ac.uk, Tel. +44 1603 59 1796.

1 Introduction

Many social, political and economic situations require people to coordinate their behavior with others in order to realize mutual gains. The coordination on one of multiple options, however, is challenging as it constitutes a situation with multiple equilibria. In his seminal work, Schelling (1960) suggests that in pure coordination games subjects select strategies based on focal points. Experimental results confirm that subjects are indeed able to increase coordination rates by identifying focal points (Mehta, Starmer and Sugden, 1994; Bacharach and Bernasconi, 1997). Various influences on this ability, such as communication, learning, payoff- and risk-dominance have been studied in the literature.[1] We are particularly interested in Crawford, Gneezy and Rottenstreich's (2008, henceforth CGR) investigation of the robustness of focal points in two-player coordination games with asymmetric payoffs. These more realistic settings are analogous to the battle of the sexes game. In these games, subjects yield significantly lower coordination rates, raising doubts about the strength of focal points and the correct way of modeling subjects' behavior.

The success of coordination naturally depends on the way people reason about these coordination situations. In this study, we uncover subject's reasoning and discriminate between the two following, most prominent theories of coordination behavior. The collective rationality approach of "team reasoning" (Sugden, 1993; Bacharach, 2006) is based on Schelling's (1960) idea of focal points. Team reasoning assumes that subjects look for a rule or strategy of selection that tends to produce successful coordination if both players followed this rule. CGR introduce an alternative model of reasoning: a level-k model that incorporates salience in both labels and payoffs. As in standard level-k models, the types differ by the number of iterated best responses to a non-strategic level-0 belief (Nagel, 1995; Stahl and Wilson, 1995; Costa-Gomes et al., 2001; Camerer et al., 2004). The level-0 player, however, is modelled to both play label salient actions with higher probability in symmetric games and payoff salient actions with higher probability in asymmetric games. This model can then account for the drop of coordination rates in asymmetric games.

Both models make predictions based on concrete and plausible ways of reasoning in situations of coordination. Any attempt to differentiate between the two models is inconclusive. CGR investigate experimentally payoff symmetric and payoff asymmetric coordination games that feature either choices X and Y (X-Y games) or three pie slices of which one is distinct

---

[1]Many factors that influence the ability to coordinate have been studied, such as payoff- or risk dominance (Cooper et al., 1990; Van Huyck et al., 1990, 1991), salient labels (Mehta et al., 1994; Rubinstein, 1999; Bardsley et al., 2010), communication (Farrell, 1987; Cooper et al., 1989, 1992; Van Huyck et al., 1992; Ellingsen and Östling, 2010), gender (Holm, 2000), forward induction (Van Huyck et al., 1993; Cachon and Camerer, 1996; Crawford and Broseta, 1998), and learning (Crawford, 1995; Huyck et al., 1997; Camerer and Ho, 1998).

in terms of position and color (Pie games).[2] Overall, their results are not fully in accordance with any one of the two theories and conclude "that a judicious combination of these explanations, possibly incorporating other considerations, should help to predict the effectiveness of focal points" (CGR, p. 1456). Bardsley, Mehta, Starmer and Sudgen (2010) conduct experiments with symmetric one-shot text and number based coordination games. The results from one laboratory location mainly support team reasoning, those from another level-k reasoning. The authors conclude that "so far the search for a unified theory has been unsuccessful." (p. 78)

In order to make progress on this empirical question and to differentiate between the two candidate modes of reasoning, we propose to investigate coordination behavior with the help of eight X-Y and Pie games. We extend these CGR's games by an experimental design introduced by Burchardi and Penczynski (2014) that features intra-group communication.[3] The communication protocol reflects a simple version of a group discussion and it incentivizes subjects to demonstrate their individual reasoning process comprehensively within the messages. Thus, it allows us to observe the reasoning processes which govern decisions in coordination games. Our results and estimations are based on characteristics of the reasoning that two research assistants classified based on the messages.

Our data shows coordination rates that are similar to those in CGR. In the X-Y games, subjects realize high coordination rates in the payoff symmetric games, and frequent miscoordination in the asymmetric games. The results of our Pie games show a more diverse pattern, but again subjects achieve slightly higher coordination rates in the symmetric than in the asymmetric games.

The subjects' elicited levels of reasoning follow a clear pattern. The level-k distribution in payoff asymmetric games is significantly different from the one in payoff symmetric games. A set of panel regressions and MLE estimates of the level-k distribution indicate lower levels of reasoning in payoff symmetric games than in payoff asymmetric games. Analyzing subjects' level-0 beliefs, we find strong tendencies in payoff symmetric games to start reasoning based on salient action labels. In these cases, we identify team reasoning approaches frequently. In contrast, the large majority of messages in payoff asymmetric games exhibits payoff salience, not label salience. In particular, most subjects assume that level-0 players choose their own high payoff actions. These results confirm the level-k mechanism that CGR proposed to be behind the reduced effectiveness of focal points in payoff asymmetric coordination games.

---

[2]CGR's experimental Pie games are based on Blume and Gneezy (2000, 2010), who use a circular plastic plate divided by lines into three, five, or nine equally-sized sectors in order to study optimal learning and cognitive forward induction in a coordination context.

[3]Games are played by two groups of two subjects each. Team reasoning then concerns the "team" of two opposing groups, analogous to teams of two players in games of individual subjects.

Our results are in line with insights of related papers. In an investigation of neutrally framed Pie games, Faillo et al. (2013) find that neither of the two theories is consistently successful in describing behavior, but the relative success of team reasoning can be linked to features of the predicted equilibria such as payoff symmetry or non-Pareto-dominatedness. Isoni et al. (2013) study focal points in tacit bargaining with the help of spatial clues. In games similar to the X-Y games, the spatial framing influences the probability of coordination. It might be interesting to combine the communication protocol with such other games for further insights.

The remainder of the paper is structured as follows. Section 2 describes the experimental setup and the classification procedures. Section 3 summarizes the results of the experiment and of the classification. The estimation and its results are presented in section 4. The method and results are discussed in section 5. Section 6 concludes.

2 Empirical methods and experimental procedures

2.1 Coordination game

CGR test the robustness of focal points in two different kinds of coordination games with asymmetric payoffs. Each coordination task requires two randomly chosen players $g \in \{1, 2\}$ to individually and simultaneously choose among abstract labeled actions $a \in \mathcal{A}$. The players get a payoff of 0, except when they select the same action and coordinate their behavior successfully; then they receive the payoff $\pi_g(a, a)$. The X-Y games present subjects with a binary choice between two actions labeled X and Y, with X being considered the label salient option. We reproduce CGR's X-Y games with symmetric payoffs (SL) and with varying degrees of payoff asymmetry (ASL, AML, ALL). The Pie games differ from the X-Y games in terms of the framing and the design of the coordination task. The subjects choose one of three pie slices, from which the bottom slice (B) is distinct from the upper left (L) and the upper right (R) slices in terms of position and color. We reproduce two Pie games with symmetric payoffs (S1, S2) and two Pie games with payoff asymmetry (AM2, AM4). Table 1 summarizes the payoff structure of all eight coordination games. The visual representation of the actions is slightly different from the representation in CGR's experimental design. For operational reasons, we mark each pie slice with an additional abstract label \$, #, and §.[4] Appendix B.3 shows screenshots of a X-Y and a Pie game.

While CGR conduct the experiment using separate subject groups for each coordination game, in our experimental design all subjects face four X-Y and four Pie games. This within-subject design allows us to investigate individual reasoning across payoff symmetric and asymmetric games. In order to control for potential carryover or learning effects we change the

---

[4]We address this additional labeling in the results section 3.1.

Table 1: Payoff structure of coordination games.

X-Y games (CGR notation)	a	π_1, π_2	Pie games (CGR notation)	a	π_1, π_2
Symmetric Payoffs (SL)	X	5, 5	Symmetric Payoffs (S1)	L (\$)	5, 5
	Y	5, 5		R (#)	5, 5
				B (§)	5, 5
Slight Asymmetry (ASL)	X	5, 5.1	Symmetric Payoffs (S2)	L (\$)	6, 6
	Y	5.1, 5		R (#)	6, 6
				B (§)	5, 5
Moderate Asymmetry (AML)	X	5, 6	Moderate Asymmetry (AM2)	L (\$)	5, 6
	Y	6, 5		R (#)	6, 5
				B (§)	6, 5
Large Asymmetry (ALL)	X	5, 10	Moderate Asymmetry (AM4)	L (\$)	6, 7
	Y	10, 5		R (#)	7, 6
				B (§)	7, 5

sequence of games between the experimental sessions as shown in table 2. Our main results hold within each sequence.

Table 2: Sequence across sessions (symmetric games in bold).

	Sequence index $t \in \{1, \ldots, 8\}$ Game type														
	1		2		3		4		5		6		7		8
Session	X-Y		Pie		X-Y		Pie		X-Y		Pie		X-Y		Pie
1-3	ASL	$\rightarrow$	AM2	$\rightarrow$	AML	$\rightarrow$	AM4	$\rightarrow$	ALL	$\rightarrow$	**S2**	$\rightarrow$	**SL**	$\rightarrow$	**S1**
4-6	ALL	$\rightarrow$	AM4	$\rightarrow$	AML	$\rightarrow$	**S2**	$\rightarrow$	ASL	$\rightarrow$	AM2	$\rightarrow$	**SL**	$\rightarrow$	**S1**
7-9	**SL**	$\rightarrow$	**S1**	$\rightarrow$	ASL	$\rightarrow$	**S2**	$\rightarrow$	AML	$\rightarrow$	AM2	$\rightarrow$	ALL	$\rightarrow$	AM4

2.2 Incentivized communication protocol

The experimental design incorporates the concept of group decision making and adapts the intra-group communication protocol from Burchardi and Penczynski (2014). Two randomly selected anonymous subjects form a group that acts as one entity throughout the eight games and that produces a joint "group action" within a three step process.[5]

In the first step, the subjects face all eight coordination games individually. In each round, players write a message to their group partner. The message consists of a "suggested decision" (SD) that represents a subject's proposal for the joint "group action", and a justifying text message that explains the suggested decision to the group partner. The text message is unlimited in size and its writing is not limited in time. Subjects do *not* receive the message of their group partner after each game, they directly face the next coordination game. In the second step, once

---

[5]The experimental instructions are reproduced in appendices B.1 and B.2.

the players passed through all eight games, they receive the messages of their group partner separately for each game. Both group members can revise their suggested decision and state their "final decision" individually. In the third step, the computer chooses one of the two final decisions randomly to obtain the joint "group action" for each of the eight games. Only if two randomly matched groups both decide on identical group actions, they coordinate their behavior successfully.

The intra-group communication protocol constitutes a simple version of group discussion. The messages are written without previous intra-group communication and are only exchanged after each subject indicates a SD in all games. Hence, they reflect individual reasoning. Because the software chooses the group action randomly among individual final decisions, the messages represent the only opportunity for subjects to convince their group partner of their reasoning. This experimental setting incentivizes subjects to demonstrate their reasoning comprehensively within the messages (Burchardi and Penczynski, 2014). Section 5 presents a detailed discussion of this setting.

2.3 Experimental procedure

We conducted the experimental sessions in the Experimental Economics Laboratory at the University of Mannheim (mLab) and in the Experimental Economics Laboratory at the University of Heidelberg (AWI Lab).[6] Across nine sessions, a total of 142 undergraduate and postgraduate students participated. We recruited all students from the general student population of the corresponding host institutions. Out of the 142 students 31 were studying Economics, 28 of them enrolled in an undergraduate and three in a postgraduate program. To ensure that subjects are familiar with the experimental structure and understand the messaging system, all participants went through two unrelated test periods and had to answer four comprehension questions individually. We compensated all participants based on all games. The payoffs π_g indicate individual team members' payoffs in Taler where 1 Taler corresponds to 0.40 Euro. The average payoff per subject was 8.60 Euro. The subjects received their payoff subsequent to the experiment in private and cash.

2.4 Level-k model

In a standard level-k model, types differ by the number $k \in \mathbb{N}$ of iterated best responses that they apply to their belief of what level-0 players do. Level-k models often assume that level-0 players randomize their action uniformly over the action space (see Stahl and Wilson, 1995 or Camerer et al., 2004).

---

[6]The experiment was programmed and conducted with the software z-Tree (Fischbacher, 2007) and subjects were recruited with ORSEE (Greiner, 2004).

CGR model the level-0 player to both play label salient actions with higher probability in symmetric games and payoff salient actions with higher probability in asymmetric games. Following CGR, we propose a level-k model that we can later estimate on the basis of both action and classification data. Our model assumes that payoff salience governs level-0 beliefs with probability o, and governs label salience with the complementary probability $1 - o$. Level-0 payoff salience $\rho \in \{h, l\}$ reflects an attraction to either own high payoff actions (high payoff salience, h) or to the other group's high payoff action (low payoff salience, l). In contrast, level-0 label salience $\sigma \in \{\mathcal{A}\}$ reflects an attraction to a particular label $a \in \mathcal{A}$.

The best responses of players with level $k > 0$ are anchored in their level-0 belief. The responses imply choice probabilities P that reflect the predictions of the level-k model. $P_\rho(a, g, k, \rho)$ indicates the probability of an action a being taken by a group g member with level k that is in line with level-0 payoff salience ρ. $P_\sigma(a, k, \sigma)$ indicates the probability of an action a from a group g member with level k that is in line with label salience σ.

For example, in an asymmetric X-Y game, the belief that a level-0 player favors the action that gives her a high payoff ($\rho = h$) makes a group 1 level-1 player choose X and a group 2 player choose Y. With every best response iteration, the predicted action alternates. Similarly, the belief that a level-0 favors the label salient action X ($\sigma = X$) makes higher level players play X independent of the group assignment. Table 3 illustrates the choice probabilities P_ρ and P_σ for this example. Appendix A.5 shows P_ρ and P_σ for the other games.

Table 3: Choice probabilities $P_\rho(a, g, k, \rho)$, $P_\sigma(a, k, \sigma)$ in the asymmetric X-Y games.

	$P_\rho(a, g, k, \rho)$								$P_\sigma(a, k, \sigma)$			
	$\rho = h$				$\rho = l$				$\sigma = X$		$\sigma = Y$	
	$g = 1$		$g = 2$		$g = 1$		$g = 2$					
a	X	Y	X	Y	X	Y	X	Y	X	Y	X	Y
$k = 0$	0	1	1	0	1	0	0	1	1	0	0	1
$k = 1$	1	0	0	1	0	1	1	0	1	0	0	1
$k = 2$	0	1	1	0	1	0	0	1	1	0	0	1
$k = 3$	1	0	0	1	0	1	1	0	1	0	0	1
$k = 4$	0	1	1	0	1	0	0	1	1	0	0	1

2.5   Five classification criteria

We classify each written message according to five classification criteria: lower and upper bound of the level of reasoning, payoff and label salience in the level-0 belief, and team reasoning. We allow messages to be classified as level-k reasoning and team reasoning independently, because it is an empirical question whether and how they occur together. We ask for a lower and an upper bound on the level of reasoning in order to accommodate ambiguities in the messages

with respect to the level of reasoning.[7] Table 4 presents ten exemplary messages (ex. 1-10) and their classification in terms of level of reasoning and team reasoning. Appendix A.1 reports further examples for team reasoning.

The first classification criterion indicates the lower bound in the level of reasoning $\underline{k}_i$ that a message corresponds to most closely. The lower bound represents the lowest possible level of reasoning which is clearly stated in a message. Each message is assigned a lower bound of reasoning with $\underline{k}_i \in \{0, 1, 2, 3, 4, 5, \varnothing\}$, where $\varnothing$ indicates that the message does not reveal a bound.

The second classification criterion specifies the upper bound in the level of reasoning $\overline{k}_i$ that a message corresponds to most closely. The upper bound represents the highest level of reasoning which can possibly be interpreted into the messages.[8] Similar to the lower bound of reasoning, each message is assigned an upper bound of reasoning with $\overline{k}_i \in \{0, 1, 2, 3, 4, 5, \varnothing\}$ where $\overline{k}_i \geq \underline{k}_i$.

Each message is classified by two research assistants (RAs). In our analysis, we use agreeing classifications of the two RAs. In order to make use of further classification data in a cautious way, we incorporate disagreeing classifications from the RAs "A" and "B" as follows: $\underline{k}_i = \min\{\underline{k}_i^A, \underline{k}_i^B\}$ and $\overline{k}_i = \max\{\overline{k}_i^A, \overline{k}_i^B\}$.[9] This use of disagreeing classifications does not change our results.

The third classification criterion refers to the payoff salience ρ_i in level-0 beliefs of players with a level of reasoning $\overline{k}_i > 0$ (ex. 6-9). Level-0 players can be believed to react to payoff asymmetry with an attraction to their own high payoff actions (high payoff salience, h) or to the other group's high payoff action (low payoff salience, l). They might also be indifferent across payoffs ($\sim$) or not show payoff salience ($\varnothing$). We asked the RAs to indicate payoff salience in level-0 beliefs based on pre-defined criteria $\rho_i \in \{h, l, \sim, \varnothing\}$.

The fourth classification criterion indicates any form of label salience σ_i in the level-0 beliefs (ex. 1-6). Level-0 players can be attracted to a particular label ($a_i \in \mathcal{A}$) or be indifferent across labels ($\sim$). In case their level-0 belief does not incorporate label salience, they might explicitly express that fact or might not mention label salience in the written message at all ($\varnothing$). We asked the RAs to indicate label salience in level-0 beliefs based on pre-defined criteria $\sigma_i \in \{\mathcal{A}, \sim, \varnothing\}$.

For the two level-0 criteria, we use agreeing classifications of the two RAs. In order to carefully make use of further classification data, we use one RA's classification when the other

---

[7]On the basis of 43 classifications of the same set of messages, Eich and Penczynski (2016) show that level bounds are a useful measure to capture the content of ambiguous messages.

[8]The relevance of the upper bound as observed in the message relies on the assumption that subjects have sufficient incentives to articulate their full reasoning. In the light of the results, the discussion in section 5 addresses concerns that this might not be the case.

[9]We do not define these rules for an argument of $\varnothing$ since such classification is never involved in disagreement.

Table 4: Examples for level-k and level-0 belief classification.

Ex.	Subj.	Game	g	SD	Classification				TR	Message
					$\underline{k_i}$	$\overline{k_i}$	ρ_i	σ_i		
1	38	SL	1	X	0	0		X	1	X because it is in first place.
2	95	AM2	1	L	0	1	$\sim$	L	3	For this case, I'd definetely take the dollar-piece, because there's two other slices with 6,5.
3	135	S1	1	B	1	1		B	1	I suggest the paragraph, because its coloring is unique. It attracts attention easily and provides us with an opportunity to coordinate with the other team.
4	27	S1	1	B	1	1	$\sim$	B	1	It would again be the simplest to choose B. The payoffs are the same anyways. I don't think that anybody will think much more about this.
5	71	S2	1	B	1	1	$\sim$	B	3	I'd pick 5,5, because the answer is distinct and the other team might therefore choose 5,5 as well.
6	13	S2	1	L	1	1	h	L	2	Same game again, we should neglect the paragraph, cause it's unattractive. I would choose the $ slice, as it's proven that people usually prefer the first mentioned response option over the second one.
7	89	ALL	2	Y	1	2	h			Should we take the answer that is worse for us? I think it is likely that they take their best answer and then we are identical. I hope they don't think as laterally as we do...
8	80	ALL	1	Y	2	2	h			Most teams offer the other team more money and less for themselves because they think that the other will be more okay with that. I would therefore take Y because I think that the other teams decides like that.
9	81	ASL	2	X	2	2	h			Ok, for 4ct one can also give in. That is what they will think as well, therefore let us take X.
10	130	ASL	1	Y	3	5				It doesn't really matter what to choose. The other team will decide based on what they think how we might think about their action. It depends on how far we think.

classified no payoff salience $\varnothing$. This use does not change our results. We leave messages unclassified if both RAs indicate a different kind of salience.

The fifth classification criterion specifies team reasoning approaches that the written messages exhibit. In team reasoning, players look for a rule or strategy of determining an action that tends to produce successful coordination if both groups followed that rule. The two central elements in team reasoning are (1) the identification of a decision rule that "suggests itself" or seems obvious and (2) the intention to look for ways to coordinate behavior.[10] Following CGR, we distinguish three team reasoning approaches that can be used to improve the coordination probability.

The first class of rules uses labels to single out an action that could present a focus of convergence (TR1). This approach can be used in all our games. For example, in an asymmetric Pie game, players might perceive the B to stand out as the distinctly colored alternative and therefore rely on it for coordination (ex. 1, 3, 4). The second class of rules eliminates Pareto

---

[10]Bacharach (2006) proposes that team identification can prompt team reasoning, leading to the question "What should we do?" rather than "What should they do?" (p. 136). We only found two possible instances of this change of perspective in our communication transcripts and thus believe that the current games do not generate much team identification.

inferior or weakly Pareto inferior strategies to limit the scope of the action space (TR2). This team reasoning approach is of particular relevance in the Pie games S2 and AM4 (ex. 6). The third class of rules uses payoff characteristics to single out an action that could present a focus of convergence (TR3). For example, coordination on action L leads to a unique payoff pair of $(5, 6)$ in the asymmetric Pie game AM2 (ex. 2, 5).

We asked the RAs to indicate team reasoning based on the three pre-defined strategies presented above. In order to account for further team reasoning approaches that subjects might apply beyond these three strategies, the RAs could specify additional strategies in an open-ended response option. The RAs identified only 7 additional team reasoning approaches, such as "always choose first mentioned option" or "grant each group the higher payoff in turns", which we report under "other" TR strategies. We use those classifications that coincide between both RAs. As before, if one RA indicates some form of team reasoning TR while the other indicates ∅, we assign the message the TR classification. We leave messages unclassified if both RAs indicate different TR classifications.

The percentage of subjects with empty messages varies between 21% and 30% by game. Aligned with our aim of not interpreting more into a message than it in fact expresses, empty messages are left unclassified (∅). Alternative specifications could assign the most neutral category for some criteria, such as a lower bound of 0 or indifference ($\sim$) in the level-0 salience. In any case, this would not add new information and would therefore not alter the main message of our paper. Importantly, our approach does not lead to a focus on a particular subsample. The choice statistics only for those observations with classified messages in appendix A.3 show that this subsample behaves very similarly to the overall sample. Furthermore, including action information from empty message observations in the structural estimation does not lead to any surprising changes (table 26 in appendix A.7).

2.6 Classification process

The two RAs separately and independently read the communication transcripts and classified the content according to the five criteria. Both research assistants were postgraduate economics students at the University of Mannheim. We provided detailed instructions of the experimental setting and reproduced the main features of the level-k and team reasoning models. The classification instructions are reprinted in appendix B.4.

The RAs' remuneration is based on the number of matches for each initial classification that coincided between both assistants (Houser and Xiao, 2011). The RAs received 0.03 EUR for each *initial* classification that coincided between both assistants.[11]

---

[11]One might be worried that these coordination incentives distract RAs from the correct classification and rather have them focus on the other RA's judgement. For two reasons we believe such a concern to be unfounded. First,

Across nine sessions, 142 subjects sent a total of 851 messages to their group partner out of 1136 possible messages. In their initial classification, the RAs yielded an average agreement of 60% (2.540 matches/4.255≘5·851 classifications), ranging from 57% - 65% across the eight coordination games. In case the RAs did not agree on a classification, we informed both of them about the classification of the other assistant and gave them the – now unpaid – possibility to *individually* and *simultaneously* reconsider their classification. Note that only when exactly one of the two RAs changes the classification, the agreement rate improves. After revision, the agreement increased to an average of 82% (3.513/4.255). While the classification process is certainly non-trivial, these numbers indicate that messages can be used as effective coordination devices.

Two related studies support the fact that messages contain valuable information and allow for replicable classification. In Eich and Penczynski (2016), one dataset of 78 messages is classified 43 times by workers on Amazon Mechanical Turk. Even fix-paid non-experts yield substantial agreement of on average 33 out of 43 (77%) exact level agreements per message. In Penczynski (forthcoming), human classifications train a machine learning algorithm that correctly classifies 65% of messages out-of-sample. Thus, there is a mapping from plain word-counts to categories that describes and successfully replicates the classification effort, including a qualitative replication of all results of the present study.

Wordcounts can be used to show how messages and their categories relate to each other. "Wordclouds" visualize more frequent words with a larger fontsize. Figure 1 provides a snap-shot of the differences in messages between positive levels of reasoning and team reasoning TR1.[12] Apart from the overall common word "team", in messages classified with level reasoning $k \geq 1$ as shown in figure 1a, the most common words used are that ("dass") and think ("denk"). In contrast, messages classified with TR1 in figure 1b feature different words very frequently, such as white ("weiss"), first ("erst"), field ("feld"), and take ("nehm"). Appendix A.2 reports such wordclouds for all categories in this study.

_____

cognitively, the belief about the other RA's judgement will necessarily be informed through the own judgement since there is neither another access to a judgement nor a way to know that the other RA systematically judges differently. Second, other studies have used alternative remuneration schemes, such as no explicit payment (Cooper and Kagel, 2005), payment per hour (Burchardi and Penczynski, 2014) or payment per message (Eich and Penczynski, 2016), and have observed similar rates of initial agreement across various games.

[12]Results show original German words, stemmed to the word root. Common "stopwords" like articles ("the") and conjunctions ("and"), etc. are dropped.

(a) Level-k, $k \geq 1$.   (b) TR1.

Figure 1: Word frequencies by message characteristic.

3 Results

3.1 Suggested decisions and coordination rates

In order to compare the effectiveness of reasoning across the eight coordination games, we calculate the *expected* percentage of successful coordination in all possible pairings of individuals (as in Mehta et al., 1994; Crawford et al., 2008; Bardsley et al., 2010).[13] We use this index as measure of the extent of coordination. On the basis of the suggested decisions, we quantify the expected coordination between groups with:

$$c \equiv \sum_{a \in \mathcal{A}} \frac{N_1(a) \cdot N_2(a)}{N_1 \cdot N_2},$$

where $N_g(a)$ denotes the number of players from group $g \in \{1, 2\}$ that propose a strategy $a \in \mathcal{A}$, and N_g the total number of subjects in groups $g \in \{1, 2\}$.

The coordination rates and suggested decisions in the four X-Y games are presented in table 5.[14] Subjects realize frequent coordination in the payoff symmetric X-Y game (SL), but yield frequent miscoordination in X-Y games with asymmetric payoffs. Even slight payoff asymmetries cause the coordination rate to drop from 93% (SL) to 49% (ASL). The coordination rate across asymmetric games remains stable at around 50%. While most subjects (96%) suggest

---

[13]The *actual* percentage of coordination is just one realization of this expectation and represents a more noisy measure. Here, the process of group deliberation makes the eventual realization of coordination from final decisions even more noisy and less immediately relevant.

[14]Note that $N_1 = 76$ subjects were assigned group 1, and $N_2 = 66$ subjects were assigned group 2. The number of participating groups was odd in five experimental sessions. In such cases, two groups 1 were randomly paired with one group 2. That approach is innocuous as no sequential interaction takes place between groups.

the salient label X as joint group action in the symmetric game SL, they suggest group actions more heterogeneously in asymmetric games. Between 53% and 61% of all subjects favor their own payoff salient action (Y for group 1, X for group 2) across all degrees of payoff asymmetry.

Table 5: Suggested decisions in X-Y Games.

	X-Y Game			
	SL	ASL	AML	ALL
$\pi_1(X,X), \pi_2(X,X)$	5, 5	5, 5.1	5, 6	5, 10
$\pi_1(Y,Y), \pi_2(Y,Y)$	5, 5	5.1, 5	6, 5	10, 5
$N_1(X)$	71	34	36	36
$N_2(X)$	66	37	40	39
$N_1(Y)$	5	42	40	40
$N_2(Y)$	0	29	26	27
Coordination rate c	93%	49%	49%	50%
Coordination rate c in CGR	64%	38%	46%	47%

The suggested group decisions and coordination rates in the four Pie games are presented in table 6. Subjects achieve slightly higher coordination rates in symmetric games (S1 and S2) than in asymmetric games (AM2 and AM4). CGR observe higher coordination in S1 with 94% of the subjects opting for the label salient B, but their subjects miscoordinate more frequently in S2 and asymmetric games. While most subjects (55%) favor the label salient B in S1, the response pattern reverses in S2, in which only 25% of subjects choose the label salient, but dominated strategy B. Similar to the X-Y games, in the payoff asymmetric games AM2 and AM4 a substantial fraction of subjects proposes actions that are not payoff salient, such as the L slice for group 1.

For technical reasons, we introduce additional labels in the Pie games, among them the $-label for the L action. One concern is that the particular meaning of the $-label might raise the choice frequency of L in our experiment compared to CGR. In our data, we see in S1 that L is less frequently chosen than B and that overall only a dozen subjects mention the specific meaning of the $-label in their message.[15] Further, the analyses in the next sections show that in the presence of the $-label other factors shift the choice of the mode of reasoning across the Pie games. Hence, our particular choice of label is not influencing our main results.

***

[15]Of all 58 subjects with a label salience on L in the level-0 beliefs, a total of 38 subjects mention the $-label in their messages. Of those 38 messages, only 13 messages explicitly associated the $-labeled L slice with "money" or "profit". The remaining messages mostly mentioned the $-label along with the #- and §-label to refer to one specific slice ("definitely not §, let's choose between # and $").

Table 6: Suggested decisions in Pie Games

	Pie Game			
	S1	S2	AM2	AM4
$\pi_1(L, L), \pi_2(L, L)$ (\$)	5, 5	6, 6	5, 6	6, 7
$\pi_1(R, R), \pi_2(R, R)$ (#)	5, 5	6, 6	6, 5	7, 6
$\pi_1(B, B), \pi_2(B, B)$ (§)	5, 5	5, 5	6, 5	7, 5
$N_1(L)$	26	46	32	29
$N_2(L)$	27	37	24	24
$N_1(R)$	5	8	8	31
$N_2(R)$	6	14	10	32
$N_1(B)$	45	22	36	16
$N_2(B)$	33	15	32	10
Coordination rate c	44%	43%	40%	37%
Coordination rate c in CGR	89%	35%	31%	25%

3.2 Levels of reasoning

The classification of messages fully determines subjects' level of reasoning in 614 cases (72% with $\underline{k}_i = \overline{k}_i$), and assigns 237 messages a range in the level of reasoning (28% with $\underline{k}_i < \overline{k}_i$).[16]

Table 7 presents elicited level bounds of reasoning for the four X-Y games and the four Pie games. The marginal distributions of lower and upper level bounds show significant differences between payoff symmetric and payoff asymmetric games. In the payoff symmetric games, the level bound distributions show no entries (SL, S1) or very few entries (S2) for levels higher than 1. The large majority of subjects is classified not to best respond at all, with $\underline{k}_i = \overline{k}_i = 0$. In strong contrast, the level bounds are higher in payoff asymmetric games. Across the payoff asymmetric X-Y and Pie games, an average of 77% of the subjects show an upper bound on the level of reasoning of at least 1 and only around one quarter of all subjects act non-strategically. The level bound distributions have more entries for levels higher than 1 compared to the distributions in payoff symmetric games. Appendix A.4 reports results of the Fisher exact tests that reflect those differences in level distributions across games.

3.3 Salience considerations in the level-0 belief

The classification assigns 561 out of 851 written messages an upper level bound of reasoning of at least 1. Table 8 summarizes the classification of label salience in the level-0 beliefs of players

---

[16]The RAs agreed on the lower bound in the level of reasoning $\underline{k}_i$ in 702 of 851 cases (83%). Our analysis incorporates the remaining 149 disagreeing classifications from the two RAs "A" and "B" with $\underline{k}_i = \min\{\underline{k}_i^A, \underline{k}_i^B\}$. The RAs agreed on the upper bound in the level of reasoning $\overline{k}_i$ in 688 of 851 cases (81%). Our analysis uses the remaining 163 disagreeing classifications with $\overline{k}_i = \max\{\overline{k}_i^A, \overline{k}_i^B\}$.

Table 7: Overview levels of reasoning.

X-Y SL

no message: 37

		Level upper bound						
		0	1	2	3	4	5	$\sum$
	0	62	15	–	–	–	–	77
Level	1		28	–	–	–	–	28
lower	2			–	–	–	–	0
bound	3				–	–	–	0
	4					–	–	0
	5						–	0
	$\sum$	62	43	0	0	0	0	105

X-Y ASL

no message: 30

		Level upper bound						
		0	1	2	3	4	5	$\sum$
	0	25	11	4	–	1	–	41
Level	1		14	17	1	–	–	32
lower	2			30	4	–	–	34
bound	3				3	1	1	5
	4					–	–	0
	5						–	0
	$\sum$	25	25	51	8	2	1	112

X-Y AML

no message: 36

		Level upper bound						
		0	1	2	3	4	5	$\sum$
	0	22	10	4	2	–	–	38
Level	1		12	13	6	–	2	33
lower	2			25	3	1	1	30
bound	3				3	1	–	4
	4					–	1	1
	5						–	0
	$\sum$	22	22	42	14	2	4	106

X-Y ALL

no message: 38

		Level upper bound						
		0	1	2	3	4	5	$\sum$
	0	17	9	7	–	–	–	33
Level	1		23	14	2	–	–	39
lower	2			28	3	–	–	31
bound	3				1	–	–	1
	4					–	–	0
	5						–	0
	$\sum$	17	32	49	6	0	0	104

Pie S1

no message: 35

		Level upper bound						
		0	1	2	3	4	5	$\sum$
	0	64	13	–	–	–	–	77
Level	1		30	–	–	–	–	30
lower	2			–	–	–	–	0
bound	3				–	–	–	0
	4					–	–	0
	5						–	0
	$\sum$	64	43	0	0	0	0	107

Pie S2

no message: 32

		Level upper bound						
		0	1	2	3	4	5	$\sum$
	0	41	32	–	–	–	–	73
Level	1		33	2	1	–	–	36
lower	2			1	–	–	–	1
bound	3				–	–	–	0
	4					–	–	0
	5						–	0
	$\sum$	41	65	3	1	0	0	110

Pie AM2

no message: 42

		Level upper bound						
		0	1	2	3	4	5	$\sum$
	0	36	11	2	–	–	–	49
Level	1		35	7	–	–	–	42
lower	2			7	–	–	–	7
bound	3				2	–	–	2
	4					–	–	0
	5						–	0
	$\sum$	36	46	16	2	0	0	100

Pie AM4

no message: 35

		Level upper bound						
		0	1	2	3	4	5	$\sum$
	0	23	11	4	3	–	–	41
Level	1		28	16	–	–	–	44
lower	2			18	–	–	1	19
bound	3				2	–	–	2
	4					1	–	1
	5						–	0
	$\sum$	23	39	38	5	1	1	107

with a level of reasoning $\overline{k}_i > 0$.[17] The level-0 beliefs show that subjects strongly respond to the labeling of actions in the payoff symmetric SL and S1 game. Almost all considered subjects indicate a label salience, predominantly on the salient X or B. Conversely, the large majority of strategic subjects is found to *not* exhibit any label salience in the asymmetric games ASL, AML, ALL, and AM4. Whatever subjects focus on to take their decisions, it is clearly not the labeling of the options.

Table 9 gives information about payoff salience considerations detected in the messages.[18] No subject believes that level-0 players are attracted to the action with a lower own payoff. Because the payoff symmetric SL and S1 games do not feature any payoff differences across actions, these games never lead to an indicated payoff salience. On the other extreme, the large majority of strategic subjects assumes that level-0 players are attracted to the high payoff action in payoff asymmetric games ASL, AML, ALL, and AM4. In the X-Y games, the strength of this tendency does not depend on the degree of payoff asymmetry.

Table 8: Label salience by game.

Game	No salience	Prefer X	Prefer Y		Indifference	Disagree	Total
SL	1	41	1		0	0	43
ASL	83	4	0		0	0	87
AML	77	6	1		0	0	84
ALL	83	4	0		0	0	87
Game	No salience	Prefer B	Prefer L	Prefer R	Indifference	Disagree	Total
S1	0	25	17	1	0	0	43
S2	11	24	25	5	3	1	69
AM2	24	24	12	3	1	0	64
AM4	61	12	4	4	3	0	84
Total	340	140	60	13	7	1	561

Taking the two tables together, subjects assume that non-strategic level-0 players do not respond to salient labels in payoff asymmetric games, but rather choose actions in favor of their own high payoff actions.[19] Considerations in the X-Y games are switching strongly from label salience considerations in the symmetric game to payoff salience considerations in the asymmetric games. This switch is more gradual in the Pie games, possibly because of the more

---

[17]The RAs agreed on the label salience criterion in 537 of 561 cases (96%). In 23 of the remaining 24 cases one RA indicates some form of label salience ($\sigma_i \in \{\mathcal{A}, \sim\}$), while the other RA does not ($\varnothing$). 1 message remains unclassified as both RAs indicate different label salience.

[18]The RAs showed consistent payoff salience classifications in 461 of 561 cases (82%). In 74 of the remaining 100 cases, one RA indicates some form of payoff salience ($\rho_i \in \{h, l, \sim\}$), while the other does not ($\varnothing$). 26 messages remain unclassified as both RAs indicate different payoff salience.

[19]Looking at both kinds of salience over all games, the two kinds of salience are almost mutual exclusive. Only 38 messages exhibit simultaneously a payoff and a label salience in the level-0 beliefs.

Table 9: Payoff salience by game.

Game	No salience	High payoffs	Indifference	Disagree	Total
SL	31	0	12	0	43
ASL	12	69	0	6	87
AML	9	67	0	8	84
ALL	8	74	0	5	87
S1	33	0	10	0	43
S2	37	25	7	0	69
AM2	26	29	6	3	64
AM4	17	63	0	4	84
Total	173	327	35	26	561

(The header cell "Payoff salience" spans the columns No salience, High payoffs, Indifference, Disagree.)

complex setting. While basically everybody believes that non-strategic level-0 players respond to the salient B and L in the symmetric game S1, only roughly 20% of the subjects do so in the asymmetric game AM4 where 75% of subjects focus on payoff salience.

3.4 Level-0 actions

Table 10 reflects the actions of players whose messages have been classified with coinciding level-k bounds of 0. For the symmetric SL game, both groups' level-0 actions strongly favor the label salient action X. This regularity is much weaker in the asymmetric X-Y games. While for group 2 the label and payoff salience point in the same direction X, group 1 players are split between X and Y with a slight preference for the payoff salient Y. This is the first result in the paper that shows significant differences between groups 1 and 2. This observation suggests that level-0 players are indeed influenced by salient payoffs. In contrast to the level-0 beliefs, however, label salience still matters as well.

In the symmetric Pie games, the actions L and B are chosen most often. In AM2, group 1 embraces the salient B's higher payoff while group 1 is split between the B and the more favorable L. In AM4, the payoffs of L and R seem to have different effects on level-0 players of the two groups.

3.5 Team reasoning strategies

The results on team reasoning in table 11 indicate how subjects intend to achieve coordination between the two groups.[20] The intention to use labels as a coordination device appears in more than 200 messages (TR1). It can be observed that these intentions are frequently arising in

---

[20]The RAs agreed in 739 of 851 cases (87%). In 83 of 112 cases, one RA indicates some form of team reasoning (TR), while the other does not ($\varnothing$). 29 messages remain unclassified as both RAs indicate different team reasoning approaches.

Table 10: Actions of level-0 players.

| | Group 1 | | | | Group 2 | | | |
Game	X	Y		Total	X	Y		Total
SL	26	2		28	34	0		34
ASL	5	6		11	10	4		14
AML***	3	7		10	12	0		12
ALL	5	4		9	7	1		8
Game	B	L	R	Total	B	L	R	Total
S1	21	11	2	34	16	10	4	30
S2	5	9	3	17	4	14	6	24
AM2	8	4	3	15	9	9	3	21
AM4	3	4	6	13	2	5	3	10

Notes: *** indicates a statistically significant difference between groups at the 0.01 confidence level (Fisher exact test).

payoff symmetric games that feature a strong label salience in the level-0 belief and only rarely in payoff asymmetric X-Y games that feature a strong payoff salience in the level-0 belief.

In contrast to TR1, team reasoning strategies TR2 and TR3 are used very infrequently and game-specifically. The intention to eliminate Pareto inferior strategies (TR2) predominantly appears in the payoff asymmetric game AM4. In this game, subjects realize that the label salient bottom slice B is weakly Pareto inferior to R. Consistently, the concept of Pareto inferiority is not found in the X-Y games, because eliminating a strategy does not yield further Pareto improvements. The intention to rely on payoff characteristics to single out an action (TR3) is applied least often and mainly appears in the asymmetric game AM2. Subjects realize that L leads to a unique payoff pair that could present a focus of convergence for both teams.

Table 11: Team reasoning by game.

	Team reasoning						
Game	None	TR1 Label	TR2 Pareto	TR3 Payoff	Other	Disagree	Total
SL	60	45	0	0	0	0	105
ASL	105	4	0	1	1	1	112
AML	99	4	0	1	1	1	106
ALL	96	4	0	1	3	0	104
S1	40	66	0	0	0	1	107
S2	42	42	7	4	0	15	110
AM2	52	32	0	10	0	6	100
AM4	68	12	19	1	2	5	107
Total	562	209	26	18	7	29	851

3.6 Individual data

Our within-subject design enables a comparison of reasoning across games on the individual level. The panel regressions in table 12 show how the classification outcomes relate to the game features.[21] As expected from the results before, the symmetric game dummy is significantly associated with lower level bounds (regressions 1 and 2) and a higher incidence of team reasoning (5). Naturally, payoff salience is not detected in symmetric games (3), but a label salient level-0 belief on X or B is significantly more prevalent in symmetric games (4). Whether or not the symmetric games were played first within the session only matters very slightly for label salience in an expected way: the focus on labels at the beginning of the session raises its average relevance for the entire session. The effects of the X-Y game dummy mostly reflect the relevance of the payoff asymmetry in three games compared to only two Pie games. Effects of the game sequence are only significant for the level lower bound, but they are comparatively small.

Table 12: Panel regression.

Dependent variable	$\underline{k}$	$\overline{k}$	$\rho_i = h$ (dummy)	$\sigma_i \in \{X, B\}$ (dummy)	any TR (dummy)
	(1)	(2)	(3)	(4)	(5)
Symmetric game	-0.50***	-0.74***	-0.43***	0.19***	0.26***
	(0.05)	(0.06)	(0.03)	(0.03)	(0.03)
Symmetric game first	0.01	0.05	-0.00	0.09**	0.02
(Sessions 7-9, dummy)	(0.07)	(0.09)	(0.04)	(0.04)	(0.04)
X-Y game	0.15***	0.25***	0.12***	-0.02	-0.23***
	(0.05)	(0.05)	(0.03)	(0.03)	(0.03)
Sequence index t	-0.04***	-0.01	0.00	-0.01	0.01
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Constant	0.95***	1.19***	0.43***	0.09***	0.29***
	(0.09)	(0.10)	(0.05)	(0.03)	(0.04)
N	851	851	851	851	851
Subjects	133	133	133	133	133
R^2 overall	0.16	0.20	0.24	0.08	0.18

Notes: Panel random-effects regressions. Cluster-robust standard errors (subject level) are provided in parentheses. ***, ** and * indicate significance at the 1%, 5% and 10% level.

---

[21]For further illustration, table 23 in appendix A.6 presents the raw sets of classifications of all subjects that have been level classified in at least 7 out of 8 games.

19

4 Estimation

We analyse the level-k distribution and level-0 belief in more detail with the help of a maximum likelihood estimation (MLE). The estimation uses individual data on the actions $a_i \in \mathcal{A}$, group $g_i \in \{1, 2\}$, level bounds $K_i = \{\underline{k}_i, \overline{k}_i\}$, payoff salience $\rho_i \in \{h, l\}$, and label salience $\sigma_i \in \{\mathcal{A}\}$.

Based on this data, we calculate the likelihood for the estimates of the level-k fraction l_k, $k \in \{0, 1, 2, 3, 4\}$, the probability of payoff salience o vs. label salience $1 - o$, the probability of salience of the own high or low payoff action r_j, $j \in \{h, l\}$ and the probability of salience on a particular action s_a, $a \in \mathcal{A}$. The choice probabilities $P_\rho(a_i, g_i, k, \rho_i)$ and $P_\sigma(a_i, k, \sigma_i)$ reflect the predictions of the level-k model as illustrated in table 3 and further in appendix A.5. If one kind of salience is observed, the probability P of the respective other kind is set to 0. If both kinds of salience are observed, results are based on having them enter under payoff salience, but the alternative specification differs only slightly.[22] Further, observations that have neither a level-k nor a level-0 classification (including indifference $\sim$) are disregarded in this estimation. In order to accommodate errors, we attribute behavior that the model expects with probability 0, a small, positive likelihood ϵ. Note, we do not make assumptions about level-0 beliefs, instead we use the classification data on subjects' level-0 beliefs in order to calculate choice probabilities.

Under the assumption of independent actions, the log-likelihood function is

$$
\log L(l_k, o, r_j, s_a; a_i, g_i, k_i, \rho_i, \sigma_i) =
$$
$$
\sum_{i=1}^{N} \log \left(\sum_{k \in K_i} l_k \cdot \left\{ (1 - \epsilon) \left[o \cdot \sum_{j \in \{h,l\}} r_j \cdot P_\rho(\cdot, \rho_i = j) + (1 - o) \sum_{a \in \mathcal{A}} s_a \cdot P_\sigma(\cdot, \sigma_i = a) \right] \right. \right.
$$
$$
\left. \left. + \epsilon \left[o \cdot \sum_{j \in \{h,l\}} r_j \cdot P_\rho(\cdot, \rho_i = \{h, l\} \setminus j) + (1 - o) \cdot s_{a_i} \cdot P_\sigma(\cdot, \sigma_i = a_i). \right] \right\} \right)
$$

Table 13 presents the estimation results for the individual games. It can be observed that the level distributions differ by the game in a systematic way. The symmetric games SL, S1 and S2 have average levels below 1 and no players of a level higher than 1. In contrast, the more asymmetric games ASL, AML, ALL, AM2 and AM4 have average levels higher than 1 and mostly substantial fractions of level 2 players. The level averages are negatively correlated with the coordination rates, suggesting a link between miscoordination in payoff-asymmetric coordination games and higher levels of reasoning.

---

[22]Appendix A.7 presents alternative specifications on the basis of coinciding classifications only (table 25), the full sample (table 26) and subjects with two saliences entering the estimation under their label salience (table 27).

Table 13: Estimation results by game.

		XY games					Pie games			
		SL	ASL	AML	ALL		**S1**	**S2**	AM2	AM4
Level-0	l_0	0.64	0.26	0.22	0.27	l_0	0.64	0.48	0.40	0.27
Level-1	l_1	0.36	0.30	0.29	0.34	l_1	0.36	0.48	0.45	0.41
Level-2	l_2	0.00	0.36	0.34	0.34	l_2	0.00	0.03	0.13	0.27
Level-3	l_3	0.00	0.06	0.11	0.05	l_3	0.00	0.01	0.02	0.04
Level-4	l_4	0.00	0.01	0.04	0.00	l_4	0.00	0.00	0.00	0.01
Average	$\hat{k}$	0.36	1.27	1.44	1.16	$\hat{k}$	0.36	0.56	0.76	1.12
Payoff/label sal.	o	0.00	0.78	0.82	0.86	o	0.00	0.60	0.50	0.75
Payoff sal.	r_h	–	1.00	1.00	1.00	r_h	–	1.00	1.00	1.00
Label sal.	s_X	0.98	0.60	0.67	0.80	s_B	0.57	0.80	0.57	0.48
	s_Y	0.02	0.40	0.33	0.20	s_L	0.38	0.00	0.28	0.33
						s_R	0.06	0.20	0.15	0.19
	N	105	111	106	104	N	107	110	100	107

Notes: 38 messages with both payoff and label salience enter estimation with their payoff salience.

Regarding the level-0 belief, around 80% of players are estimated to be influenced by the payoff salience in the asymmetric X-Y games (o). In the Pie games, these numbers are slightly lower in S2, AM2, and AM4. For all these games, players believe that the level-0 player is attracted to the action that yields the higher payoff for herself, $r_h = 1$. As expected, the label salience results predominantly from the X rather than the Y. In the Pie games, the two most salient options are L and B, reflecting the competing salience between these two actions.

5 Discussion

Our results show that incentivized messages yield rich information about subjects' reasoning. For the interpretation of our results we need to be aware of some limitations. In particular, reasoning can be inferred immediately from articulated arguments in a message – for example, the lower bound of level reasoning – but it is not straightforward to infer which reasoning has not been applied. For such inference, we need to assume that incentives are sufficiently high for subjects to indeed communicate their full reasoning. In this study, we interpret the observation of low level upper bounds in the symmetric games as evidence that further iterative best responses are not applied and that reasoning systematically differs between symmetric and asymmetric games. But what about the possibility that not the reasoning differs by games, but the communication changes between games, and a possibly *unchanged* level of reasoning is simply not articulated and thus not detectable in some games?

We will not be able to fully exclude this latter possibility, but we can offer some evidence that prove it to be unlikely. First, both in our asymmetric games and in other studies, the communication uncovers higher level reasoning showing that subjects indeed articulate in other contexts what we do not see articulated in symmetric games (Burchardi and Penczynski, 2014; Penczynski, 2016b). Why would they stop? Any answer to this question that refers to the lacking incentives of articulation has to also state why – given lacking incentives – these considerations would even arise. We revisit this interesting aspect in a moment. Second, across various studies, no systematic discrepancy between the proposal and the written message has been found that could indicate a lack of incentives to complete the written argument. For example, the beauty contest analysis in Penczynski (2016a) estimates level-0 beliefs on the basis of observed proposals and level classifications and yields an estimated level-0 belief mean of 53.45, within 2 units of the average classification of explicitly mentioned level-0 belief means. Any omission and level underclassification would have led to a systematic underestimation of the starting point of reasoning. Admittedly, the likelihood of incomplete statements is higher when partial arguments lead to the same proposal, but a lack of incentives and an implicit assumption that the partner understands a partial account of arguments could exist already in other situations.

Third, evidence suggests that individual subjects' levels of reasoning usually vary between games (Georganas et al., 2015), therefore, no strong prior makes us expect individual reasoning to be stable across games. Further, intuition and preliminary evidence have been brought forward that the average sophistication might well vary between games, for example, as a function of game complexity (Crawford and Iriberri, 2007; Arad and Rubinstein, 2012; Koch and Penczynski, 2018). Finally, indirect evidence about what is not communicated exists from analyses of the final decision after the communication. Penczynski (2016a) finds that subjects

predominantly adopt the received proposal when the team partner puts forward higher level arguments than they do, but subjects do not adopt otherwise. If subjects, however, were systematically more sophisticated than they articulate, the adoption of a more sophisticated choice should not depend in that way on the other's sophistication.

This indirect evidence from final decisions suggests that the intra-group communication is not used strategically between interest-aligned group partners to the effect that reasoning and message differ. If a subject deliberately simplifies the message or misrepresents her sophistication to compel her group partner to accept the proposal, then it is not clear why the final decision should depend as much on the proposal of a more articulate group partner as observed in Penczynski (2016a). More fundamentally, if different arguments lead to different decisions, it is not clear how one should have a better way of making the own proposal convincing other than the own reasoning. And, if an easy and a difficult argument led to the same proposal, naming both is usually more persuasive than mentioning only the easy one. Finally, given the limited strategic sophistication observed in the level-k literature and within this framework, it is doubtful whether in these group-incentivized situations subjects approach the incentive-aligned group partner in a very strategic way.

Overall, the interpretation of the results in favor of reasoning differences between games is in accordance with many observations in the literature. The possibility that the completeness of the communication differs between games or that the strategic situation within groups distorts the communication cannot be ruled out but lacks convincing supporting evidence.

6 Conclusion

Our analysis of coordination games replicates the stylized facts of successful coordination under symmetric payoffs and of frequent miscoordination under asymmetric payoffs. Thanks to incentivized written accounts of reasoning from an experimental intra-group communication design (Burchardi and Penczynski, 2014), we can furthermore observe the mechanisms behind these observations.

As conjectured by CGR, the miscoordination in asymmetric games comes along with clearly detected level-k reasoning that features the typical characteristics of hump-shaped distributions and level means between 1 and 1.5. Level-0 players are mostly believed to play the action that yields the higher payoff for them and are only rarely deemed to be influenced by label salience. Only in the absence of payoff asymmetries the reasoning is focused on the framing of the game and features team reasoning. Indeed, the coordination attempts in symmetric games come along with a strong focus on label salience and the intention to use this for coordination. The level distribution is atypically low and does not feature levels beyond 1.

If we had to write an algorithm for predicting the mode of reasoning in payoff-asymmetric coordination games, we would set level-k reasoning and payoff salience as the default mode. Despite its ineffectiveness for coordination, it is rarely abandoned in payoff-asymmetric games, probably only in games with obvious and high expected benefits of team reasoning (Faillo et al., 2013). In symmetric games without payoff salience, team reasoning approaches fruitfully use labels for coordination and constitute a useful default for prediction.

Our study clearly illustrates how small differences between strategic situations have a strong influence on reasoning. The case of team reasoning vs. level-k reasoning in coordination games thus highlights the importance of understanding how subjects approach a game. We believe that in future work the combination of written accounts of reasoning and these theories will improve our understanding of the paths of strategic reasoning.

References

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).

Bacharach, Michael, *Beyond individual choice: teams and frames in game theory*, Princeton University Press, 2006.

__ **and Michele Bernasconi**, "The Variable Frame Theory of Focal Points: An Experimental Study," *Games and Economic Behavior*, April 1997, *19* (1), 1–45.

Bardsley, Nicholas, Judith Mehta, Chris Starmer, and Robert Sudgen, "Explaining Focal Points: Cognitive Hierachy Theory *Versus* Team Reasoning," *The Economic Journal*, 2010, *120 (543)*, 40–79.

Blume, Andreas and Uri Gneezy, "An Experimental Investigation of Optimal Learning in Coordination Games," *Journal of Economic Theory*, 2000, *90*, 161–172.

__ **and** __ , "Cognitive forward induction and coordination without common knowledge: An experimental study," *Games and Economic Behavior*, 2010, *68*, 488–511.

Burchardi, Konrad B. and Stefan P. Penczynski, "Out of your mind: Eliciting individual reasoning in one shot games," *Games and Economic Behavior*, 2014, *84* (0), 39 – 57.

Cachon, Gerard P. and Colin F. Camerer, "Loss-avoidance and forward induction in experimental coordination games," *The Quarterly Journal of Economics*, 1996, pp. 165–194.

Camerer, Colin and Teck-Hua Ho, "Experience-weighted attraction learning in coordination games: Probability rules, heterogeneity, and time-variation," *Journal of mathematical psychology*, 1998, *42* (2), 305–326.

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.

Cooper, David J. and John H. Kagel, "Are Two Heads Better than One? Team versus Individual Play in Signaling Games," *American Economic Review*, June 2005, *95* (3), 477–509.

Cooper, Russell, Douglas V DeJong, Robert Forsythe, and Thomas W Ross, "Communication in the battle of the sexes game: some experimental results," *The RAND Journal of Economics*, 1989, pp. 568–587.

Cooper, Russell W., Douglas V. DeJong, Robert Forsythe, and Thomas W. Ross, "Selection Criteria in Coordination Games: Some Experimental Results," *American Economic Review*, March 1990, *80* (1), 218–33.

__ , __ , __ , **and** __ , "Communication in Coordination Games," *The Quarterly Journal of Economics*, May 1992, *107* (2), 739–71.

Costa-Gomes, Miguel A., Vincent P. Crawford, and Bruno Broseta, "Cognition and Behavior in Normal-Form Games: An Experimental Study," *Econometrica*, September 2001, *69* (5), 1193–1235.

Crawford, Vincent P., "Adaptive dynamics in coordination games," *Econometrica*, 1995, pp. 103–143.

__ **and Bruno Broseta**, "What Price Coordination? The Efficiency-Enhancing Effect of Auctioning the Right to Play," *The American Economic Review*, 1998, *88* (1), 198–225.

__ **and Nagore Iriberri**, "Fatal Attraction: Focality, Naïveté and Sophistication in Experimental 'Hide and Seek' Games," *American Economic Review*, December 2007, *97* (5), 1731–1750.

__ **, Uri Gneezy, and Yuval Rottenstreich**, "The Power of Focal Points Is Limited: Even Minute Payoff Asymmetry May Yield Large Coordination Failures," *American Economic Review*, July 2008, *98* (4), 1443–1458.

Eich, Theresa and Stefan P. Penczynski, "On the replicability of intra-team communication classification," Working Paper, University of Mannheim 2016.

Ellingsen, Tore and Robert Östling, "When Does Communication Improve Coordination?," *American Economic Review*, 2010, *100* (4), 1695–1724.

Faillo, Marco, Alessandra Smerilli, and Robert Sugden, "The roles of level-k and team reasoning in solving coordination games," *University of Trento Cognitive and Experimental Economics Laboratory Working Paper*, 2013, pp. 6–13.

Farrell, Joseph, "Cheap talk, coordination, and entry," *The RAND Journal of Economics*, 1987, pp. 34–39.

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

Georganas, Sotiris, Paul J Healy, and Roberto A Weber, "On the persistence of strategic sophistication," *Journal of Economic Theory*, 2015, *159*, 369–400.

Greiner, Ben, "The online recruitment system orsee 2.0-a guide for the organization of experiments in economics," *University of Cologne, Working paper series in economics*, 2004, *10* (23), 63–104.

Holm, Håkan J, "Gender-based focal points," *Games and Economic Behavior*, 2000, *32* (2), 292–314.

Houser, Daniel and Erte Xiao, "Classification of natural language messages using a coordination game," *Experimental Economics*, 2011, *14* (1), 1–14.

Huyck, John B. Van, Ann B. Gillette, and Raymond C. Battalio, "Credible assignments in coordination games," *Games and Economic Behavior*, October 1992, *4* (4), 606–626.

Huyck, John B. Van, Joseph P. Cook, and Raymond C. Battalio, "Adaptive behavior and coordination failure," *Journal of Economic Behavior & Organization*, April 1997, *32* (4), 483–503.

Huyck, John B. Van, Raymond C. Battalio, and Richard O. Beil, "Tacit Coordination Games, Strategic Uncertainty, and Coordination Failure," *The American Economic Review*, 1990, *80* (1), 234–248.

__ , __ , **and** __ , "Strategic Uncertainty, Equilibrium Selection, and Coordination Failure in Average Opinion Games," *The Quarterly Journal of Economics*, 1991, *106* (3), 885–910.

__ , __ , **and** __ , "Asset markets as an equilibrium selection mechanism: Coordination failure, game form auctions, and tacit communication," *Games and Economic Behavior*, 1993, *5* (3), 485–504.

Isoni, Andrea, Anders Poulsen, Robert Sugden, and Kei Tsutsui, "Focal points in tacit bargaining problems: Experimental evidence," *European Economic Review*, 2013, *59*, 167–188.

Koch, Christian and Stefan P. Penczynski, "The Winner's Curse: Conditional Reasoning & Belief Formation," *Journal of Economic Theory*, 2018, *174*, 57–102.

Mehta, Judith, Chris Starmer, and Robert Sugden, "The Nature of Salience: An Experimental Investigation of Pure Coordination Games," *American Economic Review*, June 1994, *84* (3), 658–673.

Nagel, Rosemarie, "Unraveling in Guessing Games: An Experimental Study," *American Economic Review*, December 1995, *85* (5), 1313–1326.

Penczynski, Stefan P., "Persuasion: An experimental study of team decision making," *Journal of Economic Psychology*, 2016, *56*, 244–261.

__ , "Strategic Thinking: The Influence of the Game," *Journal of Economic Behavior and Organization*, 2016, *128*, 72–84.

__ , "Using machine learning for intra-team communication classification," *Experimental Economics*, forthcoming.

Rubinstein, Ariel, "Experience from a Course in Game Theory: Pre- and Postclass Problem Sets as a Didactic Device," *Games and Economic Behavior*, July 1999, *28* (1), 155–170.

Schelling, Thomas C., *The Strategy of Conflict*, Harvard University Press, 1960.

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.

Sugden, Robert, "Thinking as a team: Towards an explanation of nonselfish behavior," *Social Philosophy and Policy*, 1993, *10* (01), 69–89.

A Appendix

A.1 Team reasoning examples

Table 14: English translation of selected messages that show team reasoning approaches

TR1 - Labels as coordination device

Subj.	Game	g	SD	$\underline{k}_i$	$\overline{k}_i$	ρ_i	σ_i	TR	Message
7	SL	2	X	1	1		X	1	"every team earns the same payoff again.... X, because it is read first and probably chosen by both teams."
134	ALL	2	X	2	2	h	X	1	"oh! to choose x would be consistent, but does team 1 really think that way and give up the higher payoff? they might think the same way we do. let's choose x, because it is mentioned first and it would be good decision rule to choose the first mentioned alternative."
24	S1	1	B	1	1	~	B	1	"All options are similar, it's just a guessing game. The $ slice possibly stands out most of all, because it is coloured white - the other team might select this slice."
142	S2	2	L	0	1	h	L	1	"We should pic one of the slices with the higher payoffs. In my opinion, people usually prefer the left one over the right one."
8	AM4	2	B	0	1		B	1	"Allthough the gray fields show a fair payoff structure, we should go with the visually predominent one..."

TR2 - Pareto dominance arguments

Subj.	Game	g	SD	$\underline{k}_i$	$\overline{k}_i$	ρ_i	σ_i	TR	Message
13	S2	1	L	1	1	h	L	2	"Same game again, we should neglect the paragraph, cause it's unattractive. I would choose the $ slice, as it's proven that people usually prefer the first mentioned response option over the second one."
29	S2	1	L	1	1	h		2	"6:6 no doubt! The question on hand is whether to choose $ or #. There's a 50-50 chance of coordination. I don't think that the other team has a clear preference for one of the two options."
5	AM4	1	R	1	1	h		2	"Definitely not the bottom one. In addition, I want much Taler, which suggests the # piece. The other team also receives 6 instead of 5 Taler for the # piece."
76	AM4	1	L	0	1	h		2	"I think we can exlude §from further consideration, because it offers the lowest total amount of taler - only 12 - compared to the other two slices, which offer 13 taler. I would choose $ because I preferred higher payoffs last round."
26	AM4	2	R	2	2	h		2	"either # or $. The other team won't pic §, because they think that we would definitely not choose that slice and would not accept lower payoffs. Let's go with #."

TR3 - Unique payoff structure

Subj.	Game	G	SD	$\underline{k}_i$	$\overline{k}_i$	ρ_i	σ_i	TR	Message
27	S2	1	B	0	1		B	3	"with two 6-6 alternatives there's a lower chance of coordination, as they are similar."
71	S2	1	B	1	1	~	B	3	"I'd pick 5,5, because the answer is distinct and the other team might therefore choose 5,5 as well."
10	AM2	1	L	1	2	h		3	"The other team will mostly likely act rationally and go for the left one, because the payoff distribution between both teams plays a minor role. Another complication is that, if the other team wants to grant us the bigger share, they can't be sure whether we pick right or bottom. I think they won't base their decision on colours."
95	AM2	1	L	0	1	~	L	3	"For this case, I'd definetely take the dollar-piece, because there's two other slices with 6,5."
14	AM4	1	L	1	1	h		3	"I think the other team will pick $ again. First, because 7 taler is the highest possible payoff for them; Second: otherwise there are two alternatives that offer 7 taler for our team."

Table 15: Original version of selected messages that show team reasoning approaches

TR1 - Labels as coordination device

Subj.	Game	g	SD	$\underline{k}_i$	$\overline{k}_i$	ρ_i	σ_i	TR	Message
7	SL	2	X	1	1		X	1	"jedes team bekommt wieder da gleiche... X. weil das als erstes gelesen wird und deshalb vllt gewählt wird wenn beides das selbe ist."
134	ALL	2	X	2	2	h	X	1	"oh. x wäre konsequent aber denkt team 1 auch so und verzichtet? sie werden das gleiche über uns denken und x vor y ist der einzige anhaltspunkt für eine gemeinsame entscheidung ohne kommunkikation. daher x"
24	S1	1	B	1	1	$\sim$	B	1	"Alle Optionen sind gleich, daher ist das ganze nur ein Ratespiel. Evtl. sticht \$ am meisten heraus, da dieses Feld weiß ist - kann daher sein, dass das andere Team daher dieses nimmt."
142	S2	2	L	0	1	h	L	1	"Höherer Betrag und dann ist das linke m.M.n. das erste Feld, das man auswählt."
8	AM4	2	B	0	1		B	1	"Wieder das Auffälligste, auch wenn die anderen grauen Felder fairer von der Aufteilung sind..."

TR2 - Pareto dominance arguments

Subj.	Game	G	SD	$\underline{k}_i$	$\overline{k}_i$	ρ_i	σ_i	TR	Message
13	S2	1	L	1	1	h	L	2	"Selbes Spiel wie vorhin, Paragraphen außer Acht lassen, da unattraktiv. Würde \$ nehmen, da es nachgewiesen ist, dass eher die erste Antwortmöglichkeit genommen wird als die zweite - bei selber Bedeutung."
29	S2	1	L	1	1	h		2	"6:6 keine Frage. Die Frage ist nur \$ oder #. Ich denke mal, dass das ne 50/50-Entscheidung ist. Glaube nicht, dass die anderen da auch so ne eindeutige Präferenz für eins der beiden Möglichkeiten haben."
5	AM4	1	R	1	1	h		2	"Mittleres Feld lässt sich ausschließen. Und dann möchte ich einfach mehr Taler, also Raute-Feld. Schließlich bekommen in dem Feld Team 2 ganze 6 anstatt nur 5."
76	AM4	1	L	0	1	h		2	"Ich denke mal §können wir ausschließen, weil es insgesamt gesehen die niedrigste auszahlung ist, nämlich 12 und die anderen beiden jeweils 13 taler ergeben ich würde \$ wählen, weil ich bei der letzten entscheidung unsere höhere auszahlung bevorzugt habe."
26	AM4	2	R	2	2	h		2	"also entweder # oder \$ weil §werden sie nicht wählen, da wir das auf gar keinen fall wählen würden, da wir uns damit viel schlechter stellen würden also ich bin für #."

TR3 - Unique payoff structure

Subj.	Game	g	SD	$\underline{k}_i$	$\overline{k}_i$	ρ_i	σ_i	TR	Message
27	S2	1	B	0	1		B	3	"bei 6,6 alternativen ist die wahrscheinlichkeit niedriger da es zwei von solchen alternativen gibt."
71	S2	1	B	1	1	$\sim$	B	3	"ich würde 5,5 nehmen, da sich die antwort von den anderen unterscheidet und das andere team sie deshalb nehmen könnte."
10	AM2	1	L	1	2	h		3	"Ich denke sie werden rational handeln und sich für links entscheiden, da hier die Verteilung zwischen beiden Teams hier eine geringere Rolle spielt. es gibt ja die zusätzliche Verkomplizierung für das andere Team, wenn sie großzügig sein wollen, dass wir per zufall die andere option aus den 2 mit gleichen auszahlungen auswählen mit den farben werden sie denke ich nicht spielen."
95	AM2	1	L	0	1	$\sim$	L	3	"Hier würde ich jetzt definitiv auf die Dollar-Variante klicken, da nur hier sichergestellt ist, das wir bei dieser Entscheidung das gleiche Feld treffen. Bei Entscheidung für 6,5 gibt es ja zwei Felder."
14	AM4	1	L	1	1	h		3	"Ich denke sie würden wieder \$ auswählen, zum einen weil 7 das größte payoff für sie ist zum anderen weil wir ansonsten 2 Alternativen haben mit dem selben payoff von 7 was das maximum ist."

A.2 Wordclouds

Showing the word frequencies for each level classification of the message, one can observe an interesting transition from level-0 (2a) to level-3 (2d). While take ("nehm"), white ("weiss"), same ("gleich"), first ("erst") are some of the most common words in level-0 messages, levels 1 and 2 feature most prominently "team" and that ("dass"). The incidence of think ("denk") is steadily rising in levels 1 and 2, becoming the most frequent word in level-3. Table 16 summarizes the most frequent used words of all wordclouds and provides their English translation.

(a) Level-0.

(b) Level-1.

(c) Level-2.

(d) Level-3.

Figure 2: Word frequencies by level.

Figure 3 shows that words frequencies between level-0 belief categories "no payoff salience" and "high payoff salience" are very similar. Still, the computer classification can predict 66% of the human classification correctly (Penczynski, forthcoming).

(a) No payoff salience.

(b) High payoff salience.

Figure 3: Word frequencies by payoff salience.

It can further be shown that both in X-Y games (4a, 4b) as well as in Pie games (5a, 5b), label salience on X or on the white B Pie slice is quite distinct from no label salience.

(a) No label salience.

(b) Label salience on X ($X \succ Y$).

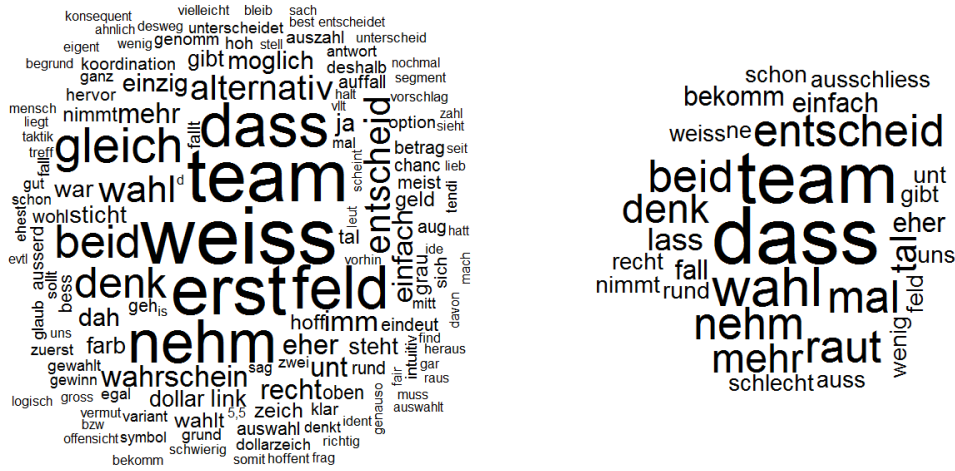Figure 4: Word frequencies in X-Y games by label salience.

(a) No label salience.
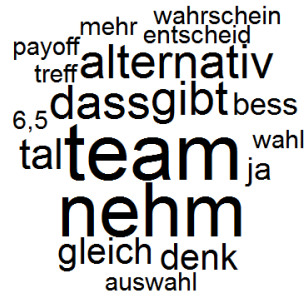
(b) Label salience on B (White).

Figure 5: Word frequencies in Pie games by label salience.

Finally, figure 6 illustrates word frequencies across different categories of team reasoning.

(a) TR1.

(b) TR2.

(c) TR3.

Figure 6: Word frequencies by team reasoning.

Table 16: Translation of the most frequent used words

Word stem (German)	Full word examples (German)	English translation
alternativ	alternativ (adjective), Alternative (noun)	alternative
beid	beide (pronoun)	both
bess	besser (adverb), verbessern (verb)	better, to improve
dass	dass (conjunction)	that
denk	denken (verb), Denken (noun)	to think, thought
denkt	denkt (verb, 3rd person singular), denken (verb)	he/she/it thinks, to think
egal	egal (adverb)	irrelevant
einfach	einfach (particle, adjective), Einfachheit (noun)	easy, simple, simplicity
entscheid	entscheiden (verb), Entscheidung (noun)	to decide, decision
erst	erst (adverb), erstens (adverb)	at first, first of all
feld	Feld (noun)	field, area ("pie slice")
gibt	gibt (verb, 3rd person singular), geben (verb)	he/she/it gives, to give
gleich	gleich (adverb, preposition, adjective)	same, equal, similar
imm	immer (adverb)	always, throughout
ja	ja (particle, noun)	yes, certainly, indeed
mal	mal (adverb), Mal (noun)	sometime, times
mehr	mehr (adverb, pronoun), Mehr (noun)	more, further
nehm	nehmen (verb)	to take, to have
raut	Raute (noun)	hash mark ("#")
tal	Taler (noun)	taler ("ECU")
team	Team (noun)	group
unt	unten (adjective), unter (adverb)	below, under ("bottom slice")
wahl	Wahl (noun), wählen (verb)	choice, option, to choose
weiss	weiß (adjective)	white

A.3 Results for alternative treatment of empty messages

Table 17 and table 18 illustrate the suggested group decisions with dropped empty message observations. The coordination rate is recalculated as follows:

$$c^* = \sum_{a \in \mathcal{A}} \frac{N_1^*(a) \cdot N_2^*(a)}{N_1^* \cdot N_2^*},$$

where $N_g^*(a)$ denotes the number of players from group $g \in \{1, 2\}$ that propose a strategy $a \in \mathcal{A}$ and sent a message, and N_g^* the total number of subjects in groups $g \in \{1, 2\}$ that sent a message.

Table 17: SD in X-Y Games: dropped empty message observations

| | X-Y Game | | | |
	SL	ASL	AML	ALL
$\pi_1(X, X), \pi_2(X, X)$	5, 5	5, 5.1	5, 6	5, 10
$\pi_1(Y, Y), \pi_2(Y, Y)$	5, 5	5.1, 5	6, 5	10, 5
$N_1(X)$	51	20	26	27
$N_2(X)$	51	32	32	30
$N_1(Y)$	3	34	28	25
$N_2(Y)$	0	26	20	22
Empty messages	37	30	36	38
Coord. rate c (all observations - table 5)	93%	49%	49%	50%
Coord. rate c^* (empty message obs. dropped)	94%	49%	50%	50%

Table 18: SD in Pie Games: dropped empty message observations

| | Pie Game | | | |
	S1	S2	AM2	AM4
$\pi_1(L, L), \pi_2(L, L)$ ($)	5, 5	6, 6	5, 6	6, 7
$\pi_1(R, R), \pi_2(R, R)$ (#)	5, 5	6, 6	6, 5	7, 6
$\pi_1(B, B), \pi_2(B, B)$ (§)	5, 5	5, 5	6, 5	7, 5
$N_1(L)$	17	30	22	21
$N_2(L)$	21	29	17	21
$N_1(R)$	3	7	6	25
$N_2(R)$	4	9	8	21
$N_1(B)$	35	21	21	11
$N_2(B)$	27	14	26	8
Empty messages	35	32	42	35
Coord. rate c (all observations - table 6)	44%	43%	40%	37%
Coord. rate c^* (empty message obs. dropped)	46%	41%	39%	37%

A.4 Differences in level distributions

Table 19 illustrates the p-values of Fisher exact tests for equality of the level bound distributions across the eight games. In the payoff symmetric games, the similarity of the distributions between Pie games (S1, S2) and the X-Y game (SL) stands out. The level bound distributions of payoff asymmetric X-Y games are not significantly different from each other, but significantly different from the distribution of payoff symmetric games. The distribution for AM4 is relatively similar to those of the asymmetric X-Y games, the test are marginally significant.

Table 19: p-values of Fisher exact tests for equality of level distribution.

Game	SL	ASL	AML	ALL	S1	S2	AM2	AM4
SL	–							
ASL	0.000	–						
AML	0.000	0.694	–					
ALL	0.000	0.630	0.312	–				
S1	0.923	0.000	0.000	0.000	–			
S2	0.003	0.000	0.000	0.000	0.001	–		
AM2	0.000	0.000	0.000	0.000	0.000	0.001	–	
AM4	0.000	0.058	0.035	0.181	0.000	0.000	0.024	–

A.5 Choice Probabilities

Table 20: Choice probabilities $P_\rho(a, g, k, \rho)$ in the AM2 game.

| | $\rho = h$ | | | | | | $\rho = l$ | | | | | |
| | $g = 1$ | | | $g = 2$ | | | $g = 1$ | | | $g = 2$ | | |
a	L	R	B	L	R	B	L	R	B	L	R	B
Level-0	0	0.5	0.5	1	0	0	1	0	0	0	0.5	0.5
Level-1	1	0	0	0	0.5	0.5	0	0.5	0.5	1	0	0
Level-2	0	0.5	0.5	1	0	0	1	0	0	0	0.5	0.5
Level-3	1	0	0	0	0.5	0.5	0	0.5	0.5	1	0	0
Level-4	0	0.5	0.5	1	0	0	1	0	0	0	0.5	0.5

Table 21: Choice probabilities $P_\rho(a, g, k, \rho)$ in the AM4 game.

| | $\rho = h$ | | | | | | $\rho = l$ | | | | | |
| | $g = 1$ | | | $g = 2$ | | | $g = 1$ | | | $g = 2$ | | |
a	L	R	B	L	R	B	L	R	B	L	R	B
Level-0	0	0.5	0.5	1	0	0	1	0	0	0	0	1
Level-1	1	0	0	0	0.5	0.5	0	0	1	1	0	0
Level-2	0	0.5	0.5	1	0	0	1	0	0	0	0	1
Level-3	1	0	0	0	0.5	0.5	0	0	1	1	0	0
Level-4	0	0.5	0.5	1	0	0	1	0	0	0	0	1

Table 22: Choice probabilities $P_\sigma(a, k, \sigma)$ in the Pie games.

| | $\sigma = L$ | | | $\sigma = R$ | | | $\sigma = B$ | | |
a	L	R	B	L	R	B	L	R	B
Level-0	1	0	0	0	1	0	0	0	1
Level-1	1	0	0	0	1	0	0	0	1
Level-2	1	0	0	0	1	0	0	0	1
Level-3	1	0	0	0	1	0	0	0	1
Level-4	1	0	0	0	1	0	0	0	1

A.6 Individual Data

The individual raw data in table 23 reflects the main differences in reasoning between games. For example, subjects 26, 49, and 98 individually show the patterns of levels of reasoning that are observed in the aggregate. Furthermore, it can be seen that levels of reasoning are varying strongly across games. Even across very similar games such as ASL, AML, and ALL, stable levels within individuals are the exception. Two factors might facilitate this instability. First, the simplicity of the game makes best-responding relatively easy. More importantly, however, any derived belief can hardly be deemed more reasonable than another due to the cycling of the optimal action. This can be seen in a possible reformulation of the game if it is approached with a payoff salient level-0 belief $\rho = h$. In this case, the choice probabilities in table 3 show that a X-Y game is transformed into a meta-coordination game not on actions X and Y, but rather on even and odd levels as depicted in table 24. Therefore, with small costs and benefits of deliberation, the realized level and resulting decision might at the end be the result of a rather arbitrary process.

Table 23: Level bound, level-0 belief and team reasoning classification for selected subjects by game.

Game	Subject 5 (g = 1)						Subject 6 (g = 1)						Subject 13 (g = 1)						Subject 14 (g = 1)					
	SD	$\underline{k}_i$	$\overline{k}_i$	ρ_i	σ_i	TR	SD	$\underline{k}_i$	$\overline{k}_i$	ρ_i	σ_i	TR	SD	$\underline{k}_i$	$\overline{k}_i$	ρ_i	σ_i	TR	SD	$\underline{k}_i$	$\overline{k}_i$	ρ_i	σ_i	TR
SL	X	1	1	~	X	1	X						X	1	1		X	1	X	0	0	~	~	
ASL	Y	2	2	h			Y	2	2	h			Y	2	2	h			Y	0	0			
AML	X	1	1	h			X	1	1	h			Y	2	2	h			Y	0	0			
ALL	Y	0	2	h			X	1	1	h			Y	2	2	h			X	2	2	h		
S1	B	1	1		B	1	B	0	0				L	1	1		L	1	R	0	0			
S2	B	1	1	~	B		R	1	1	h		2	L	1	1	h	L	2	R	0	0	h		
AM2	B	2	2	h	B		B	1	1		B	1	L	0	0		L		L	1	1	h		3
AM4	R	1	1	h		2	R	1	1	h			L	0	1	h		2	L	1	1	h		3

Game	Subject 26 (g = 2)						Subject 49 (g = 2)						Subject 90 (g = 2)						Subject 98 (g = 2)					
	SD	$\underline{k}_i$	$\overline{k}_i$	ρ_i	σ_i	TR	SD	$\underline{k}_i$	$\overline{k}_i$	ρ_i	σ_i	TR	SD	$\underline{k}_i$	$\overline{k}_i$	ρ_i	σ_i	TR	SD	$\underline{k}_i$	$\overline{k}_i$	ρ_i	σ_i	TR
SL	X	0	0				X	0	0		X	1	X	1	1		X	1	X	1	1		X	1
ASL	Y	2	2	h			X	2	2	h			Y	1	1	h			X	2	2	h		
AML	X	2	2	h			X	2	2	h			X	2	2	h			Y					
ALL	X	2	2	h			Y	1	1	h			Y	1	1	h			X	1	1	h		
S1	B	0	0	~	B	1	L	0	0		L	1	B	0	0		B	1	L	1	1		L	1
S2	L	0	0				L	0	0		L		B	1	1		B	1	L	0	0		L	
AM2	B	1	1	h	~		L	0	0		L		B	1	1		B	1	B	1	1	h	B	
AM4	R	2	2	h		2	L	2	2	h		2	B	1	1		B	1	L	2	2	h		

Game	Subject 134 (g = 2)						Subject 138 (g = 2)					
	SD	$\underline{k}_i$	$\overline{k}_i$	ρ_i	σ_i	TR	SD	$\underline{k}_i$	$\overline{k}_i$	ρ_i	σ_i	TR
SL	X	0	0		X		X	0	0	~	~	
ASL	X	0	0		X		Y	1	1			
AML	X	0	0		X		X	2	2	h		
ALL	X	2	2	h	X	1	Y	1	1	h		
S1	L	0	0	~	L		R	0	0	~	~	
S2	L	0	0		L		L	1	1			
AM2	L	0	0				R	1	1			
AM4	R	0	0		R		L	1	1	h		

Table 24: Meta-coordination game of ALL in terms of levels of reasoning.

		Group 2	
		odd level	even level
Group 1	even level	5, 10	0, 0
	odd level	0, 0	10, 5

A.7 Alternative Estimations

In table 25, we report results of the alternative estimation that is exclusively based on the coinciding classifications between both RAs. The results are qualitatively not different.

Table 25: Estimation results by game (Coinciding classifications).

| | | XY games | | | | | Pie games | | | |
		SL	ASL	AML	ALL		**S1**	**S2**	AM2	AM4
Level-0	l_0	0.65	0.30	0.27	0.24	π_0	0.65	0.45	0.43	0.26
Level-1	l_1	0.35	0.26	0.24	0.33	π_1	0.35	0.45	0.43	0.36
Level-2	l_2	0.00	0.34	0.34	0.36	π_2	0.00	0.05	0.11	0.25
Level-3	l_3	0.00	0.08	0.10	0.04	π_3	0.00	0.03	0.02	0.06
Level-4	l_4	0.00	0.02	0.06	0.03	π_4	0.00	0.03	0.00	0.05
Average	$\hat{k}$	0.35	1.26	1.44	1.28	$\bar{k}$	0.35	0.73	0.73	1.28
Payoff/label sal.	o	0.35	0.86	0.89	0.91	o	0.00	0.47	0.53	0.75
Payoff sal.	r_h	–	1.00	1.00	1.00	r_h	–	1.00	1.00	1.00
Label sal.	s_X	0.99	0.59	0.73	0.67	s_B	0.56	0.37	0.54	0.44
	s_Y	0.01	0.41	0.27	0.33	s_L	0.37	0.42	0.30	0.37
						s_R	0.06	0.20	0.15	0.19
	N	99	91	79	81	N	102	86	83	85

Furthermore, table 26 reports results of an alternative specification with the full sample, including observations whose messages have neither a level-k nor a level-0 classification. The level estimation is more noisy, but the main results of the paper are still reflected.

Table 26: Estimation results by game (Full Sample).

| | | XY games | | | | | Pie games | | | |
		SL	ASL	AML	ALL		**S1**	**S2**	AM2	AM4
Level-0	l_0	0.37	0.23	0.21	0.21	l_0	0.38	0.34	0.27	0.23
Level-1	l_1	0.26	0.25	0.24	0.28	l_1	0.26	0.33	0.29	0.30
Level-2	l_2	0.12	0.28	0.27	0.27	l_2	0.12	0.12	0.17	0.23
Level-3	l_3	0.12	0.13	0.16	0.13	l_3	0.12	0.11	0.13	0.13
Level-4	l_4	0.12	0.11	0.12	0.11	l_4	0.12	0.10	0.13	0.11
Average	$\hat{k}$	1.35	1.64	1.73	1.66	$\hat{k}$	1.33	1.30	1.55	1.59
Payoff/label sal.	o	0.02	0.61	0.62	0.61	o	0.00	0.78	0.39	0.56
Payoff sal.	r_h	–	1.00	1.00	1.00	r_h	–	1.00	1.00	1.00
Label sal.	s_X	0.97	0.67	0.57	0.53	s_B	0.50	0.84	0.54	0.26
	s_Y	0.03	0.33	0.43	0.47	s_L	0.41	0.00	0.37	0.31
						s_R	0.09	0.16	0.09	0.42
	N	142	142	142	142	N	142	142	142	142

Finally, table 27 reports results of the specification in which observations with both label and payoff salience are entering according to their label salience. Section 4 reports the estimation in which they entered according to their payoff salience.

Table 27: Estimation results by game.

		XY games					Pie games			
		SL	ASL	AML	ALL		**S1**	**S2**	AM2	AM4
Level-0	l_0	0.64	0.26	0.22	0.27	l_0	0.64	0.48	0.40	0.27
Level-1	l_1	0.36	0.30	0.29	0.34	l_1	0.36	0.48	0.45	0.41
Level-2	l_2	0.00	0.36	0.34	0.34	l_2	0.00	0.03	0.13	0.27
Level-3	l_3	0.00	0.06	0.11	0.05	l_3	0.00	0.01	0.02	0.04
Level-4	l_4	0.00	0.01	0.04	0.00	l_4	0.00	0.00	0.00	0.01
Average	$\hat{k}$	0.36	1.27	1.44	1.16	$\hat{k}$	0.36	0.56	0.76	1.12
Payoff/label sal.	o	0.00	0.78	0.80	0.85	o	0.00	0.51	0.37	0.67
Payoff sal.	r_h	–	1.00	1.00	1.00	r_h	–	1.00	1.00	1.00
Label sal.	s_X	0.98	0.60	0.71	0.81	s_B	0.57	0.70	0.58	0.47
	s_Y	0.02	0.40	0.29	0.19	s_L	0.38	0.13	0.24	0.26
						s_R	0.06	0.18	0.18	0.27
	N	105	111	106	104	N	107	110	100	107

Notes: 38 messages with both payoff and label salience entering estimation under label salience.

B Experiment Materials

B.1 Experiment Instructions in English translation

Welcome

Introduction

Welcome to the experiment. The experiment is funded by the University of Mannheim and the "German Association for Experimental Economic Research" ("Gesellschaft für experimentelle Wirtschaftsforschung"). Please follow the instructions carefully. You may earn a considerable amount of money. Your decisions and the decisions of the other participants determine the amount of money you will receive. 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.

The experiment consists of a test round, four comprehension questions and the main experiment, which consists of eight rounds. 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. The way you interact as a team to take decisions will be the same throughout all eight rounds. The experiment is carried out anonymously. Neither your, nor any other participants' identity will be disclosed. You and your team partner will earn the exact same amount of money.

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**.

Test period

A participant in this room is now randomly chosen to be your teammate. The test period has two rounds. Since this is only a test, your earnings will not depend on anything that happens now. In both test rounds you will need to answer a question related to the date of a specific historical event. The winning team will be the one whose answer is closest to the correct date of the historical event.

You can propose to your partner a *Suggested Decision and send him one justifying text message*. 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. Your final decision is not limited to your or your team partner's suggested decision, you can choose your answer freely. The computer will choose randomly which one of your two final decisions counts as your *team's action*.

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 only 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.

When you are ready, please click the "Ready" button to start the **Test Period**.

Experimental Procedure

A new participant in this room is now randomly chosen to be your teammate throughout the eight experimental rounds. You can earn a considerable amount of money in each round. We will inform you of the amount you earned in each round at the end of the experiment. You will receive *0,40 Euro* for each Taler you earned in the experiment. Your task is the following:

You and your teammate will randomly be assigned to play against another team throughout all eight experimental rounds. It is in your best interest to *coordinate your team's action with the action of the other team*. Your team can only earn a specific amount of Taler, if your team's action is identical with the other team's actions. Both teams coordinated their actions successfully and earn a specific amount of Taler in case your team's action is identical with the

42

other team's actions. ***If your team's action is different from the other team's action, no team will receive any payoff.*** You will go through the eight experimental rounds in sequence.

In each round, you can propose to your partner a ***Suggested Decision and send him one and only one text Message***. After you finish entering your suggested decision and your message, you will ***directly*** proceed to the next round.

As soon as you entered your suggested decisions and your messages in each of the eight experimental rounds, you will receive your ***partner's suggested decisions and messages*** in the same order. Likewise, your team partner will receive your suggested decisions and messages. Both of you will then make your ***final decision***. The computer will choose randomly which one of your two final decisions counts as your ***team's action***.

Please note again that your team can only earn a specific amount of Taler if your team's action is identical with the other team's actions. ***We inform you of the amount you earned as soon as you passed through all eight experimental rounds.***

In summary: You and your teammate will randomly be assigned to play against another team throughout all eight experimental rounds. You will pass through the eight experimental rounds sequentially. In each round, you can propose to your partner a Suggested Decision and send him one and only one text Message. As soon as you entered your suggested decisions and your messages in each of the eight experimental rounds, you will receive your partner's suggested decisions and messages. Both of you will then make your final decision in all eight rounds. The computer will choose randomly which one of your two final decisions counts as your team's action. Your team can only earn a specific amount of Taler, if your team's action is identical with the other team's actions. You will receive ***0,40 Euro*** for each Taler you earned in the experiment.

If you have any questions at this point, please raise your hand. Before the experiment begins, we ask you to answer ***four comprehension questions.*** Each question presents you some hypothetical team actions. It is your task to determine the winning team. When you are ready, please click the "Ready" button to start the **Comprehension questions**.

The experiment begins!

↓ Round	1	
↓	2	
↓	3	
↓	4	You will send a **suggested**
↓	5	**decision** and a **justifying**
↓	6	**message** to your teammate.
↓	7	
↓	8	
↓ Back to round	1	
↓	2	
↓	3	You will receive your partner's
↓	4	suggested decisions and messages.
↓	5	You will then make your **final decisions**
↓	6	in all eight rounds individually.
↓	7	
↓	8	
↓ Back to round	1	
↓	2	
↓	3	In each round, the computer
↓	4	will choose randomly which one
↓	5	of your two final decisions
↓	6	counts as your **team's action**.
↓	7	
↓	8	

B.2   Experiment Instructions in German original version

Willkommen

Einleitung

Ich begrüße Sie zum heutigen Experiment. Das Experiment ist finanziert durch die Universität Mannheim und die Gesellschaft für experimentelle Wirtschaftsforschung. Bitte befolgen Sie die Anweisungen sorgfältig. Sie haben die Möglichkeit, einen variablen Geldbetrag zu verdienen. Ihre Entscheidungen und die Entscheidungen anderer bestimmen diesen Betrag. Sie werden im Detail unterrichtet, wie dieser Betrag von Ihren Entscheidungen und den Entscheidungen anderer abhängt. Was Sie verdienen wird Ihnen nach der Sitzung privat und in bar ausgezahlt. Es ist wichtig, dass Sie während der Sitzung nicht reden und nicht auf andere Bildschirme schauen. Wenn Sie Fragen haben oder Hilfe brauchen, heben Sie die Hand und jemand wird zu Ihnen kommen. Wenn Sie sprechen, laut werden, etc. werden Sie aufgefordert, das Experiment zu verlassen.

Das Experiment besteht aus einer Testrunde, vier Verständnisfragen und dem eigentlichen Experiment, das sich in 8 Runden gliedert. Die Entscheidungen in diesem Experiment werden *im Team* getroffen. Sie formen dementsprechend mit einem zufällig ausgewählten Versuchsteilnehmer in diesem Raum ein Team, das als eine Einheit agiert. Die Interaktion im Team ist im gesamten Experiment die gleiche. Das Experiment wird anonym durchgeführt, Identitäten von Ihnen, Ihren Teampartnern oder anderen Mitspielern werden nicht preisgegeben. Ihr Teampartner und Sie werden stets den gleichen Geldbetrag verdienen.

Wie wird nun die *Teamentscheidung* getroffen? Ihr Teampartner und Sie werden beide eine individuelle *endgültige Entscheidung* treffen. Der Computer sucht zufällig eine der beiden Entscheidungen als *Teamentscheidung* aus. Die Wahrscheinlichkeit, dass die endgültige Entscheidung Ihres Teampartners gewählt wird, ist gleich der Wahrscheinlichkeit, dass Ihre endgültige Entscheidung gewählt wird, d.h. die Chancen sind 50:50.

Sie haben allerdings die Möglichkeit, die Entscheidung Ihres Partners auf die folgende Weise zu beeinflussen: Bevor Sie Ihre endgültige Entscheidung treffen, können Sie Ihrem Partner einen *Entscheidungsvorschlag machen und ihm genau eine Nachricht schicken*. Beachten Sie, dass diese Nachricht die einzige Möglichkeit ist Ihren Partner von den Gründen hinter Ihrer Entscheidung zu überzeugen. Nutzen Sie deshalb die Nachricht um Ihren Entscheidungsvorschlag zu erklären. Nachdem Sie Ihren Entscheidungsvorschlag und Ihre Nachricht eingegeben haben, werden diese Ihrem Partner gezeigt, der dann die endgültige Entscheidung treffen wird. Gleichzeitig werden Sie die Nachricht und den *Entscheidungsvorschlag Ihres Partners empfan-*

gen und können dann Ihre endgültige Entscheidung treffen. Wie beschrieben wird der Computer dann eine der beiden endgültigen Entscheidungen zufällig als Teamentscheidung auswählen.

Wenn Sie jetzt Fragen haben, heben Sie bitte Ihre Hand. Um das Kommunikationssystem kennenzulernen, werden Sie nun eine **Testrunde** absolvieren.

Testrunde

Ein Teilnehmer in diesem Raum wird nun zufällig als Ihr Partner ausgesucht. Die Testrunde besteht aus zwei Perioden, in denen jeweils eine Frage zu beantworten ist. Da dies lediglich ein Test ist, können Sie nun kein Geld gewinnen. In beiden Runden werden Sie eine Frage zur Jahreszahl eines historischen Ereignisses beantworten. Das Team gewinnt, das am nächsten an der korrekten Jahreszahl liegt.

Wie beschrieben werden Sie die Möglichkeit haben, einen *Entscheidungsvorschlag* mit der vorgeschlagenen Jahreszahl zu machen und eine *erklärende Nachricht* zu schicken. Nachdem Sie den Entscheidungsvorschlag und die Nachricht Ihres Partners gelesen haben, werden Sie Ihre *endgültige Entscheidung* eingeben. Die Wahl der endgültigen Entscheidung ist nicht auf die zwei Entscheidungsvorschläge beschränkt, sie können Ihre endgültige Entscheidung frei wählen. Entweder Ihre oder die endgültige Entscheidung Ihres Partners wird zufällig als *Teamentscheidung* ausgewählt.

Das Kommunikationssystem erlaubt Nachrichten beliebiger Länge. Sie müssen die Nachricht allerdings Zeile für Zeile eingeben, da das Eingabefeld nur einzeilig ist. Innerhalb dieser Zeile können Sie mit der normalen *Backspace* Taste Ihres Keyboards Eingaben löschen. Sie fügen den Text in der Eingabezeile zu Ihrer Nachricht hinzu, indem Sie *Enter* drücken. Bitte beachten Sie, dass ausschließlich Eingaben, die Sie der Nachricht hinzugefügt haben, von Ihrem Partner gesehen werden. Die *Eingaben in der blauen Eingabezeile werden nicht übermittelt*. Hinzugefügte Nachrichten können Sie jederzeit löschen indem Sie den *Eingabe löschen* Button auf dem Bildschirm klicken. Die Anzahl der Zeilen ist nicht limitiert. Sie können dementsprechend Nachrichten jeglicher Länge senden. Sie senden die Nachricht endgültig, wenn Sie den *Nachricht senden* Button klicken.

Wenn Sie bereit sind, klicken Sie bitte *Bereit*, um die **Testrunde** zu starten.

Beschreibung des Experiments

Sie werden nun zufällig mit einem neuen Teampartner gematcht. Im Experiment durchlaufen Sie *acht Runden*, die Sie durchgängig mit dem gleichen Partner im Team gegen das gleiche Team spielen werden. In jedem Spiel können Sie eine bestimmte Anzahl an Taler gewinnen.

Sie werden am Ende des Experiments über Ihren Erfolg informiert. Für jeden gewonnen Taler erhalten Sie *0,40 Euro*. Ihre Aufgabe ist die Folgende:

Ihr Team spielt acht Runden gegen eines der anderen Teams. Es ist in Ihrem Interesse *die Handlung zwischen den Teams zu koordinieren*. Beide Teams müssen die gleiche Teamentscheidung treffen, um eine bestimme Anzahl Taler zu gewinnen. Ist die Entscheidung beider Teams identisch, so haben Sie ihre Handlung erfolgreich koordiniert und erhalten eine bestimmte Anzahl Taler. *Treffen die Teams jedoch voneinander verschiedene Entscheidungen, erhält keins der beiden Teams Taler.* Sie werden die acht Runden der Reihe nach durchlaufen.

Wie zuvor können Sie in jeder Runde einen *Entscheidungsvorschlag und eine erklärende Nachricht* zu Ihrem Teampartner senden. Nachdem Sie Ihren Entscheidungsvorschlag und eine Nachricht zu Ihrem Teampartner gesendet haben, gelangen Sie *direkt* zur jeweils nächsten Runde.

Nachdem Sie die acht Runden durchlaufen haben, empfangen Sie für jede der acht Runden in der gleichen Reihenfolge den *Entscheidungsvorschlag und die Nachricht Ihres Partners*. Sie werden nun Ihre *endgültige Entscheidung* treffen. Entweder Ihre oder Ihres Partners endgültige Entscheidung wird durch den Computer zufällig als *Teamentscheidung* ausgewählt.

Bitte beachten Sie nochmals, dass Ihr Team nur dann eine bestimmte Anzahl Taler erhält, wenn Ihr Team die gleiche Entscheidung wie das andere Team trifft. *Sobald Sie in allen acht Runden Ihre endgültige Entscheidung getroffen haben, werden Sie über den Erfolg Ihres Teams informiert*.

Noch einmal zusammengefasst: Sie spielen mit dem gleichen Partner im Team acht Runden gegen ein anderes Team. Sie werden die Spiele der Reihe nach durchlaufen. In jeder Runde können Sie einen Entscheidungsvorschlag und eine erklärende Nachricht zu Ihrem Teampartner senden. Nachdem Sie die acht Runden durchlaufen haben, empfangen Sie die Entscheidungsvorschläge und die Nachrichten Ihres Partners. Nun treffen Sie für alle acht Runden Ihre endgültige Entscheidung. Entweder Ihre oder Ihres Partners endgültige Entscheidung wird durch den Computer zufällig als Teamentscheidung ausgewählt. Sie erhalten nur dann eine bestimmte Anzahl Taler, wenn Ihr Team die gleiche Entscheidung trifft wie das andere Team. Für jeden gewonnen Taler erhalten Sie 0,40 Euro.

Wenn Sie jetzt Fragen haben, heben Sie bitte Ihre Hand. Sie werden vor dem eigentlichen Experiment *vier Verständnisfragen* beantworten. In den Verständnisfragen werden Ihnen die Teamentscheidungen in verschiedenen Situationen gezeigt. Ihre Aufgabe ist das Bestimmen der Gewinnerteams. Wenn Sie bereit sind, klicken Sie bitte *Bereit*, um die **Verständnisfragen** zu starten.

↓		1	
↓		2	
↓		3	
↓		4	Sie senden einen **Entscheidungs-**
↓	Runde	5	**vorschlag** und eine **erklärende**
↓		6	**Nachricht** zu Ihrem Teampartner.
↓		7	
↓		8	
↓		1	
↓		2	
↓		3	Sie empfangen nun den Entscheidungs-
↓		4	vorschlag und die Nachricht Ihres
↓	Zurück zu Runde	5	Teampartners. Anschließend treffen
↓		6	Sie Ihre **endgültige Entscheidung.**
↓		7	
↓		8	
↓		1	
↓		2	
↓		3	Entweder Ihre oder die endültige
↓		4	Entscheidung Ihres Teampartners
↓	Zurück zu Runde	5	wird zufällig vom Computer als
↓		6	**Teamentscheidung** ausgewählt.
↓		7	
↓		8	

Das Experiment beginnt!

Wenn Sie bereit sind, klicken Sie bitte *Bereit*, um das **Experiment** zu starten.

B.3   Screenshots

Sie beginnen nun Runde 3. Ihr Teampartner und Sie bilden Team 2.
Betrachten Sie das folgende Spiel und die Anzahl Ihrer Taler im Falle einer Koordination mit Team 1.
Bitte schlagen Sie Ihrem Teampartner eine der beiden nachstehen Alternativen als Teamentscheidung vor.

Alternative **X**   [5 Taler für Team 1 und 6 Taler für Team 2]

Alternative **Y**   [6 Taler für Team 1 und 5 Taler für Team 2]

1. Bitte klicken Sie um Ihrem Teampartner eine *Entscheidung* vorzuschlagen.

| X | | Y |
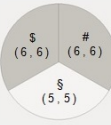
2. Hier können Sie Ihre *Nachricht* eingeben um Ihrem Partner den *Entscheidungsvorschlag* zu erklären.
Drücken Sie die <Enter>-Taste um die Eingabe zur Nachricht hinzuzufügen.
Klicken Sie den "Eingabe löschen"-Button um die letzte hinzugefügte Eingabe von der Nachricht zu löschen.

Hier werden Nachrichten eingegeben [...] [...]

Eingabe löschen

Text in der blauen Eingabezeile ist nicht Teil der Nachricht und wird nicht gesendet. Platzieren Sie den Cursor in der
Eingabezeile und drücken Sie die <Enter>-Taste um die Eingabe zur Nachricht hinzuzufügen!

Sie beginnen nun Runde 2. Ihr Teampartner und Sie bilden Team 1.
Bitte schlagen Sie Ihrem Teampartner eines der drei abgebildeten Kreissegmente als Teamentscheidung vor.
Die erstgenannte Zahl zeigt die Anzahl der Taler für Team 1, die zweitgenannte Zahl die Anzahl der Taler für Team 2 **bei erfolgreicher Koordination.**

$
\begin{array}{cc}
\$ & \# \\
(6,6) & (6,6) \\
& \S \\
& (5,5)
\end{array}
$

1. Bitte klicken Sie um Ihrem Teampartner eine *Entscheidung* vorzuschlagen.

| $ | | # |
| | § | |

2. Hier können Sie Ihre *Nachricht* eingeben um Ihrem Partner den *Entscheidungsvorschlag* zu erklären.
Drücken Sie die <Enter>-Taste um die Eingabe zur Nachricht hinzuzufügen.
Klicken Sie den "Eingabe löschen"-Button um die letzte hinzugefügte Eingabe von der Nachricht zu löschen.

Hier werden Nachrichten eingegeben [...]

Eingabe löschen

Text in der blauen Eingabezeile ist nicht Teil der Nachricht und wird nicht gesendet. Platzieren Sie den Cursor in der
Eingabezeile und drücken Sie die <Enter>-Taste um die Eingabe zur Nachricht hinzuzufügen!

49

B.4   Classification Instructions

Welcome!

Thank you for participating in this experiment. In this section you find instructions as to how this experiment works. To take part in the experiment, we assume that you are familiar with the level-k model as it has been introduced by Nagel (1995) and also with the concept of team reasoning as it has been introduced by Schelling (1960). In the experiment, subjects play pure coordination games with symmetric and asymmetric payoffs. We assume that you are familiar with the concept of coordination games as they have been carried out by Crawford, Gneezy and Rottenstreich (2008).

However, in order to clarify potential questions of terminology, we reproduce the main features of the level-k model and the concept of team reasoning. In addition we provide detailed experimental instructions, which explain the game and also give you a short introduction to coordination games. Please read all information carefully in order to know how the original experiment proceeded.

Experimental Setting

Introduction

This section describes the main features of the experiment. Subjects are randomly assigned into teams of two players. For a given strategic situation, each player makes suggestions for the team action at two points in time. First, the so-called "suggested decision" and a justifying **written message** are exchanged between the team partners simultaneously. After this, the "final decision" is taken individually by each team player. The computer chooses randomly one of the two final decisions to obtain the "team's action."

All teams play a series of eight coordination games. Coordination games are characterised by situations in which all parties can realize mutual gains, but only by making mutually consistent decisions. Each team is randomly matched with another team. If a matched pair of teams both decide on identical team actions, they coordinate their behavior successfully and are rewarded with a payoff.

However, if both teams choose different team actions, they fail to coordinate their behavior and do not receive any payoff. Thus both teams are motivated solely to coordinate their strategies in order to obtain an outcome that is best for them. The following example illustrates a random coordination game in which each team decides on one strategy X, Y or Z simultaneously. Only if both teams make mutually consistent decisions they receive a payoff of 2 units each.

		Team 2		
		X	Y	Z
	X	(2,2)	(0,0)	(0,0)
Team 1	Y	(0,0)	(2,2)	(0,0)
	Z	(0,0)	(0,0)	(2,2)

The payoff is represented through an experimental currency unit ("Taler"). One Taler is worth 0,40 Euro. In a **symmetric coordination** game each team is rewarded the *same* payoff if they coordinate their behavior successfully. In **asymmetric coordination** games players usually disagree on which action they prefer to coordinate. There may be one outcome where *one team disproportionately* benefits in comparison to the other team.

"X-Y Coordination Games"

All subjects face a series of eight coordination games composed of **four "X-Y Games"** and **four "Pie Games"**. We reproduce the main features and attributes of those games in the following. **"X-Y Games"** are characterised by a binary choice option "X" or "Y". The assignment of payoffs for successful coordination is indicated in brackets. Example:

X   [6 Taler for Team 1 and 5 Taler for Team 2]
Y   [5 Taler for Team 1 and 6 Taler for Team 2]

If a matched pair of teams both decide on the identical team action "X", team one receives *6 Taler* and team two receives *5 Taler*. If both teams chose "Y", the assignment of payoffs would be reversed. If both teams chose decisions with different labels "X" and "Y", neither team receives any payoff. The payoff differences vary within the four "X-Y treatments".

"Pie Coordination Games"

"Pie Games" are characterised by a visual representation of different choice options as indicated in the following figure. Each team simultaneously selects one of the three "pie slices". Each slice is labeled with an abstract decision label §, **$ or #**. The assignment of payoffs for successful coordination is indicated in brackets within the three slices. The first number represents the quantity of Taler for team one, the second number the quantity of Taler for team two.
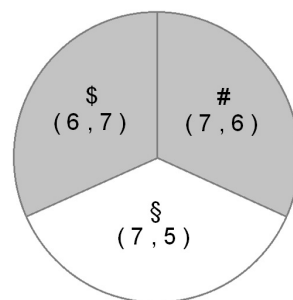
If a matched pair of teams both decide on the identical team action "#", team one receives *7 Taler* and team two receives *6 Taler*. If both teams chose decisions with different labels §, $

51

or #, neither team receives any payoff. The payoff differences alternate within the four "Pie treatments".

Note that the "X-Y Game" and the "Pie Game" might both contain one alternative that is visually distinctive from another alternative. For instance, the unshaded buttom slice is visually distinctive from the two upper slices ($ and #) that are shaded in a light grey color. We refer to a visually distinctive alternative as **label salient**. Moreover an alternative might be **payoff salient** in a way that it is distinctive with respect to its payoff structure. The **concept of label and payoff salience is important** for the classification process.

Treatment Overview

We conducted six sessions in Mannheim and three sessions Heidelberg. All sessions consist of the same eight treatments (four "X-Y games" and four "Pie games"), however the sequence of treatments in Mannheim is different from the sequence of treatments in Heidelberg. The following two tables provide a brief overview over the sessions conducted in Mannheim (session 1-3, session 7-9 [rounds 7 and 8 moved to the beginning]) and the sessions conducted in Heidelberg (session 4-6). The payoff for successful coordination is indicated in brackets. The first number represents the quantity of Taler for team one, the second number represents the quantity of Taler for team two, if both teams coordinate their behavior.

Sessions 1-3 Mannheim:

Round 1	X	(5 ; 5.1)	**Round 2**	$	(5 ; 6)	**Round 3**	X	(5 ; 6)	**Round 4**	$	(6 ; 7)
	Y	(5.1 ; 5)		#	(6 ; 5)		Y	(6 ; 5)		#	(7 ; 6)
				§	(6 ; 5)					§	(7 ; 5)
Round 5	X	(5 ; 10)	**Round 6**	$	(6 ; 6)	**Round 7**	X	(5 ; 5)	**Round 8**	$	(5 ; 5)
	Y	(10 ; 5)		#	(6 ; 6)		Y	(5 ; 5)		#	(5 ; 5)
				§	(5 ; 5)					§	(5 ; 5)

Sessions 4-6 Heidelberg:

Round 1	X	(5 ; 10)	**Round 2**	$	(6 ; 7)	**Round 3**	X	(5 ; 6)	**Round 4**	$	(6 ; 6)
	Y	(10 ; 5)		#	(7 ; 6)		Y	(6 ; 5)		#	(6 ; 6)
				§	(7 ; 5)					§	(5 ; 5)
Round 5	X	(5 ; 5,10)	**Round 6**	$	(5 ; 6)	**Round 7**	X	(5 ; 5)	**Round 8**	$	(5 ; 5)
	Y	(5,10 ; 5)		#	(6 ; 5)		Y	(5 ; 5)		#	(5 ; 5)
				§	(6 ; 5)					§	(5 ; 5)

Sessions 7-9 Mannheim:

Round 1	X	(5 ; 5)	**Round 2**	$	(5 ; 5)	**Round 3**	X	(5 ; 5.1)	**Round 4**	$	(5 ; 6)
	Y	(5 ; 5)		#	(5 ; 5)		Y	(5.1 ; 5)		#	(6 ; 5)
				§	(5 ; 5)					§	(6 ; 5)
Round 5	X	(5 ; 6)	**Round 6**	$	(6 ; 7)	**Round 7**	X	(5 ; 10)	**Round 8**	$	(6 ; 6)
	Y	(6 ; 5)		#	(7 ; 6)		Y	(10 ; 5)		#	(6 ; 6)
				§	(7 ; 5)					§	(5 ; 5)

Classification Process

Remember: Each player makes suggestions for the team action at two points in time. First, the so-called "suggested decision" and a justifying **written message** are exchanged between the team partners simultaneously. After this, the "final decision" is taken individually by each team player. The computer chooses randomly one of the two final decisions to obtain the "team's action." **Your task is to classify the written messages into different categories.** In the following we will describe the classification process for the analysis of the experiment. Please read the classification instructions carefully.

Level k Model

Notation of the level k model

It is assumed that you are familiar with the level-k model as it has been introduced by Nagel (1995) or represented by Camerer (2004). The model here is extended to incorporate salience in the level-0 belief according to Bacharach and Stahl (2000). In order to clarify potential questions of terminology and introduce the main features of the model we quickly reproduce the main features of the model in the terminology used in this document. The level-k model of bounded rationality assumes that players only think through a certain number (k) of best responses. The model has four main ingredients:

Population distribution: This distribution reflects the proportion of types with a certain level $k \in N_0 = \{0, 1, 2, 3, 4, 5, \ldots\}$.

Level-0 distribution: By definition, a level-0 player does not best respond. Hence, his actions are random to the game and distributed randomly over the action space. In our case, the action space is $\mathcal{A} = \{\{X\}, \{Y\}\}$ or $\mathcal{A} = \{\{§\}, \{\$\}, \{\#\}\}$. The model incorporates salience by assuming higher probabilities in the level-0 distribution for actions that are visually distinctive (salient). An action might be salient in terms of payoffs and in terms of labels. In the "X-Y" treatments, the level-0 distribution would not assign a uniform probability of 0.5 to each possible action, but $p > 0.5$ to the salient one and $q_i < p$ for the remaining actions. In the

"Pie" treatments, the level-0 distribution would not assign a uniform probability of $1/3$ to each possible action, but $p > 1/3$ to the salient one and $q_i < p$ for the remaining actions.

Level-0 belief: In the model, the best responses of players with $k > 0$ are anchored in what they believe the level-0 players play. Their level-0 belief might not be consistent with the level-0 distribution. For best responding, all that matters is the expected payoff from choosing an action from the action space $\mathcal{A} = \{\{X\}, \{Y\}\}$ or $\mathcal{A} = \{\{\S\}, \{\$\}, \{\#\}\}$. A subject would therefore decide on a particular action, when the probability is highest, that the other team chooses the same action.

Population belief: Players do not expect other players to be of the same or a higher level of reasoning. For a level-k player, the population belief is therefore defined on the set of levels strictly below k. It follows that level-0 players have no defined belief, level-1 players have a trivial belief with full probability mass on $\{0\}$, level-2 players have a well defined belief on $\{\{0\}, \{1\}\}$. From level 3 higher order beliefs are relevant as level-3 players have to form a belief about level-2's beliefs.

Characterisation of the different levels

Level 0 The player does not exhibit any strategic reasoning whatsoever. Different versions of this might be randomly chosen or purely guessed actions, misunderstanding of the game structure or other non-strategic 'reasons' for picking a location, e.g. by taste or salience. It is important that no best-responding to the other's play occurs. There could be considerations of what others might play, but without best responding to it. Examples: "Well, it's a pure guess", "There are no arguments. Simply choose any."

Level 1 This player best responds to some belief about the other teams' action. However, he does not realise that others will be strategic as well. Example: "They are probably picking X, so we do as well", "The other team would naturally go for the visual distinctive buttom slice, no?"

Level 2 This player not only shows the basic strategic consideration of playing best response (matching/mismatching), but also realises that other players best respond as well according to the belief they entertain. A level-2 player clearly contemplates how the other player might best respond to his frame. The player plays a best response to this hypothesised consideration. Example: "The other team may think we are most attracted to the alternative # with the highest payoff. In order to coordinate our behavior we should also choose the # slice."

Level 3 This player realises that others could be level-2 and reacts by best responding to the associated expected play. Put differently, he realises that others realise that others best respond

to their initial belief. Therefore, a level-3 player clearly states that his opponent expects that he (the level-3 player at question) best-responds to a certain belief.

Level 4, 5, ... The process goes on in a similar fashion. A level k player realises that other subjects could be level $k - 1$ and reacts by best responding to the associated expected play.

Category 1: Lower and upper bound on the level of reasoning

Your aim

is to classify the written messages into the underlying level $k \in N_0 = \{0, 1, 2, 3, 4, 5, \ldots\}$ of reasoning. For a given statement it might not be possible to exactly determine the underlying level of reasoning. To extract as much information as possible, we ask you to indicate a lower and an upper bound on the level of reasoning.

For the lower bound on the level of reasoning, you should ask yourself: "What is the minimum level of reasoning that this statement clearly exhibits?" Once noted, you should be able to say to yourself: "It seems impossible that the players' level of reasoning is below this number!" Here we ask you to be very cautious with the classification, not giving away high levels easily.

The upper bounds should give the maximum level of reasoning that could be interpreted into the statement. Therefore, you should ask yourself: "What is the highest level of reasoning that can be underlying this statement?" Once noted, you should be able to say: "Although maybe not clearly communicated, this statement could be an expression of this level. If the player reasoned higher than this number, this was not expressed in the statement!" For both lower and upper bound, please refer to the characterisation of the different levels.

There are two necessary conditions for a player to exhibit a level greater than 0. First, the player has to be responsive to the salience of the games' framing. Secondly, the player has to be strategic in best-responding to his level-0 belief, which is shaped by label or payoff salience. If he did not react to salience, he would have no reason to chose one over the other object, resulting in random level 0 play.

For this category, the excel-sheet for the classification will feature a drop-down menu where you can choose upper and lower bounds between 0 and 5. If no inference can be made since nothing or nothing to the point is written, you can choose not applicable (n/a).

Category 2: Level-0 belief

Your aim

is to indicate the underlying level-0 belief that is connected with the lowest possible level of reasoning. **If level reasoning is observed in the statement, there has to be a starting point in the argument which states an attraction or aversion to one alternative.** This is then not derived by strategic reasons, but is an intuitive reaction to the framing of the coordination game.

Otherwise, level reasoning would not occur. Please indicate the underlying level-0 belief that is connected with the lowest possible level of reasoning. Note that the level-0 belief of a person reasoning on an odd level, i.e. level 1, 3, 5, etc. is always with respect to how a player of the opposite side intuitively reacts to the framing. The belief of a person reasoning on an even level, i.e. level 2, 4 etc. is always with respect to what the opposite type believes about the own type's intuitive reaction.

There are two kinds of framing in these games. On the one hand, subjects might react to the **framing of the coordination game (label salience)**. Imagine a subject that you classify to be level-1. It might communicate that the other team is most attracted to the visual distinctive white buttom slice $\S$ and therefore proposes $\S$ as team decision. A subject that you classify to be level-2 might indicate that the other team believes that one's own team is more likely to choose "X", because this alternative is mentioned first on the screen. To reflect a level-0 belief of an attraction to X or Y, or to #, $\S$, or \$, the excel-sheet features a drop-down menu that allows to indicate such a preference or an indifference. If such a preference or indifference over labels is not indicated, or if the subjects' level of attractiveness cannot be distinguished or is not expressed clearly within the message, please indicate that the level-0 belief from the message does not exhibit any label salience.

On the other hand, subjects might **respond to the payoffs (payoff salience)**. For example, consider a subject that you classify to be level-1. It might communicate that the other team is most likely to choose alternative X as it offers the highest payoff to this very team. Or, a subject that you classify to be level-2 might indicate that the other team remains of the conviction that one's own team is not attracted to the action that gives one's own team high payoffs. To reflect the exhibited level-0 beliefs you can indicate in the excel-sheet whether the team that the level-0 belief is formed about is believed to be attracted to a) the action that yields – under coordination – a higher payoff for this team, to b) the action that yields – under coordination – a higher payoff to the other team or c) is indifferent. If no such preference or indifference over salient payoff actions is indicated, please indicate that the level-0 belief from the message does not exhibit any payoff structure.

Please note that payoff and label salience are not mutually exclusive, please indicate both if both is expressed in the message. Finally, for players whose lower bound is 0, the level-0 belief classification can be used to indicate whether a level-0 player states for his action a preference with respect to label or payoff salience.

Team Reasoning

Basic concept of team reasoning

It is assumed that you are familiar with the concept of team reasoning as it has been introduced by Schelling (1960). In order to clarify potential questions of terminology we reproduce the main features of the concept.

The central idea behind the concept of team reasoning in coordination games is to find some criterion that distinguishes one particular label from others. In team reasoning, players begin by asking themselves if there is a decision rule that would be better for both teams than individualistic rules, if both team followed that rule. When someone is playing a coordination game, he/she will look for a rule of selection which tends to produce successful coordination if this rule is followed by both teams. A rule of selection - and by extension, the label or strategy that is identified - "suggests itself" or seems obvious or natural to people who are looking for ways to solve coordination games.

Category 3: Team reasoning

Your aim is to indicate any approach of team reasoning that can be obtained from the written messages. There are several examples that illustrate the concept of team reasoning with respect to the treatments in the experiment. Three possibilities will be given in the drop-down menu of the excel-sheet. First, subjects might ignore payoffs completely and choose according to visually distinctive attributes, with the underlying expectation that the other team thinks and behaves analogously. Individuals might ask if there is any rule that reliably breaks the asymmetry in the payoffs. In these cases, please choose "Labels as coordination device". Second, a similar approach might lead to coordination on actions that feature a distinctive and unique payoff pair, as in round 2 and round 6 of the Mannheim sessions. In these cases, please choose "Payoffs as coordination device". Third, a different approach of team reasoning might be to minimize the risk of miscoordination between two teams and to apply pareto-dominance arguments to the payoff structure. If an action is seen as irrelevant since both teams could earn more from another action, please indicate the existence of "Pareto dominance arguments".

As described above, the concept of team reasoning can be expressed in a number of different ways. Please indicate any other decision rule that fits the concept of team reasoning by noting "other" in the excel sheet and by writing a short comment about the nature of the decision rule. If the subjects' message does not clearly express a rule, label or strategy of team reasoning, please do not indicate any rule.

Classification Summary

In coordination games both teams are motivated solely to coordinate their strategies in order to obtain an outcome that is best for them. For a given strategic situation, each player proposes a suggested decision and writes a justifying **written message** to the team partner. Your task is to classify the written messages into different categories that are summarized in the following:

Category 1 Please classify the written messages into the underlying level $k \in N_0 = \{0, 1, 2, 3, \ldots\}$ of reasoning. Provide the lower and an upper bounds on the level of reasoning as described.

Category 2 Please indicate the underlying level-0 belief that is connected with the lowest conceivable level of reasoning. Information about the underlying level-0 belief that one might obtain out of the communication is how subjects respond to payoffs (payoff salience) and how subjects react to the framing (label salience) of the coordination games.

Category 3 Please indicate any approach of team reasoning that can be obtained from the written messages. Team reasoning might be expressed through a rule, label or strategy that tends to produce successful coordination if it is followed by both teams.

Please read the messages of each player, taking into account his action, and note for each player every possible level of reasoning. It is important that you 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 <u>might</u> have thought. **When you think that the information does not clearly lend itself to any inference, simply do not note any classification. Consequently, do not note anything if no statement has been made! Please note only those classifications for which you are certain.**

Remuneration

For each individual classification, the benchmark will be the classification and assessment of another classifier. Your remuneration is based on the number of matches for each classification: level lower bound, level upper bound, level-0 belief and team reasoning. A match is a classification that is congruent with the classification of another independend classifier. **Each match will be remunerated with 0.03 Euro.**

Next steps: We give you the transcripts of the messages which were as well provided to all other participants. To classify the written messages more easily, you are provided an Excel Sheet that is tailored to your particular need and task. Before you start, please make sure that you fully understood the level-k model and the concept of team reasoning. Please feel free to resolve any question you might have about the classification process and the Excel file.

Thank you.

References to these instructions

Bacharach, Michael and Dale O. Stahl, "Variable-Frame Level-n Theory", Games and Economic Behavior, August 2000, 32(2), 220-246.

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

Crawford, Vincent P., Gneezy, Uri, and Rottenstreich, Yuval, "The Power of Focal Points Is Limited: Even Minute Payoff Asymmetry May Yield Large Coordination Failures", American Economic Review, 2008, 98(4), 1443-1468.

Nagel, Rosemarie, "Unraveling in Guesing Games: An Experimental Study", American Economic Review, December 1995, 85 (5), 1313-1326.

Schelling, Thomas C., "The Strategy of Conflict", Cambrige, MA: Harvard University Press.