Carnegie Mellon University
Research Showcase @ CMU

Department of Social and Decision Sciences | Dietrich College of Humanities and Social Sciences

11-2007

Classification of Natural Language Messages Using A Coordination Game

Daniel Houser
George Mason University

Erte Xiao
University of Pennsylvania, exiao@andrew.cmu.edu

Follow this and additional works at: http://repository.cmu.edu/sds

Published In
Experiemental Economics (forthcoming), .

This Article is brought to you for free and open access by the Dietrich College of Humanities and Social Sciences at Research Showcase @ CMU. It has been accepted for inclusion in Department of Social and Decision Sciences by an authorized administrator of Research Showcase @ CMU. For more information, please contact research-showcase@andrew.cmu.edu.

CLASSIFICATION OF NATURAL LANGUAGE MESSAGES USING A COORDINATION GAME

BY

DANIEL HOUSER[a] & ERTE XIAO[b]

[a]GEORGE MASON UNIVERSITY
DHOUSER@GMU.EDU

[b]UNIVERSITY OF PENNSYLVANIA
EXIAO@SAS.UPENN.EDU

November, 2007

Abstract: The role of natural language communication in economic exchange has been the focus of substantial experimental analysis. Recently, scholars have taken the important step of investigating whether certain types of communication (e.g., promises) might affect decisions differently than other types of communication (e.g., empty talk). Doing this requires the classification of natural language messages. Unfortunately, no broadly accepted method is available for this purpose. We here describe a coordination game for objective classification of natural language messages. The game is similar in spirit to the 'ESP' game that has proven successful for the classification of tens of millions of internet images. We illustrate our game using data reported by Charness and Dufwenberg (2006). We demonstrate that our objective classification procedure allows one to infer more from the data than is possible with subjective approaches.

Acknowledgements: A grant to the first author from the International Foundation for Research in Experimental Economics funded this research. We thank John Dickhaut, Daniel Schunk and Vernon Smith for valuable feedback. We appreciate Jason Aimone's able assistance with writing instructions, running experiments and organizing the data. We thank Kail Padgitt and Dorina Tila for assistance with experiments. We are especially grateful to Gary Charness and Martin Dufwenberg for their thoughtful comments on several earlier drafts of this paper. Our thinking has been substantially improved as a result of their input. Any remaining shortcomings are entirely our responsibility.

I. Introduction

The role of natural language communication in economic exchange has been the focus of substantial experimental analysis. By exogenously varying the possibility of communication, an early and continuing experimental literature provides rigorous data on its importance to economic outcomes (see, e.g., Dawes, et al., 1977; Isaac and Walker, 1988, Ben-Ner, et al., 2007). More recently, economists have taken the important step of investigating whether certain types of communication (e.g., promises) might affect decisions differently than other types of communication (e.g., empty talk). Doing this requires the classification of natural language communication. Unfortunately, no broadly accepted method is available for this purpose. Consequently, many excellent papers in this literature have adopted reasonable but subjective procedures. We here describe an objective procedure for the classification of natural language messages. We implement the procedure on an interesting data set reported by Charness and Dufwenberg (2006), and in so doing make clear that objective classification procedures allow one to learn more from the same set of data than is possible from subjective approaches..

Formal theory results on communication as an abstract signal emerged early (e.g., Von Neumann and Morgenstern, 1944, especially section 6.4.3; Luce and Raiffa, 1957, chapter 6). Pioneering experimental research by Fouraker and Sigel (1963, chapter 4) provided early evidence that even restricted forms of communication could have substantial impact. Economic experiments allowing subjects to communicate using natural language appeared shortly thereafter (see the brief surveys by Ledyard, 1995, section III.3, and Holt, 1995, section VIII.D), and convincingly demonstrated that allowing communication can foster cooperation (or collusion) in a variety of

environments. Ledyard (1995, p. 158) points out, however, that left unanswered by this research is why.

Very recently, a number of scholars have taken up the challenge of teasing out reasons for communication's effects (e.g., Xiao and Houser, 2005, 2007; Charness and Dufwenberg, 2006; Kimbrough, et al., 2007; Schotter and Sopher, 2007; Ellingsen and Johannsen, 2007). Each of these papers follows the approach of classifying natural language messages written during experiments in ways that substantively inform the authors' scientific hypotheses. The procedures they use to do this differ.

For example, Charness and Dufwenberg (2006, discussed further below), Schotter and Sopher (2007) and Kimbrough et al., (2007) used subjective procedures to classify their natural language message data. That is, messages were classified according the authors' own opinions. These classification data were then used to draw, respectively, the following scientific conclusions: promises increase economic efficiency, fairness is a justification for rejections in ultimatum games, and people distinguish between personal and impersonal exchange environments.

We argue here that objective classifications can, at very small cost, substantially clarify and extend the nature of the conclusions reached. For example, in the Charness and Dufwenberg (2006) data we study below, our objective classification allows us to define the "promise" category more rigorously than is possible with subjective approaches. We find that Charness and Dufwenberg's (2006) results are sensitive to reasonable changes in how the "promise" category is constructed, and in a way that adds important clarity to their original findings.

II. Objective Classification of Natural Language Messages

Xiao and Houser (2005, henceforth XH) suggest a laboratory coordination game for the classification of natural language messages. To the best of our knowledge, ours is the first such suggestion to be made within the context of message classification.[1] However, it is related to the well-known 'ESP' game (von Ahn, 2005) which has been licensed by Google and is functioning successfully as their Google Image Labeler.[2] The XH game has been used for objective message classification by Xiao and Houser (2005, 2007, and in section III below) as well as Ellingsen and Johannsen (2007).

The XH classification game proceeds as follows. A group of N>2 evaluators is given a list of M>0 messages and a set of K categories (note K can be greater than, equal to, or less than M). The goal is to assign each message to a single category. Evaluators are told that after they have all completed this task a set of J ($M \geq J>0$) messages will be chosen at random. Each evaluator earns $Y>0 for each of the chosen messages such that their own message classification matches a most popular classification.

The XH game is a coordination game (for surveys see Ochs, 1995; Crawford, 1997; Camerer, 2003). In particular, there are a large number of Nash equilibria varying according to the beliefs people bring to the experiment. For example, if "category 1"

[1] It is common in the psychology literature to use external participants (e.g., not the principle investigator) to assist with a variety of evaluation tasks. To our knowledge, that literature has not used saliently rewarded decisions of independent participants in a coordination game to obtain classifications.

[2] In the ESP game participants are randomly paired. Matched participants are anonymous and cannot communicate. Each member of the pair is shown the same image. They both begin to enter words as candidate labels for the image. A word that is entered by both (not necessarily at the same time) becomes the images label (or part of the label.) The task is to label as many images as possible within a fixed amount of time. Thus, the ESP game is a coordination game with multiple equilibria. For example, if both people in the pair believe both people in the pair will use the first word on page 124 of Epstein (2007) as an image's label, then using the word "uniform" as the label is the equilibrium. This is in our view unlikely. More likely is that people will coordinate on the content of the image, and suggest labels that are sensible with respect to expected community standards and opinions. In fact, the now tens of millions of successful labels indicate that this is exactly what happens. It seems this is also what happens in our classification game.

becomes a focal point so that all evaluators believe most evaluators will assign all messages to "category 1", then that strategy is an equilibrium of this game.

Like the ESP game, the equilibrium of the XH game of interest emerges when all evaluators believe that all evaluators will classify the natural language messages according to the content of the message. Under these beliefs evaluators must read and consider each message and classify it in a way that they think is sensible by "community" standards.[3] For most cases of practical importance, including the issues concerning all of the papers discussed above, it is precisely this standard that is relevant when trying to draw an inference regarding the uncertain meaning of a natural language message.

Given that there is a single equilibrium of interest, but many that exist, it is natural to ask how people actually play this game. This is an empirical question, and the empirical evidence gathered to date, including the data in analysis described in section III below, make a cogent argument that evaluators coordinate on message content.

An alternative objective procedure for message classification is to define word-based rules that determine the category into which a message falls. For example, one could specify that a message is to be classified as a 'Promise' if the message contains the word 'promise', but only if 'promise' is not preceded by 'not'.[4] The XH game provides an objective approach to message classification when scholars are either unwilling or unable to specify rules linking the words of a message to its classification.

[3] Google's goal with the ESP game is to make image searches faster and more accurate by associating with each image a set of key-words that are sensible by "community" standards. Google's continued investment in this project suggests they are optimistic that this works.

[4] Efforts to classify by text-parsing were used early-on in the empirical reputation literature, where people attempted to gather vast amounts of auction data (usually from eBay) using 'spyders' and then categorize those data using automated text-parsing routines. Doing this turns out to be extremely difficult and unreliable. Houser and Wooders (2006) suggest an alternative approach.

The next section illustrates the XH game using data recently reported by Charness and Dufwenberg (2006). We demonstrate that objective classification is an efficient approach to shedding light on hypotheses that subjective classifications cannot inform.

III. Promises and Economic Efficiency

In a recent contribution, Charness and Dufwenberg (2006) (henceforth C&D) design laboratory experiments to test the impact of communication in a one-shot principal-agent game. They find that messages sent from principals to agents affect beliefs, and that different beliefs lead to different actions (p. 1587-1589). Moreover, they find that beliefs change when messages contain promises or statements of intent (p. 1590-1591). C&D conclude that people are sensitive to guilt in their experiment, so that promises sent from principals to agents enhance trust, cooperation and efficiency (p. 1594).

In C&D's analysis, whether a given message includes a promise or statement of intent is determined by C&D's own evaluation. Consequently, like any subjective approach, C&D need not directly confront the issue of what makes a promise a promise. Indeed, C&D's entire description of their classification procedure is: "The promise category is broad, including any statement of intent that we found (p. 1590)." Below we provide objective evidence that the classification strategy actually used by C&D required more for a message to be considered a 'promise' than this description might suggest.

Below we use the XH message classification game to reanalyze data from just one of the several treatments reported by C&D. We narrow our focus even further by restricting our attention to inferences regarding exchange efficiency, which is only one of

several dimensions C&D report in their paper. This limited analysis is sufficient for our purpose of demonstrating the scientific merit of objective classification procedures.

III.1. Overview

C&D study one-shot sequential two-player games using the strategy method. Player "A" chooses *In* or *Out*, and player "B" chooses to either *Roll* or *Don't Roll* a six-sided die. Player B's choice affects payoffs only if "A" chooses *In*. Player "B" makes her decision without knowing player A's actual choice, but under the hypothetical condition that "A" chose *In*. Table 1 details the payoffs in C&D's "(5,5)" treatment.

The unique efficient outcome of the game occurs when "A" chooses *In* and "B" chooses "*Roll*." C&D investigate behavior in treatments where players "B" can send cheap-talk pre-play messages to player "A". They find that "B" is more likely to *Roll*, and "A" is more likely to choose *In*, when "B" sends a pre-play promise to *Roll*, Thus, based on their subjective evaluations of the messages, C&D conclude that promises enhance economic efficiency.

We report below two objective analyses of C&D's message data. In one analysis evaluators are instructed to classify a message as a promise if it includes "any statement of intent." As it turns out, objective classifications under this condition do not accord well with the classifications reported by C&D. Also, we find that messages classified as "promises" in this condition do not statistically significantly enhance economic efficiency.

In addition, we obtain objective classifications of the same messages under a treatment where evaluators are instructed to classify a message as a promise only if that

message is a clear promise. The resulting classifications are more closely aligned with C&D. Moreover, under these classifications we obtain the result that promises enhance economic efficiency.

III.2. Message Classification Procedures

We recruited a total of 50 evaluators, 25 for each treatment, from the general student population at George Mason University. Evaluators were seated at visually separated desks where they worked independently. Before receiving any messages, subjects were acquainted with the C&D game and were provided with a transcript of C&D's instructions for their (5,5) 'messages' treatment. We used a quiz to ensure everyone understood the game.

Messages were transcribed directly from the C&D supplementary material on the *Econometrica* website, and presented to evaluators in different and random orders. Evaluators were not given any information regarding the decisions of the message writers or their counterparts. Also, evaluators were not given any information regarding the purpose of the study, the hypotheses of interest or C&D's previous evaluations. Evaluators were instructed to classify each message as either "Promise or Intent" or "Empty Talk" (subjects' instructions are reproduced in the appendix.)

Our 'weak promise' treatment follows C&D directly. Evaluators were told, "You should classify a message as 'Promise or Intent' if, in your opinion, it includes any statement of intent." Recall C&D described their classification procedure as: "The promise category is broad, including any statement of intent that we found (p. 1590)."

In our ‘strong promise’ treatment we instead asked evaluators to classify a message as a ‘promise’ only if they were confident that it was in fact a promise. In particular, evaluators were instructed, “You should classify a message as ‘Promise or Intent’ only if, in your opinion, it clearly is a promise.”

Messages were classified using the XH message classification game. Each evaluator was instructed to evaluate every message, and was told she would earn $5 for doing so. Evaluators knew that three messages would be randomly chosen at the end of each session. For each, if an evaluator’s classification matched the most popular classification she earned an additional $5 (so up to $15 could be earned this way). The evaluation procedure took about an hour and median earnings were $22. In both treatments an evaluator earned $7 for arriving to the lab on time.

III.3. Results

III.3.A. Coordination on Content of Message

We noted above that XH is a coordination game, and that the equilibrium of interest occurs when evaluators coordinate on the content of each message. Our evidence is that this is indeed what occurs. In particular, in both treatments when the message is a clear promise (e.g., it includes the word “promise”) evaluators labeled it a promise. On the other hand, when the message is clearly empty talk (e.g., entirely unrelated to the game) evaluators labeled it accordingly. In less clear cases opinions broke towards “promise” in the “weak promise” treatment, and towards “empty talk” when the requirement was stronger. One would predict this to occur when people coordinate using the content of the

message.[5] We are unable to discover any explanation for the patterns in our classification data other than coordination on content.

III.3.B. Strong Promises Promote Economic Efficiency

C&D obtained 42 observations in their (5,5) treatment. However, four of these included no message. All 38 written messages were classified by our evaluators. In aggregate, C&D labeled 24 messages 'Promise' and 14 messages "Empty Talk." Our 'Strong Promise' treatment yields identical frequencies. On the other hand, we obtain 31 'Promise' and 7 'Empty Talk' classifications in our "Weak Promise" treatment. In addition, for the purpose of statistical analyses, we follow C&D and group the no-message cases with those messages classified as "Empty Talk".

In both treatments most classifications are the same as those obtained by C&D. Table 2 lists the seven messages in the "weak promise" treatment, and the four messages in the "strong promise" treatment, whose classifications switched in relation to C&D's. In the "weak promise" treatment, all switches are from C&D's label of "Empty Talk" to our evaluators' labels of "Promise." In the "strong promise" case, in relation to C&D's classifications, there are two switches to "Empty Talk" and two switches to "Promise."

C&D report that the frequency of economically efficient outcomes increases from 27% to 67%, between cases where the message did not and did include a promise, and that this increase is statistically significant. ($Z=2.5$, $p<0.01$, one-tailed test). Using their same procedures (a standard proportions test), we find that "weak promises" have a

[5] Evaluators also completed a questionnaire while waiting to be paid. Their responses are additional evidence that evaluations were based on the messages' content. For example, one evaluator wrote: "There were a few messages that made it difficult to determine the intent of player B. Most messages were pretty straightforward."

somewhat smaller effect on efficiency, increasing it from 36% to 55%, and that this increase is statistically insignificant (Z=1.1, p>0.10, one-tailed test). On the other hand, "strong promises" are found to increase efficiency from 39% to 58%, and this increase is (weakly) statistically significant (Z=1.3, p=0.10, one-tailed test).[6]

III.4. Discussion

Significant differences emerged between objective classifications in our 'weak promise' treatment and C&D's subjective classifications, and all differences involved switches from "empty talk" (C&D) to "promise" (evaluators). This is evidence that evaluators in our 'weak promise' treatment used a weaker criterion for "promise or intent" than C&D had in mind. When combined with the fact that our 'strong promise' treatment leads to classifications that line up better with those provided by C&D, the evidence seems to point to the efficiency enhancing effect of promises first documented by C&D requiring something more similar to a pledge than a milder statement of intent.

Convergent evidence on this point is provided in a very recent paper by Charness and Dufwenberg (2007) that addresses whether 'watered-down' promises can promote efficiency. Their new paper uses the C&D (2006) game but eliminates free-form communication, instead allowing only 'bare-bones' promise statements. They report that such restricted statements are ineffective which, they argue, suggests "full-blooded" promises are necessary to foster cooperative economic exchange.

It is worth emphasizing that we were able to reach the C&D (2007) result using C&D (2006) data. Our fast and inexpensive message classification session obviated the

[6] We report one-tailed tests at the 10% significance level in order to be consistent with C&D, who reported one-tailed tests at the 10% significance level in their original contribution.

need for a much larger number of additional sessions with a new design. Objective message classification with the XH game is an efficient procedure to inform hypotheses that cannot be addressed using subjective classifications.

IV. Conclusion

We here described and implemented a laboratory game to generate classifications of natural language messages. Scholars have developed a variety of algorithms for classifications of behavioral decision strategies (see, e.g., Houser et. al., 2004, and cites therein), but these approaches do not easily extend to the domain of natural language. The game we described in this paper might be useful to the rapidly increasing number of scholars exploring the role of language and communication in economic exchange, especially when investigators are either unwilling or unable to adopt explicit rules linking the words in a message to that message's classification.

To shed light on mechanisms underlying communication effects, ideally we would like to collect data on messages' "true" intended meaning as well as "true" interpretations by targets of messages. Such data, unfortunately, is difficult or impossible to acquire. Self-reports are not useful here. A reason is that it can be difficult to provide incentives for truthful revelation on either meaning or interpretation (e.g., subjects report meanings or interpretations to justify their actions, independent of their actual views).

The objective classification procedure we described here offers a scientific mechanism to test hypotheses regarding both the impact of types of messages on target subjects as well as how a message writer's decisions change according to the way she expects her target to interpret her message. The reason objective classification helps is

that in typical experiments on communication subjects write messages to strangers. The average opinion of a large number of evaluators, also a stranger to the message writer, is a reasonable way to infer both the way the message was likely interpreted, as well as the way the message writer expected the message to be interpreted. The XH classification game generates this information and thus, for at least this reason, seems useful.

Very little is known about “optimal” ways to classify natural language messages. The XH classification game is objective and easily replicable, and we have pointed to evidence that it works well. Still, a systematic comparison with alternatives would be valuable. How does one assess whether a message classification game is optimal, and how does one design such a game? How do classifications arising from saliently rewarded “lay” evaluators, especially those whose demographic characteristics closely resemble the message writers, compare to professional evaluations? In ongoing research we take steps towards answering these questions.

REFERENCES

Ben-Ner, A., Putterman, L., and Ren, T. (2007). “Lavish Returns on Cheap Talk: Non-binding Communication in a Trust Experiment.” Mimeo.

Camerer, C. (2003). Behavioral Game Theory. Princeton University Press. Princeton, NJ.

Charness, G., and Dufwenberg, M. (2006). “Promises and Partnership,” *Econometrica*, 74(6), 1579-1601.

Charness, G. and Dufwenberg, M. (2007). “Broken Promises: An Experiment.” Working paper, University of Arizona.

Crawford, V. (1997). “Theory and experiment in the analysis of strategic interactions.” In D. Kreps and K. Wallis (Eds.), Advances in Economics and Econometrics: Theory and Applications. Seventh World Congress, Vol. I. Cambridge University Press. Cambridge, U.K.

Dawes, R.M., McTavish, J. and Shaklee, H. (1977). “Behavior, communication and assumptions about other people’s behavior in a commons dilemma situation.” *Personality and Social Psychology*, 35 (1), 1-11.

Ellingsen, T., and Johannesson, M. (2007). “Anticipated Verbal Feedback Induces Altruistic Behavior,” *Evolution and Human Behavior*, Forthcoming.

Epstein, J. (2007). Generative Social Science. Princeton University Press, Princeton, NJ.

Fouraker, Lawrence and Sidney Siegel. (1963). Bargaining Behavior.McGraw-Hill. New York, NY.

Holt, C. (1995). “Industrial Organization: A Survey of Laboratory Research.” In Kagel, J. and Roth, AE (eds). The Handbook of Experimental Economics. Princeton University Press, Princeton, NJ

Houser, D., Keane, M. and McCabe, K. (2004). “Behavior in a Dynamic Decision Problem: An Analysis of Experimental Evidence using a Bayesian Type Classification Algorithm,” *Econometrica*, 72:3, 781-822.

Houser, D. and Wooders, J. (2006). “Reputation in Auctions: Theory, and Evidence from eBay.” *Journal of Economics and Management Science.*

Issac, R.M. and Walker, J. (1988). “Communication and Free Riding Behavior: The Voluntary Contributions Mechanism," Economic Inquiry 26: October 1988.

Kimbrough, E., Smith, V. and Wilson, B. "Historical Property Rights, Sociality, and the Emergence of Impersonal Exchange in Long Distance Trade." *American Economic Review*, Forthcoming.

Luce, R. D. and Raiffa, H. (1957). Games and Decisions: Introduction and Critical Survey. John Wiley and Sons, Inc.

Ledyard, J. O. (1995). "Public Goods: A Survey of Experimental Research." In Kagel, J. and Roth, AE (eds). The Handbook of Experimental Economics. Princeton University Press, Princeton, NJ

Ochs, J. (1995). "Coordination Problems." In Kagel, J. and Roth, AE (eds). The Handbook of Experimental Economics. Princeton University Press, Princeton, NJ.

Schotter, A. and Sopher, B. 2007. Advice and behavior in intergenerational ultimatum games: An experimental approach. *Games and Economic Behavior*. 58: 365-393.

von Ahn, L. (2005). Games with Purpose. PhD Dissertation. Carnegie Mellon University.

von Neumann, J. and Morgenstern, O. 1944. Theory of Games and Economic Behavior, Princeton Univ. Press, Princeton NJ.

Xiao, E., and Houser, D. (2005): "Emotion Expression in Human Punishment Behavior," *Proceedings of the National Academy of Sciences* , 102(20), 7398-7401.

Xiao, E., and Houser, D. (2007): "Emotion Expression and Fairness in Economic Exchange," Manuscript, George Mason University.

Table 1: Payoff Table for C&D "(5,5)" Game

		A: *In*		
			B: *Roll*	
	A: *Out*	B: *Don't Roll*	*Failure* [p=1/6)]	*Success* [p=5/6]
Payoff (A,B)	(5,5)	(0,14)	(0,10)	(12,10)

Table 2: Messages where C&D and Evaluators have different classifications

Message	Classification		"P" frequency (%)
	C&D	Evaluators	
Weak Promise Treatment			
— Please choose In so we can get paid more.	E	P	15(60%)
— If you stay in, the chances of the die coming up other than 1 are 5 in 6 – pretty good. Otherwise, we'd both be stuck at $5. (If you opt out)	E	P	21(84%)
— If you choose IN the first round and then I will choose Don't Roll at first. I will get $14 but then after that I will choose roll each time after the 1st role. Chances are most likely you will get $12 and I will get only $10.I will the only take 7 rolls for you to get even with me. That way we both leave with a good amount of money. Hope you have a great evening and that this works out for both of us. ☺	E	P	21(84%)
— If you choose IN, and I roll, the chances of our getting the most $ are very high. The likelyhood of my rolling a 1 is small compared to the chances of rolling a 2–6. So we both get cash.	E	P	19(76%)
— Hopefully I'll make a lucky role.	E	P	15(60%)
— Hello fair stranger, anonymous partner … Choose whatever you want. Far be it from me to influence your decision, but I think you should choose "in" and I should choose "roll" and we should take the chance at both earning as much as we can. 5 chances out of 6 say it'll work, and I'm totally broke, looking to rake in stray cash however I can. I feel the luck in the air. I don't really have much else to say. Hope you're doing well, whoever you are. Yes. That's all. Random note from random human	E	P	14(56%)
— If you choose IN you have the best opportunity to make the most money. You have a 5/7 chance of making more money! So IN would be your best bet. Cheers. ☺	E	P	16(64%)
Strong Promise Treatment			
— If I roll a 2–6 (you'll know when you receive the $, you will give $5.00 to a stranger. [[[then there is a line, under which is written "Sign here if you are so kind]]] Thanks. You'll still be gaining more than if I had chosen Don't roll.	P	E	6 (24%)
— If you choose IN the first round and then I will choose Don't Roll at first. I will get $14 but then after that I will choose roll each time after the 1st role. Chances are most likely you will get $12 and I will get only $10. I will the only take 7 rolls for you to get even with me. That way we both leave with a good amount of money. Hope you have a great evening and that this works out For both of us. ☺	E	P	19 (76%)
— Choose "In" so we can both make some $$ What are the chances me rolling a 1? I'll try my best.	P	E	10(40%)
— If you choose IN, and I roll, the chances of our getting the most $ are very high. The likelyhood of my rolling a 1 is small compared to the chances of rolling a 2–6. So we both get cash.	E	P	14(56%)

Note: E: classified as "Empty Talk"; P: classified as "Promise".

Appendix: Transcript of Instructions to Subjects

Page 1

Thank you for coming! You've earned $7 for showing up on time, and the following instructions will explain how you can make decisions to earn more money.

Your task:

In this experiment, you will be given a list of messages. Your task in this experiment is to evaluate whether each of the messages is:

- A statement of intent or promise
- Empty Talk

Your payoff:

You will earn $5 by evaluating all the messages. At the end of the experiment **3** messages will be randomly chosen. If your evaluation matches the most common evaluation, you will be paid **$5** for that message.

The messages were written by participants in a previous experiment (Experiment I). To evaluate the messages, you need to first understand Experiment I. The next few pages describe Experiment I. Please read it carefully. The message writer is in the role of **subject B.**

Page 2

Experiment I

Message Writers' (Subject B's) Instructions:

Thank you for participating in this session. The purpose of this experiment is to study how people make decisions in a particular situation. Feel free to ask us questions as they arise, by raising your hand. Please do not speak to other participants during the experiment. You will receive $5 for participating in this session. You may also receive additional money, depending on the decisions made (as described below). Upon completion of the session, this additional amount will be paid to you individually and privately. During the session, you will be paired with another person. However, no participant will ever know the identity of the person with whom he or she is paired.

Decision Tasks

In each pair, one person will have the role of A, and the other will have the role of B. The amount of money you earn depends on the decisions made in your pair. On the designated decision sheet, each person A will indicate whether he or she wishes to choose IN or OUT. If A chooses OUT, A and B each receive $5. We will collect these sheets after the choices have been indicated. Next, each person B will indicate whether he or she wishes to choose ROLL or DON'T ROLL (a die). Note that B will not know whether A has chosen IN or OUT; however, since B's decision will only make a difference when A has chosen IN, we ask B's to presume (for the purpose of making this decision) that A has chosen IN. If A has chosen IN and B chooses DON'T ROLL, then B receives $14 and A receives $0. If B chooses ROLL, B receives $10 and rolls a six-sided die to determine A's payoff. If the die comes up 1, A receives $0; if the die comes up 2–6, A receives

$12. (All of these amounts are in addition to the $5 show-up fee.) This information is summarized in the chart below:

	A Receives	B Receives
A chooses OUT	$5	$5
A chooses IN, B chooses DON'T ROLL	$0	$14
A chooses IN, B chooses ROLL, die = 1	$0	$10
A chooses IN, B chooses ROLL, die = 2,3,4,5, or 6	$12	$10

Prior to the decision by A and B concerning IN or OUT, B has an option to send a message to A. Each B receives a blank sheet, on which a message can be written, if desired. We will allow time as needed for people to write messages, then these will be collected. Please print clearly if you wish to send a message to A. In these messages, no one is allowed to identify him or herself by name or number or gender or appearance. (The experimenter will monitor the messages. Violations (experimenter discretion) will result in B receiving only the $5 show-up fee, and the paired A receiving the average amount received by other A's.) Other than these restrictions, B may say anything that he or she wishes in this message. If you wish to not send a message, simply circle the letter B at the top of the sheet.

Page 3

B

You may print a message to A below if you wish.

Page 4

ID_____ Name _____

The following are the messages to evaluate. Please put an **X** in the box, provided after each message, which corresponds with your evaluation of whether the message shows intent or a promise or shows empty talk. Please put only one **X** per message.