Understanding the Plott-Wit-Yang Paradox

Katarína Kálovcová and Andreas Ortmann

CERGE-EI*

June, 2009

Abstract

Plott et al. (2003) conduct a betting market experiment and find: First, information was aggregated. This suggests that traders updated their private information based on observed market odds. Second, a model based only on the use of private information seems to fit their data best. The authors call this paradoxical. Because the original data are lost, we replicate their experiment. Our results suggest that the paradox seems due to aggregate rather than individual level data analysis. We analyze the individual level data and explain the paradoxical results reported in Plott et al. (2003).

Keywords: experimental betting markets, private information, information aggregation

JEL classification: D81, D82, G14

*CERGE-EI is a joint workplace of the Center for Economic Research and Graduate Education, Charles University, and the Economics Institute, Academy of Sciences of the Czech Republic. Address: CERGE-EI, P.O. Box 882, Politických vězňu 7, Prague 1, 111 21, Czech Republic
E-mail: katarina.kalovcova@cerge-ei.cz, andreas.ortmann@cerge-ei.cz, aortmann@yahoo.com

Acknowledgement: We would like to thank Charles R. Plott and Peter Katuščák for invaluable comments. Support through GDN grant No. RRC VIII-89 is gratefully appreciated.

1 Introduction

A (parimutuel) betting market as typically used in horse racing and other sports events, is a system in which all bets are collected and the payoffs are then determined by dividing the total amount of money invested by the amount bet on the winning horse.

In Plott et al. (2003), the authors address experimentally two fundamental questions: first, is information aggregated on betting markets? Noting that there is no clear theoretical reason why betting markets should aggregate information at all, the authors report that the implicit prices on their experimental markets are very close to the prices that would exist if all agents pooled their information and made decisions on the basis of the pooled data. This observation suggests that the information in their markets does aggregate. Second, which model explains best how information is aggregated? The theoretical model which seems to fit their data best (the Decision Theory Private Information, or DTPI, model) does not rely on information aggregation whatsoever. The authors call this paradoxical. We refer to their result below as the Plott-Wit-Yang (PWY) paradox.

We replicate their experiment with minor changes and find, first and like Plott et al., the paradoxical result that information is aggregated while the data seem to be explained best by a theoretical model that does not require information aggregation.

We show that market odds are indeed very close to odds that would exist if traders behaved according to the DTPI model. However, our individual level data analysis suggests that, apart from the private information, traders extract significant additional information from observing the market odds. The PWY paradox seems due to aggregate rather than individual level data analysis.

We also observe a learning effect: In later rounds traders seem to understand the mechanism of the betting market better and put higher weight on the information contained in the market odds rather than private signals. One plausible explanation is that subjects become increasingly familiar with the laboratory environment.

Finally, we examine the effect of risk-aversion on traders' behavior. We find that the degree of risk aversion does not have any impact on the amount of money traders bet on our experimental market.

In the next section we discuss the PWY paradox, illustrate our explanation and formally state our hypothesis. In section 3 we explain design and implementation of our experiment. Results are reported in section 4 and we conclude in section 5.

2 The PWY Paradox

The paradox consists of two results that contradict each other: first, information is aggregated on the market, i.e. traders are involved in some sort of strategic behavior. Second, if we want to simulate the behavior of traders our best bet is to use the DTPI model which is based on the use of private signals only. Our explanation of the PWY paradox is based on a detailed analysis of the second result. We show that, while aggregate level data might suggest that traders follow the DTPI model while individual data analysis might lead to a different conclusion because two different trading behaviors can lead to the same aggregate results. Our argument can be illustrated by the following example:

Example: Suppose that there are only two traders on the market (Trader 1 and Trader 2) with the same budget and only two ex-ante equally likely events A and B that traders can bet on. Further suppose that based on their private signal, Trader 1 thinks that A is the winning event and Trader 2 thinks that B is more likely to win. If both traders behave according to the DTPI model then every trader invests all the money into the more likely event and the resulting market odds are 2:1 for both events A and B. Alternatively, traders can behave strategically and by observing the other trader's actions they learn about each other's information. Consequently they both invest half of the budget into each event. Again, the resulting market odds are 2:1 for each event.

When searching for an underlying model, Plott et al. (2003) look at the aggregate level data and conclude correctly that the prediction of the DTPI model fits market odds the best. In terms of the example above, Plott et al. observe market odds 2:1 for each event and claim that the prediction of DTPI model is also 2:1 for each event and hence traders follow this model. Our evidence suggests that it is not necessarily true.

Hypothesis: *Traders do take into account information contained in their private signals and information contained in market odds.* In other words, traders observe behavior of others and based on market odds they update private signals. Through this process information is aggregated and translated into market odds.

3 Our Experimental Betting Market[1]

Because the original data are lost, we replicate the Plott et al. (2003) design. We change minor implementation details all designed to allow us to analyze the impact of risk-aversion.

3.1 Design

The design of our betting market follows the one in Plott et al. (2003). Subjects bet on six events labeled A, B, C, D, E and F which are equally likely ex ante. Each round one of the letters is drawn at random from an urn, recorded, and then placed back in to the urn. In other words, the draw of an event is independent across rounds and the history of draws holds no implications of what future draws might be. Which of the events wins is announced after the end of each round. After the winning event is chosen, each individual is privately given a noisy signal (or "clue") about the winning event. The clues are determined independently for each individual by the following procedure. Once the winning event is determined a new urn is created with five letters of the winning event and two letters from each of the other events. The participant is informed of the outcome of three random draws with replacement. The information distributed across all participants in a session is more than that of any one individual. However, this information is not sufficient to determine the winner with absolute certainty.

[1]The experiment was programmed in z-Tree (Fischbacher 2007). Instructions for this experiment and the data can be found at: http://home.cerge-ei.cz/kalovcova/research.html

3.2 Implementation

Employing 109 undergraduate students, we conducted our experiment in four sessions in February 2008 and additional five sessions in March 2009 each of which consisted of one trial round (which did not affect the earnings and was intended to make subjects familiar with the software) and then continued with 16 regular rounds. Time, in seconds, was displayed on each computer screen. The duration of each round was 120-300 seconds - the time of duration was chosen randomly and independently for each round and was unknown. At the end of the experiment four rounds were randomly chosen and subjects were paid based on their performance in the paying periods. The price of each event ticket was 1 ECU (experimental currency unit), once a ticket was bought it could be neither returned nor resold. At the beginning of each round subjects were endowed with 300 ECU which they were free to spend or to keep. That part of the endowment that was not spent declined in value as subjects were allowed to keep only three quarters of it. After subjects spent their entire endowment, they could get a loan of 600 ECU which had to be paid back after the end of each round. The payoff for each round was determined in the following way:

$$\begin{aligned} \text{Payoff} = {} & 0.75 \times \text{money on hand (part of endowment or loan not spent)} \\ & + \text{profit} \\ & - \text{loan payback (if the loan was taken)} \end{aligned}$$

where

$$\text{profit} = \frac{\text{Total ECU from all ticket sales}}{\text{Total number of winning tickets sold}} \times \text{Number of winning tickets held}$$

The implementation of our experimental betting market differs from that in Plott et al. (2003) in four respects. All four changes served the additional purpose of creating a betting market in which we could observe the effect of risk aversion. That risk aversion might have an impact is strongly suggested by the literature. For recent and comprehensive review see Harrison and Rutström (2008). We believe, and the evidence below suggests, that these differences in implementation do not affect participants' behavior to the extent that is relevant for an examination of the PWY paradox. In the section below we discuss the implementation changes. In the results section we discuss briefly the effects of the risk aversion and the specifics of the risk-aversion instrument we used. We focus, however, mostly on results directly connected to the PWY paradox.

1. Risk-aversion. To measure the level of risk aversion, we administered the assessment instrument proposed in Holt and Laury (2002) and now widely used for that purpose. Participants were financially incentivized for this part of the experiment. To control for the order effect we administrated the risk aversion measure prior to the betting market part of the experiment in first four sessions and after the betting market part in the next five sessions of our experiment.

2. House bonus. In Plott et al. a house bonus is used. A house bonus is the money added to the total amount of money invested by all the subjects. The expected payoff from the investment is thus strictly positive and gives risk-averse subjects better incentives to invest: The house bonus makes investment more profitable and the more subjects invest the more information can be aggregated. The house bonus seems responsible for the successful information aggregation on the betting markets in Plott et al. (2003). However, this mechanism calls for investing in the market as much as possible. Hence the traders who fully understand this mechanism invest all the money they have irrespective of their attitude towards risk. Only extremely risk-averse subjects would do otherwise. In real betting markets house charge is used instead of house bonus. The expected payoff from the investment is then slightly negative and risk-aversion is likely to play a significant role. In our experimental betting market neither house bonus nor house charge is used.

3. Endowment depreciation. Without house bonus, traders are less motivated to invest in the market. Pilot experiments that we conducted confirmed this hypothesis: Subjects spend 78% of all the money at their disposal in markets with house bonus whereas they spend only 64% in markets without house bonus. To enhance the process of information aggregation we wanted to make sure that traders would spend a major part of their endowment. Plott et al. use an experimental design in which the part of the endowment that is not spent is lost. This makes all subjects spend the entire endowment. In our betting market that part of the endowment that is not spent declines in value and subject are allowed to keep only three quarters of it. This design creates strong incentives for subjects to spend a major part of the endowment and thus allows for information aggregation. At the same time extremely risk-averse participants are allowed to keep all the endowment and earn a small but sure profit. Hence we enhance the process of aggregation of information while keeping risk-aversion to play a significant role.

4. Paying periods. After the experiment was finished, we randomly selected four periods for which subjects were paid and this was ex-ante known to all participants. (In Plott et al. subjects were paid in all rounds.) We implemented this payment mode to prevent subjects doing nothing and only shortly before the market is closed investing all the money into the event with the lowest odds, i.e. the most likely event. Most of the time the true event is identified successfully and hence the sniping strategy that we observed in pilot markets, leads to a large long-run profit (negative profit in a few periods is offset by a large positive profit in most of the periods). However, extensive waiting worsens information aggregation because subjects who wait keep their private information away from the market. With our payment mode, waiting and investing everything in the most likely event becomes less attractive because those periods where the profit is negative could be chosen to be paying periods.

4 Results

We start this section with two results that constitute the Plott-Wit-Yang paradox. Then we follow with the third - key result - that supports our hypothesis and we finish with three supplementary results concerning learning effect, market efficiency, and effect of risk-aversion.

Result 1: Information is aggregated. Similar to Plott et al. (2003), we find evidence in favor of information aggregation. The results are provided in table 1 below in form of the Würtz[2] measure of the distance of model predictions from AIA (Aggregated Information Available, i.e. posterior probabilities given the pooled signal of all traders). The Würtz measure is computed for aggregate data. For example, the Würtz measure of distance between DTPI and AIA is determined in the following way: first we compute what market odds would be if all traders behaved according to the DTPI model and determine the corresponding probability distribution p_i. Then we take the probability distribution given by AIA, q_i, and use the formula in footnote 2 to compute their distance.

In table 1 below we follow the notation in Plott et al. (2003):

• Decision Theory Private Information Model (DTPI) - model where traders base their decisions exclusively on their own private information and bet all their money on the most likely event.

• Competitive Equilibrium Private Information Model (CEPI) - model where traders take market odds as constants and maximize their expected profit with respect to their private information.[3]

• Average Opinion statistics - the average of individual beliefs before the market opens.

• Best Opinion statistics - the most accurate belief among traders' beliefs before betting.

• Implicit prices (IP) - market prices implicitly determined by the market odds.

[2]If the discrete distributions are described by their probability density functions $\{p_i\}_{i=1\ldots K}$ and $\{q_i\}_{i=1\ldots K}$ respectively, then the measure proposed by Würtz (1997) can be written as $W(p,q) = 0.5\sum_{i=1}^{K}|p_i - q_i|$.

[3]We use the method described in Eisenberg and Gale (1959) and `Mathematica` to compute equilibrium odds.

Table 1: Average Würtz measure of distance from AIA:

all periods:

Best Opinion	IP	DTPI	Average Opinion	CEPI
0.380 (0.163)	0.495 (0.193)	0.515 (0.102)	0.634 (0.099)	0.663 (0.101)

last 8 periods:

Best Opinion	IP	DTPI	Average Opinion	CEPI
0.427 (0.187)	0.489 (0.191)	0.511 (0.111)	0.627 (0.105)	0.657 (0.109)

The results in table 1 show that the distribution of probabilities based on IP is closer to the distribution given by AIA than the prediction of any other model except of BO. For example, in the first row of table 1 the Würtz measure of the distance between AIA and IP is 0.495 which is lower than the Würtz measure of the distance between AIA and any other model expect BO. This means that apart from BO, IP is closer to the AIA than the prediction of any other model.

We also observe that information aggregation improves over time. At the beginning of the experiment the information aggregation is weaker which is probably caused by the inexperience of participants (the average Würtz measure of the distance between AIA and IP across all periods is 0.495). As the experiment continues participants understand the mechanism better, behave more strategically, and try to update their own signal based on what happens on the market. Hence in later periods the information aggregation is more obvious (the average Würtz measure of the distance between AIA and IP across the last 8 periods is 0.489).

Result 2: DTPI model fits the data from betting markets the best.

Table 2: Average Würtz measure of distance of model predictions from IP:

all periods:

DTPI	Average Opinion	CEPI	Best Opinion	AIA
0.261 (0.124)	0.306 (0.193)	0.330 (0.134)	0.324 (0.161)	0.495 (0.134)

last 8 periods:

DTPI	Average Opinion	CEPI	Best Opinion	AIA
0.269 (0.126)	0.330 (0.123)	0.355 (0.120)	0.340 (0.169)	0.489 (0.191)

The average Würtz measure of the distance between IP and DTPI across all periods (0.261) and across the last 8 periods (0.269), is lower than the average Würtz measure of the distance between IP and any other model. This means that DTPI model fits the experimental data the best. Results 1 and 2 are in line with the results in Plott et al. (2003) and constitute the PWY paradox. In the following section we provide an explanation for this paradox.

Result 3: Our PWY paradox explanation is supported by the data.

First, we find that traders invest on average one third of their overall investment into events that they should ignore according to DTPI model. Second, we compare the observed individual distribution of bets to the distribution of bets implied by private signals (the DTPI model) and the distribution of bets implied by market odds (bets are in proportions to their probabilities implied by market odds). For this comparison we use again the Würtz measure. Our hypothesis that traders do take into account information contained in their private signals and in market odds implies that the Würtz measure (Würtz criterion, WC) of distance between observed individual behavior and private signal is approximately the same or larger than the WC of distance between observed individual behavior and behavior induced by the market odds (WC[Behavior-Signal]$\geq$WC[Behvior-Odds]. Note that the smaller WC, the lower the distance between two distributions).

We run t-test on our data and find extensive support for this hypothesis. We can reject the null hypothesis that the Würtz measure between observed behavior and market odds is the same as the Würtz measure between observed behavior and private signal (WC[Behavior-Odds]=WC[Behvior-Signal]) in favor of the alternative hypothesis that the Würtz measure between observed behavior and market odds is lower the Würtz measure between observed behavior and private signal (WC[Behavior-Odds]<WC[Behvior-Signal]) at any reasonable level of significance (p-value is 0.00). In other words, we find support for the fact that traders rely on the signal contained in market odds more than they rely on their private information.[4]

Moreover, for every trader we analyze separately periods with strong signal (at least two out of three draws are the same; i.e. the probability of the most likely event is 50% or 75%) and weak signal (all three draws are different; i.e. three most likely events are equally likely with probability to occur equal to 24% each). We find that traders follow market odds more closely than their private signal irrespective of the quality of their private signal. We can reject the null hypothesis that the WC[Behavior-Odds]=WC[Behvior-Signal] in favor of the alternative hypothesis that WC[Behavior-Odds]<WC[Behvior-Signal] at any reasonable level of significance (p-value is 0.00 in both cases).

To provide an additional insight into data we analyze the group of rounds in which the private signal is in line with market odds and the group of rounds in which private signal and market odds differ. We find that:

- in the first group: WC[Behavior-Odds]=WC[Behvior-Signal] (p=0.10 with two-sided alternative hypothesis; p=0.05 with one-sided alternative hypothesis)
- in the second group: WC[Behavior-Odds]<WC[Behvior-Signal] (p=0.00)

[4]We also run a non-parametric Wilcoxon rank-sum test. The results are qualitatively the same.

Therefore we conclude that traders form weighted average of their private signal and market odds with approximately equal weights when their signal is consistent with market odds. However, traders trust their private signal significantly less if it contradicts the market odds.

Result 4: Traders' behavior is increasingly influenced by public signals.

We analyze the first and the last eight periods of our experiment separately and we find support for a learning effect: We find that traders follow market odds more than private signals in the first eight periods and they rely on market odds even significantly more during the last eight periods. In particular, traders follow private signals slightly less in latter periods (however, this result is not statistically significant) and secondly traders follow market odds significantly more in latter periods (p-value is 0.00).

These results suggest that after the traders understand the mechanism of betting markets better and learn that the market works well identifying the winning event, traders shift weight towards the public information in form of market odds.

Result 5: Market is efficient. Betting experiment exhibits weak statistical efficiency.

Table 3: Winning probabilities assigned by the betting market and actual frequencies of winning.

Market Rank by IP	Average IP	Frequency of Winning	Standard Error of Frequency of Winning	t-statistics
1^{st}	0.517	0.660	0.150	-1.060
2^{nd}	0.191	0.132	0.077	0.522
3^{rd}	0.108	0.125	0.044	-0.205
4^{th}	0.078	0.069	0.034	0.250
5^{th}	0.060	0	0.030	2.014
6^{th}	0.046	0.014	0.026	1.220

In table 3 markets are ranked according to the average implicit price (IP) for all sessions. The average IP of the 1^{st} market is 0.517. Actual relative winning frequency of the 1^{st} market is 0.660. We can not reject the null hypothesis that the two distributions (column 2 and column 3) are the same. As a result we can not reject weak statistical efficiency of this betting market with the exception of the 5^{th} market for which the implicit price is significantly larger than the actual frequency of winning. Hence, the efficiency of the market is not so profound. We also observe a favorite long-shot bias in our markets: the market probability for favorites is understated (0.517 with actual winning frequency of 0.660) and the probability for long-shots is overstated (0.046 with actual winning frequency 0.014). However, this result is not statistically significant.

Result 6: Risk Aversion Does not Effect the Level of Investment. Out of 109 participants in our experiments, there were 24 participants for whom the level of risk-aversion could not be measured and they were omitted from further analysis.[5] We divide the remaining 88 (35 from the first and 50 from the second round of experiments) participants into two groups - 51 more (15 from the first and 36 from the second round of experiments) and 34 less (20 from the first and 14 from the second round of experiments) risk averse participants. The null hypothesis that less and more risk-averse participant spend on average the same amount of money can not be rejected at any level of significance (p-value is 0.9) in favor of the alternative hypothesis that less risk-averse traders spend more money. We do not observe any significant difference in risk-aversion distribution (p value in Wilcoxon-Mann-Whitney rank sum test is 0.43). If we look at the data from 2008 - risk-aversion measure comes first - separately we find that less risk-averse participants spend on average 20% more than more risk-averse individuals. The null hypothesis that less and more risk-averse participant spend on average the same amount of money can by rejected at 10% level of significance (p-value is 0.9) in favor of the alternative hypothesis that less risk-averse traders spend more money. We can say that more risk-averse individuals will participate less and hence

[5] These individuals made multiple switches between safe and risky lottery.

their private information will have less of an impact on implied prices, with a resulting loss in efficiency. If the risk-aversion measure comes second (2009), participants invest on average the same amount irrespective of their risk aversion. We tested for the order effect of the risk aversion assessment instrument and we did not find any differences in the participants' distribution of risk aversion among the first series and the second series of sessions.

5 Conclusion

We replicated the experimental betting market in Plott et al. (2003). Our data confirm the Plott et al. findings on their level of analysis. Specifically, our analysis showed that aggregate data suggest that traders follow the DTPI model. Individually, traders do take into account information and behavior of other traders in form of market odds, though. Based on this finding we explained the PWY paradox. Furthermore, we found a learning effect on our betting market. In later rounds traders put less weight on their private signal and rely more on the signal contained in market odds. Finally, we do not find any effect of a degree of risk aversion on traders behavior.

References

Eisenberg, E. & Gale, D. (1959). Consensus of Subjective Probabilities: The Pari-Mutuel Method. *The Anals of Mathematical Statistics* 30(1). pp.165-168.

Fischbacher, U. (2007): z-Tree: Zurich Toolbox for Ready-made Economic Experiments, *Experimental Economics* 10(2). pp. 171-178.

Harrison, G. W. et al. (2005). Risk Aversion and Incentive Effects: Comment. *The American Economic Review* 95(3). pp 897-901.

Harrison, G. W. & Rutström, E. (2008). Risk Aversion in the Laboratory, in Cox, J. C. & Harrison, G. W. (eds.), *Risk Aversion in Experiments* (Bingley, UK: Emerald, Research in Experimental Economics, Volume 12, 2008).

Holt, C. A. & Laury, S. K. (2002). Risk Aversion and Incentive Effects. *The American Economic Review* 92(5). pp 1644-1655.

Manski, C. (2004). Interpreting the Predictions of Prediction Markets. NBER Working Paper No. 10359.

Plott, C. R., Wit, J. & Yang, W. C. (2003). Parimutuel Betting Markets as Information Aggregation Devices: Experimental Results. *Economic Theory* 22(2). pp 311-351.

Roust, A. K. & Plott, R. C. (2005). The Design and Testing of Information Aggregation Mechanisms: A Two-Stage Parimutuel IAM. Social Science Working Paper No. 1245.

Wolfers, J. & Zitzewitz, E. (2005). Interpreting Prediction Market Prices as Probabilities. Mimeo, University of Pennsylvania.

Wolfers, J. & Zitzewitz, E. (2006). Five Open Questions About Prediction Markets, in I*nformation Markets: A New Way of Making Decisions in the Public and Private Sectors*, ed. Robert Hahn and Paul Tetlock, AEI-Brookings Joint Center, Washington D.C.