A Comparison of Interest Arbitrator Decision-Making in
Experimental and Field Settings

By

Craig A. Olson
Princeton University &
University of Wisconsin - Madison

Gregory G. Dell'Omo
Canisius College

Paul Jarley
University of Iowa

May 1991
Revised

The authors thank Orley Ashenfelter, Henry Farber, two anonymous reviewers and seminar participants at Princeton University for their comments. Earlier versions of this paper were presented at the 1987 Winter Meetings of the Econometrics Society and the John M. Olin-Russell Sage Conference in Dispute Settlement held at the Center For Advanced Studies in the Behavioral Sciences, November 1989. Support for this research was provided to the first author by the Vilas Foundation and the Industrial Relation Section, Princeton University.

Abstract

Recent studies have investigated arbitrator decision rules in both experimental and field settings. The external validity of experimental studies is evaluated by comparing the decisions made in an experiment with those made in actual cases for a common group of arbitrators. The results show the decision models used in the two settings are very similar when the decision problem in the two settings is carefully controlled to be the same.

A Comparison of Interest Arbitrator Decision-making in Experimental and Field Settings

The decision-making process used by arbitrators to settle new contract disputes between labor and management has been the subject of several recent empirical studies (Ashenfelter & Bloom 1984; Bazerman 1985; Bazerman & Farber 1985; Bloom 1986; Dell'Omo 1987, 1989; Farber & Bazerman 1986; Olson & Jarley 1991). This research is motivated by the central role of arbitrator decision-making in the evaluation of interest arbitration. More specifically, interest has been focused on identifying the factors that are important in determining the fair award beliefs of arbitrators and whether or not these beliefs can be influenced by the parties' final offers.

Two research strategies have been followed in recent studies of this topic. In one design (Ashenfelter & Bloom 1984; Olson & Jarley 1991) data on actual arbitration decisions are used to estimate the mean and variance of the distribution describing the fair award beliefs of arbitrators. The parameters of this distribution are estimated using data describing factors believed important in wage determination: the characteristics of the disputing parties, the bargaining environment, and the parties' positions on the unresolved issue. Aside from the parties' position on the unresolved issue, the actual information communicated by the parties to the arbitrator during the hearing is typically not known. The second, more popular, strategy (e.g. Bazerman 1985; Bazerman and Farber 1986; Farber and Bazerman 1986; Bloom 1986) is to ask arbitrators to decide one or more hypothetical disputes where the facts and offers are experimentally varied.

The popularity of the experimental approach is a function of the difficulties associated with collecting and analyzing field data. For one, analysis of field data is complicated by the existence of different types of arbitration. Studying conventional arbitration decisions is difficult for two reasons; first, there is frequently no documentation of participants' positions presented to the arbitrator and second; the effect of offers on arbitrator beliefs must be controlled when evaluating the effect of the offers on awards (Ashenfelter and Bloom 1984; Bloom 1986).

Although the use of field data to study decision-making under final offer arbitration does not suffer from these problems, the empirical results have sometimes been difficult to interpret. In the Ashenfelter & Bloom study, the coefficients on the variables that were thought to influence arbitrators' views of an appropriate award were very unstable across the three years under study. For two of the four variables, the estimated parameters were statistically significant in each year, but were of different signs from one year to the next. The signs for the other two variables were the same across the three year period, but the estimates for one year were at least half the size of those for the other years. As the authors note, this variation in results across years may be due to important variables influencing arbitrator beliefs that were not included in their analysis.

An experiment can overcome the shortcomings of field data. The inherent control in the experimental design maximizes internal validity by ensuring there are no omitted characteristics of the dispute or simultaneity between awards and offers that might bias parameter estimates. Such a design has special appeal in the case of arbitrator decision-making, since theory has failed to suggest a clear set of factors that influence arbitrator

preferences. In addition, orthogonal construction of the facts can facilitate inferences about the dominant factors in determining arbitrator preferences.

Balanced against these advantages are concerns over the external validity of simulations. Although the experimental studies in this area have tended to use actual decision-makers (ie. arbitrators) rather than college students, these studies possess many of the same characteristics that have lead some to question the generalizability of experimental research on negotiation and arbitration (Anderson 1981; Gordon, Schmitt and Schneider 1984). One criticism involves a lack of "ecological validity" (Locke 1986). Reading and deciding cases based on short scenarios differs sharply from hearing testimony during the arbitration process. In an experiment, the scenarios present "facts" to the decision-maker as if the parties had stipulated to them. For example, the composition of the outside wage comparability group is taken as undisputed by the parties. This is often not the case in actual disputes. Where the "facts" are in dispute, the arbitrator must either resolve these disputes or discount the facts when deciding on the appropriate award.

Second, decisions in an experiment are made without the pressure to make an "acceptable" award that is found in actual cases. In an experiment an arbitrator knows the decisions have no consequences for his or her future arbitration practice and, as a result concerns about the decisions' acceptability to the parties may be absent.

Finally, the practice of orthogonal construction of the independent variables may actually hamper external validity. For example, in the Bazerman and Farber study, where the offers were constructed without regard to the facts of the case, a substantial minority of conventional arbitration awards were made that were equal to or outside the boundaries defined by the

positions of the parties. Such awards, which are rare in actual disputes, could indicate that the orthogonal construction of facts and offers affected decisions in all the experimental disputes and biased the results (Bloom 1986; Olson 1988). Surely arbitrators expect the position of the parties to be influenced by the factors presented in an arbitration hearing and they may alter their decision rules upon discovering that this is not the case.

Although these limitations call into question the external validity of laboratory research, the assessment of external validity is more an inductive than deductive exercise (Locke 1986:5). Lab settings do differ from those found in the field, but empirical evidence is needed to determine whether, and how, results from field and laboratory research on arbitrator decision-making differ. This study provides such evidence by comparing similar experimental and field decisions made by a common group of arbitrators. The decisions made by nineteen arbitrators who participated in an experiment (Dell'Omo 1987, 1989) are compared with the field decisions these same arbitrators made in teacher disputes under Wisconsin's Mediation-Arbitration statute. The results show a high degree of consistency in the decisions of the arbitrators across the two settings and suggest that much can be learned about actual arbitrator decision-making through the use of experiments.

The Experimental Data

Twenty-two arbitrators were asked to decide 32 hypothetical interest disputes during the fall of 1986. The twenty-two arbitrators were selected to participate in the study because they have made numerous decisions under Wisconsin's two interest arbitration statutes.¹ Nineteen of the original twenty-two subjects were used in this study. One subject in the original

study refused to allow his or her data to be used in research beyond the initial study. Two arbitrators in the experiment did not decide any teacher disputes during the period studied so their decision-making in the two designs could not be compared. The thirty-two hypothetical disputes were between a city and police union and dealt with only a single issue, the hourly wage for police officers. The nine variables described in Table 1 were manipulated in the disputes. The variables fell into three broad categories; 1) wage levels and wage increases for "comparable" bargaining units, 2) changes in the cost of living, and 3) financial characteristics of the employer.

The 32 scenarios were constructed by a two step process designed to minimize the correlation between the treatment levels. First, the treatment values were divided into "high" and "low" groups with four values in each category. Then, the criteria were combined based on the two categories according to a 1/16 fractional replication of a 2^9 factorial design. This produced 32 scenarios where the two ordinal levels were orthogonal. Because of variation within the "high" and "low" categories, the procedure minimized but did not eliminate the correlation between the treatment levels. Table 2 shows the correlation between the interval values of the variables shown in Table 1 for the 608 scenarios (19 arbitrators x 32 scenarios). Most of the correlations are less than .10 and the largest, between cost of living changes and the percentage increase in comparable police units, is .39. The cases were presented to each arbitrator in an identical but randomly selected order.

The experiment differed from the Bazerman and Farber (1985, 1986) studies in several important respects. First, the data were collected in face to face interviews by the second author, permitting greater control than is possible in a mail survey. Second, the use of a single occupation covered by statutory

TABLE 1
EXPERIMENTAL TREATMENT VARIABLES

<u>Variable</u>	<u>Treatment Values</u>	<u>Mean</u>	<u>SD</u>
Cost of living (%), previous year and coming year (COL)	2,3,4,5 7,8,9,10	6.063	2.675
% increase in comparable police units (CPERC)	1.5, 2.5, 3.5, 4.5 6.5, 7.5, 8.5, 9.5	5.438	2.735
Wage rate in comparable police units (\$/hour) (CWAGE) ¹	10.15, 10.25, 10.35, 10.45 10.65, 10.75, 10.85, 10.95	10.54	.280
% increase for firefighters in community (FINC)	1,2,3,4 6,7,8,9	5.375	2.673
Wage rate for firefighters in community (\$/hour) (FWAGE)	10.09, 10.16, 10.23, 10.30 10.59, 10.66, 10.73, 10.80	10.434	.261
% increase for non-uniformed bargaining units in the community (OTHWAG)	1.5, 2.5, 3.5, 4.5 6.5, 7.5, 8.5, 9.5	5.500	2.851
Community property tax rate (per \$100 of full value property) (TAX) ¹	.009, .010, .011, .012 .031, .032, .033, .034	.021	.011
Police Expenditures as a % of Gov't Operating Budget (EXP)	8,9,10,11 26,27,28,29	19.094	9.128
Crime rate/1000 residents (CRIME)	16,17,18,19 40,41,42,43	29.469	12.276

1. CWAGE and TAX were transformed in the empirical analysis so these variables would be similarly scaled across the field and experimental data. CWAGE was scaled as the bargaining unit wage divided by the wage rate in comparable police units. The mean of this new variable .95 (\$10/10.54) with a SD of .95. TAX was transformed to equal the untransformed tax rate by the mean tax rate for the experimental data, producing a variable with a mean of 1.00 and SD of 1.132.

TABLE 2

CORRELATION MATRIX OF THE EXPERIMENTAL DATA

	CWAGE	CPER	CRIME	COL	FINC	FWAGE	OTHWAG	EXP
TAX	0.115	0.041	0.072	0.000	0.044	0.020	-0.181	0.045
CWAGE		-0.026	0.026	0.020	0.082	0.069	-0.029	0.090
CPERC			-0.144	0.393	0.220	-0.005	0.083	-0.126
CRIME				-0.036	0.004	-0.005	-0.066	-0.005
COL					0.087	-0.024	0.173	-0.093
FINC						0.157	0.038	0.026
FWAGE							0.140	-0.066
OTHWAG								0.074
EXP								

decision criteria that were familiar to all the participants increased the realism of the design. Third, both the current wage (\$10/hour) and the offers of the parties (\$10.30 and 10.80/hour) were constant across all disputes.² Fourth, the subjects made only one award for each case and this was made under the final offer procedure of Wisconsin's police and firefighter interest arbitration law.

Following previous research, the arbitrator decision model was estimated using a standard probit model. Let W_A , the arbitrator's view of a fair percentage wage increase, be a linear function of the nine variables:

$$W_A = X\beta + \epsilon, \text{ where} \quad (1)$$

$$\epsilon \sim N(0, \sigma^2)$$

The probability that the union's offer is chosen is the probability that W_A is greater than or equal to the mean of the final offers (MOFFER), 5.5 percent:

$$\Pr(U=1) = 1 - \Phi((5.5 - X\beta)/\sigma) \quad (2)$$

where $\Phi(\cdot)$ is the CDF for the standard normal distribution. Previous studies have been able to estimate σ because of variation in the final offers of the parties. This is not possible in these data because the offers were identical in all the cases.

Columns one and three of Table 3 report the probit estimates using two model specifications. The first column reports the results for a model that includes the nine manipulated variables and column three reports results for a model that includes dummy variables for each arbitrator. The latter specification permits $E(W_A|X\beta)$ to vary across the arbitrators while

Table 3

Estimates of Arbitrator Decision Model For
The Experimental Data
(n=608)

Variable	(1)	(2)	(3)	(4)
Constant	13.362** (3.082)	7.990** (2.212)		
Community property tax rate (TAX)	-.258* (.121)	-.236* (.107)	-.266* (.124)	-.243* (.108)
Wage rate in comparable police units (CWAGE)	-11.114** (2.925)	-9.792** (2.363)	-11.503** (2.955)	-10.196** (2.404)
Wage increase in comparable police units (CPERC)	.218** (.026)	.169** (.023)	.228** (.027)	.175** (.023)
Crime rate/1000 residents (CRIME)	.778** (.150)		.808** (.156)	
Cost of Living (COL)	.145** (.025)	.145** (.023)	.152** (.026)	.151** (.024)
Wage increase for firefighters in the community (FINC)	-.036 (.025)		-.040 (.026)	
Wage rate for firefighters in the community (FWAGE)	-.012 (.266)		-.025 (.269)	
Increase in Other B.U. in community (OTHWAGE)	.034 (.022)		.035 (.023)	
Police expenditures (EXP)	.026** (.007)		.028** (.007)	
Arbitrator Dummies	No	No	Yes	Yes
-log L	310.406	334.379	297.832	324.552

Standard errors are in parentheses.

* $p < .05$, one tail

** $p < .01$, one tail

constraining the effect of the exogenous parameters to be the same for each arbitrator.

Six of the nine variables had a significant impact on decisions. The probability the union's offer was chosen was positively related to wage increases in comparable police bargaining units (CPERC), employee working conditions as measured by the crime rate (CRIME), the cost of living (COL) and the share of the community's budget allocated to police protection (EXP). The arbitrators were also sensitive to local property tax rates (TAX) and were less likely to select the union's offer when tax rates were high. The subjects also favored wage increases that compressed wages relative to comparable units. The negative coefficient on CWAGE shows the probability the union's offer was chosen declined if the city's wage level was high relative to comparable police bargaining units. The three variables that were not significant were wage and wage changes in other bargaining units within the local government. Arbitrators were not influenced by the firefighters' wage level, (FWAGE) the increase received by firefighters (FINC) or the increase agreed to between the city and its other unionized employees (OTHWAG).

The coefficients from the arbitrator dummies are not reported in Table 3 and as a set they were not significant at conventional levels using a likelihood ratio test ($\chi^2_{19} \text{ df} = 21.382, p = .316$). Moreover, as the third column of Table 3 shows, the parameters on the other variables changed only slightly when mean shifts in arbitrator preferences were permitted. Although the set of dummy variables were not significant, some of the differences between individual arbitrators were substantial. For example, one arbitrator was significantly more likely (.10 significance level) to select the union's offer when compared to 11 of the remaining 18 subjects. Figure 1 shows the

Figure 1
Distribution of Arbitrators in Experiment By % Union Wins

Stata™

distribution of arbitrators based on the percentage of the 32 cases that were decided for the union. Since all the subjects decided the identical set of cases, the variation in this figure reflects differences in their decision models.

The Field Data

Nineteen arbitrators that participated in the experiment were also decision-makers in teacher disputes under Wisconsin's Mediation-Arbitration statute. During the first 8.5 years of the statute (January 1, 1978 through the 1985-86 school year) complete data on offers and the hypothesized determinants of W_A were available for 289 cases of which 208 or 72 percent were decided by participants in the experiment.³

While the experimental and field data involved different occupations and laws, key features of the data were similar. Both were final offer by package schemes where the statutory criteria arbitrators were to consider when making a decision were nearly identical.⁴ The major difference between the laws was that under the Med-Arb law the individual appointed as the arbitrator was first required to mediate the dispute before initiating arbitration. Although this procedure may have influenced what disputes went to arbitration, it is unlikely to have had an effect on W_A different from the arbitration hearing alone.

The variables determining the arbitrators' views of an appropriate award included recent and expected changes in the cost of living, comparable wage levels, wage increases in other relevant bargaining units, and local property taxes. Changes in the cost of living were proxied by the average annual change in the cost of living index (CPI-All Urban Consumers) for the first twelve months of the arbitrated contract. Because of time delays in the

arbitration process, awards covering a school year were usually not made until most of the year was completed.⁵ Thus, arbitrators had nearly perfect information about changes in the cost of living for the first year of the contract when making a decision. Community fiscal support for education (i.e. willingness to pay) was measured using the school property tax rate in year $t-1$.

The athletic conference has become the comparison group used by arbitrators when evaluating the offers of the parties in teacher disputes because districts within an athletic conference are of comparable size and in close proximity to one another. A study of written awards in teacher disputes from 1978-1984 found arbitrators used athletic conference wages as a comparison when making an award in 71 percent of the cases (Halm, 1985). Therefore, two measures of comparable wage settlements were constructed based on the district's standing in its athletic conference. The first comparability measure was the district's position in the conference wage distribution for the preceding school year (CWAGE). This was computed as the district's wage divided by the conference mean.

The second comparability measure was the average percentage wage increase in the current year for districts in the conference that reached a settlement without resorting to the mediation-arbitration process (CPERC). Bargaining pairs that settled without petitioning for arbitration reached agreement before arbitration awards were made because of the time delays in the arbitration process. Therefore, using the mean settlement in districts that did not resort to mediation-arbitration avoids the potential simultaneity between negotiated awards and arbitrator decisions that is created when

parties petition for arbitration but eventually settle after observing what arbitrators have done within the athletic conference.

The first column of Table 4 reports the standardized probit results where each reported coefficient is B_i/σ_f and the coefficient on the mean offer of the parties is equal to $-1/\sigma_f$. The second column reports the unstandardized parameters which are interpreted as the effect of the independent variables on W_A .⁶ Two parameters are significant at the .05 level (one-tail). The district's wage position within the district's athletic conference (CWAGE) had a substantial impact on W_A . As in the experiment, arbitrators view of W_A depends on how the district's wage compared to wages in its group of comparable districts. If a district's wage was 5 percent above the athletic conference mean, W_A was 2.4 percentage points lower than a comparable district 5 percentage points below the conference mean. Wage settlements within the conference (CPERC) also had a significant effect on arbitrator beliefs about the appropriate award. The coefficient, .96, is significantly greater than zero and not different from one ($p = .540$), showing a one-to-one relationship exists between the average percentage increase in negotiated settlements within the athletic conference and W_A . The parameter on local property taxes (TAX) is negative but not significant at conventional levels. The effect of cost of living (COL) is positive but not different from zero ($p = .956$). This suggests the effect cost of living had only a modest direct effect on arbitrator beliefs over the 8.5 years.

Uncertainty about what is viewed as a fair award or the standard deviation of σ was 8.4 percent and statistically significant at the .05 level. This implies a substantial amount of uncertainty about what is viewed as a fair award even when the facts of the case and the weight given to the facts

Table 4

Estimates of Arbitrator Decision Model For
The Field Data
(n=208)

Variable	(1)	(2)	(3)
Constant	2.898 (2.047)	24.331 (15.385)	
Mean of the Parties' Offers (MOFFER)	-.119** (.042)		-.130** (.045)
Community property tax rate (TAX)	-.155 (.430)	-1.302 (3.741)	-.280 (.461)
Wage rate in comparable units (CWAGE)	-2.841 (1.816)	-23.853 (13.821)	-3.856* (1.981)
Wage increase in comparable units (CPERC)	.114* (.049)	.959* (.390)	.112* (.052)
Cost of Living (COL)	.046* (.026)	.389* (.228)	.035 (.031)
σ		8.396** (2.476)	
Arbitrator Dummies	No	No	Yes
-log L	137.238	137.238	130.633

Standard errors are in parentheses.

* $p < .05$, one tail

** $p < .01$, one tail

are known. The predicted fair wage increase for an average case was 8.23 percent, but the estimated error variance implies there was a 50 percent probability arbitrators believed the appropriate settlement was less than 2.56 or greater than 13.89.⁷

When the arbitrator dummies are added the fit of the model improves only slightly. The hypothesis that the appropriate settlement given the facts of the case is the same across these arbitrators cannot be rejected using a likelihood ratio test ($\chi^2_{19} = 13.208$, $p = .779$).⁸ Again, however, some of the differences between individual arbitrators were significant. Finally, the identity of the arbitrator had virtually no effect on the error variance. Uncertainty declined only slightly to 7.7 percent when the arbitrator dummies were added.

A Comparison of the Decisions in the Two Samples

The test used to evaluate decision model consistency across the two samples is based on the simple premise that if the decision models are comparable, the variables common to both data sets should have a similar impact on W_A . Note, however, that there are five variables in the experimental data (CRIME, FINC, FWAGE, OTHWAG and EXP) hypothesized to influence arbitrators in police disputes that are not relevant in teacher disputes. These variables, which measure other wage settlements within the government and the budget share devoted to police, have no counterpart in teacher bargaining. A review of written teacher awards by the authors show these factors are not explicitly raised by the parties or the arbitrator. Blue collar units in school districts tend to follow the settlement reached by

teachers and, since school districts in Wisconsin are independent government units, all of their budget is devoted to education.

Two strategies exist for dealing with the unique variables in the experimental data. First, the variables unique to the experiment could be included in the analysis and only the effect of variables common to the two designs could be constrained to be the same. The second strategy was to use only the variables common to both data sets and simply ignore information on the five criteria that were unique to the experiment. Although this second strategy ignores information available to the arbitrators in the experiment, the nearly orthogonal relationship between the criteria across the experimental scenarios ensures any bias introduced by excluding the five unique variables is slight. This is shown in the second and fourth columns of Table 3. The parameters on the remaining three variables change only slightly when the five variables are excluded. The virtue of the second strategy is that decision consistency in the two samples is easier to evaluate. This was the strategy pursued in this study.

If there was variation in the final offers in the experiment, testing whether the decision-making process was identical in the two designs is based on separate tests of the following two null hypothesis (where F and E denote, respectively, field and experimental parameters):

$$H_1: \beta_F = \beta_E$$

$$H_2: \sigma_F = \sigma_E$$

H_1 tests whether the variables common to the two data sets have an identical impact on W_A . Rejection of H_2 implies the uncertainty associated with

predicting arbitrator awards in the experiment equals the uncertainty associated with predicting decisions in actual cases. Without variation in the final offers H_1 and H_2 can only be jointly tested. If the joint hypothesis is rejected, it could either be because of differences across the two designs in the β , σ or both.

The estimates that constrain all of the common field and experimental equation parameters to be the same are reported in columns 1 and 2 of Table 5. Column 2 gives the estimates for a model that includes a dummy variable equal to "1" for the experimental data. To test H_1 and H_2 jointly, a likelihood ratio test was constructed using the sum of the values of the log likelihood functions of the unconstrained models (column 2 of Table 3 and column 1 of Table 4) and the values of the log likelihood function from the constrained model (columns 1 and 2 of Table 5). When the intercept is not allowed to vary across the two samples, the joint hypothesis defined by H_1 and H_2 is rejected by the data ($\chi^2_{d.f.} = 29.29$, $p < .0001$). The joint hypothesis continues to be rejected when the intercept is allowed to vary across the two samples ($\chi^2_{d.f.} = 13.535$, $p = .0188$).

Further Tests Comparing the Two Samples

Two strategies were pursued to further investigate the differences between the two samples. First, we test whether the relative weight placed on each of the "facts" within each data set were identical across the two data sets. Specifically, we tested the following restrictions:

$$H_3: \beta_{F,tax}/\beta_{F,cola} = \beta_{E,tax}/\beta_{E,cola}$$

$$H_4: \beta_{F,cwage}/\beta_{F,cola} = \beta_{E,cwage}/\beta_{E,cola}$$

$$H_5: \beta_{F,cperc}/\beta_{F,cola} = \beta_{E,cperc}/\beta_{E,cola}$$

Table 5

Estimates of Arbitrator Decision Model For
The Pooled Experimental and Field Data
(n=816)

Variable	(1)	(2)
Constant	1.764** (1.199)	5.581** (1.556)
Mean of the Parties' Offers (MOFFER)	-.049** (.016)	-.183** (.038)
Community property tax rate (TAX)	-.243** (.101)	-.245* (.102)
Wage rate in comparable units (CWAGE)	-2.781* (1.252)	-5.588** (1.452)
Wage increase in comparable units (CPERC)	.166** (.020)	.164** (.020)
Cost of Living (COL)	.081** (.016)	.101** (.017)
Experimental Dummy Variable (DESIGN)		-1.289** (.332)
Arbitrator Dummies	No	No
-log L	486.264	478.384
Standard errors are in parentheses.		
* p < .05, one tail		
** p < .01, one tail		

If we are unable to reject all three hypotheses then we can conclude that the weight placed on taxes, comparable wage increases and comparable wage levels relative to the weight placed on cost of living was the same in the field and experimental data. If we fail to reject these three hypothesis and if $\beta_{F, \text{cola}} = \beta_{E, \text{cola}}$, it would imply the weights are the same in the two samples and only σ s differ. However, since $\beta_{E, \text{cola}}$ isn't identified in the experimental data, $H_3 - H_5$ test only whether the ratio of the weights on the different facts are the same in the experimental and field data.

Table 6 reports the tests for $H_3 - H_5$. The first two columns report the ratio of the weight placed on taxes, comparable wage levels and comparable wage increases to the weight placed on inflation for each of the two samples. The last column reports the difference between the ratio of the weights for the two samples. This column shows the difference in the ratio are not statistically significant in the field and experimental data.

The rejection of H_1 and H_2 and the inability to reject $H_3 - H_5$ is consistent with the interpretation that the weights placed on the various facts are the same in the experiment and field data but the level of uncertainty about what the arbitrator will do is different. One reason the actual variances could be different between the two settings is because of the composition of cases in the field data and the final offer by package feature of the Wisconsin law. As Olson & Jarley (1991) report, many of the teacher final offer decisions that include wages also include other economic and non-economic issues in the package. If these other issues are given some weight in the decisions of arbitrators but are ignored by the researcher, the estimate of the error variance will be biased. Olson & Jarley (1991) confirm the importance of this

Table 6

The Differences in the Ratio of the Weights Placed on The
Facts In the Field and Experimental Data

	Field Data	Experimental Data	Field-Experimental
β_{tax}/β_{col}	-4.634 (10.584)	-1.631 (.779)	-3.004 (10.612)
$\beta_{c\text{wage}}/\beta_{col}$	-7.729 (6.411)	-6.776 (1.799)	-.952 (6.659)
$\beta_{c\text{perc}}/\beta_{col}$	2.967 (2.138)	1.172 (.261)	1.794 (2.153)

Standard errors are in parentheses.

distinction and obtain a much smaller estimate of σ when the sample of awards is restricted to those cases that involve only wages.

It is possible the inclusion of multi-issue cases in the field data in this study is responsible for the rejection of H_1 and H_2 because the experimental data involved only the single issue of wages. To test this possibility the sample of field cases was restricted to cases that involved only wages. This restriction reduced the sample of cases in the field data from 208 to 85 cases. The first column of Table 7 reports the estimates of the fair wage beliefs of the arbitrators for these 85 cases. The key point to note from this table is that the estimate of σ declined from 8.4 to 2.45 percent.

Columns 2 and 3 of Table 7 show the estimates from pooling the experimental data and the 85 field observations. The results in column 3 constrain the constant to be the same for the two data sets and column 2 includes an experimental dummy variable. The joint hypothesis described by H_1 and H_2 is tested by comparing the log likelihood values from columns 2 and 3 of Table 7 with the log likelihood values for the unconstrained model. The unconstrained value is simply the sum of the log likelihood values shown in column 1 of Table 7 and column 2 of Table 3. When a constant is not included in the constrained model, the hypothesis that the weights and σ are the same in the experiment and field data is still rejected using the wage only field disputes ($\chi^2_{6\text{ df}} = 29.4$, $p < .0001$). Very different results are obtained for the model reported in column 2. The constrained model fits the data almost as well as the unconstrained model when the constant term is allowed to vary between the experimental and field data. The χ^2 statistic comparing the

Table 7

Estimates of Arbitrator Decision Model For
The Wage Only Field Data and
the Experimental Data

Variable	Wage Only Field Data	Wage Only Field Data & Experimental Data	
Constant	5.984 (3.444)	10.016** (2.108)	3.543* (1.531)
Mean of the Parties' Offers (MOFFER)	-.408** (.144)	-.488** (.116)	-.039** (.024)
Community property tax rate (TAX)	.745 (.743)	-.223* (.105)	-.220* (.104)
Wage rate in comparable units (CWAGE)	-5.188* (3.033)	-8.100** (1.828)	-4.852** (1.620)
Wage increase in comparable units (CPERC)	.102 (.085)	.163** (.022)	.161** (.021)
Cost of Living (COL)	.159* (.074)	.145** (.022)	.112** (.020)
Experimental Dummy (DESIGN)		-3.595** (.849)	
Sample Size	85	693	693
Arbitrator Dummies	No	No	No
-log L	51.536	387.684	400.616

Standard errors are in parentheses.

* $p < .05$, one tail

** $p < .01$, one tail

constrained and unconstrained model is 3.54 and far less than conventional significance levels ($p = .6177$).

Table 8 shows how well the constrained model with a design dummy variable predicts decisions in the field data. The top portion of Table 8 compares the actual final offer decisions in the wage only field data with the predicted decisions using the estimates from only the field data (e.g., estimates in column 1 of Table 7). The probit model correctly predicts 67.1 percent of decisions (e.g., 57/85). The lower portion of Table 8 shows the same calculations using the estimates that constrain all of the coefficients except the constant to be the same in the two data sets (e.g., estimates in column 2 of Table 7). A slight decline in the percent of correct predicted decisions is expected from the constrained model because it cannot fit the data any better than the unconstrained predictions shown in the top half of the Table. As Table 8 shows, the percent of correct final offer decisions predicted from the constrained model declined very slightly to 63.5 percent. These results mean the weights arbitrators placed on the different facts and the level of uncertainty about which offer was to be chosen given the facts were the same in the experimental data and the wage only field decisions.

The negative and statistically significant coefficient on the experimental dummy variable means arbitrators thought the appropriate fair wage for police disputes was lower than a comparable teacher dispute. This finding is broadly consistent with the differences between police and firefighter and teacher arbitrations in the state of Wisconsin. For example, in 1986-87 the average arbitration award for law enforcement and firefighter personnel was 5 percent and 6.2 percent for teachers (WERC, undated). The lower average arbitration award for police and firefighters and the comparable

Table 8

Predicted and Actual Final Offer Selections in The Wage
Only Field Cases

Predictions From The Unconstrained Field
Estimates (Column 1 of Table 7)

Actual Decisions	No. Of Employer Victories	No. of Union Victories	Total
No. of Employer Victories	34	11	45
No. of Union Victories	17	23	40
Total	51	34	

Predictions From The Constrained Field and
Experimental Estimates (Column 2 of Table 7)

Actual Decisions	No. Of Employer Victories	No. of Union Victories	Total
No. of Employer Victories	31	14	45
No. of Union Victories	18	22	40
Total	49	36	

union win rates across the two sectors suggests the fair wage beliefs of Wisconsin arbitrators were higher for teachers. This suggests the experimental dummy is measuring real differences in arbitrator beliefs and not simply differences between the designs.

Summary and Conclusion

This paper compares the actual decisions made by a group of practicing arbitrators over an 8.5 year period with the decisions made in 32 simulated "paper and pencil" disputes that were decided in a couple of hours. Although the experimental study offers greater internal validity than is possible using field data, the differences in the designs are so substantial that one is left wondering whether the decision process used in an experiment accurately reflects the process used in real disputes. In true interest arbitration cases the parties present oral arguments and written exhibits over a half day or more that are designed to convince the arbitrator of the fairness of their offer. The arbitrator also knows that his acceptability to the parties in the future depends on the quality of his or her decisions.

Despite the substantial differences between the two designs, our results show a high degree of consistency in the decisions made across the two environments when the same decision is being made in the two environments. The data strongly reject the hypothesis of identical weights and uncertainty levels in the two environments when the single issue wage decisions made in the experiment are compared with multi-issue field decisions under final offer by package. However, this same hypothesis is not rejected when the experimental data is compared to the sample of field cases where the wage is the only issue.

Our findings suggest arbitrators make decisions in multi-issue disputes under a final offer package scheme using a model that is different from the single issue wage model used in this paper. Faced with a more complex decision involving multiple issues, arbitrators look to an expanded set of factors in reaching a whole package decision. As a result, the experimental data look different from the entire sample of field cases because the decision process is different in the multi-issue field cases and not because there was any inherent differences between the field and experimental environments. This result suggests great care must be taken when comparing "identical" field and experimental decisions. The observed differences in the taste and color of apples reported in an experiment and the taste and color of oranges reported by subjects in the field may have more to do with differences in the fruit than the differences between experimental and field environments!

Endnotes

1. These 22 arbitrators decided 82 percent of all public sector interest arbitration cases between 1979 and 1985. See Dell'Omo (1987, 1989) for additional detail about the design of this experiment.
2. This design feature complicates the comparison between the two data sets because the uncertainty around the arbitrators' preferred wage settlement is not identified without imposing an additional constraint on the model.
3. There were a total of 395 awards during this period (Babcock, 1987). The difference between this number and the 289 awards that were available for analysis in this study was caused by two factors. First, some of the cases did not involve wage disputes. In the remaining cases it was not possible to construct the final offers of the parties from the written awards.
4. The statutes covering nonuniformed, local government and police and fire employees requires that the arbitrator give weight to the following factors when making a decision: 1) legal authority of the employer, 2) stipulations by the parties, 3) welfare of the public, 4) comparisons with other public and private sector employees, 5) changes in the cost of living, 6) current wage levels and benefits, 6) changes in the preceding factors that occur during the arbitration process, 7) other factors traditionally considered by collective bargaining. See 111.70(4)(cm) and 111.77(cm)(b), Wisconsin Statutes.
5. A survey of experience under the Mediation-Arbitration Law for contracts

in effect in 1983 found that the average time from an initial petition to invoke the procedure to an arbitration award was greater than 320 days (Wisconsin Legislative Council Staff, 1985).

6. The likelihood function was $\prod_{\text{Union Wins}} (1 - \Phi((\text{MOFFER} - X\beta)/\sigma)) \prod_{\text{Employer Wins}} \Phi((\text{MOFFER} - X\beta)/\sigma)$. This model assumes the arbitrator picks the offer closest to W_A and is not biased toward one of the parties. See Ashenfelter and Bloom (1984) for a discussion of this point.
7. The reason for this high estimate of σ is discussed later in the paper. Also see Olson & Jarley (1991).
8. One arbitrator was dropped from this analysis because his intercept term was not identified. All of his decisions in the field data were for only one party.

REFERENCES

- Anderson, John C. 1981. "The Impact of Arbitration: A Methodological Assessment." Industrial Relations, Vol. 20, No. 2, pp. 129-147.
- Ashenfelter, Orley & David E. Bloom. 1984. "Models of Arbitrator Behavior." American Economic Review, Vol. 74, No. 1, pp. 111-124.
- Babcock, Linda. 1987. "The Role of Incomplete Information in Bargaining Under the Threat of Final Offer Arbitration: A Theoretical and Empirical Analysis." Unpublished paper, University of Wisconsin.
- Bazerman, Max H. 1985. "Norms of Distributive Justice in Interest Arbitration." Industrial and Labor Relations Review Vol. 38, No. 4, pp. 558-569.
- Bazerman, Max H. & Henry S. Farber. 1985. "Arbitrator Decision Making: When are Final Offers Important?" Industrial and Labor Relations Review Vol. 39, No. 1, pp. 76-89.
- Bloom, David E. 1986. "Empirical Models of Arbitrator Behavior Under Conventional Arbitration." Review of Economics and Statistics, Vol. 68, No. 4, pp. 578-585.
- Bloom, David E. & Christopher L. Cavanaugh. 1986. "An Analysis of the Selection of Arbitrators." American Economic Review, Vol. 76, No. 3, pp. 408-422.
- Dell'Omo, Gregory G. 1987. "Arbitration Application of Wage Criteria Under The Wisconsin Police and Firefighter Arbitration Act: An Empirical Analysis." Unpublished PhD dissertation, University of Wisconsin.
- Dell'Omo, Gregory G. 1989. "Wage Disputes in Interest Arbitration: Arbitrators Weigh the Criteria." The Arbitration Journal, Vol. 44, No. 2, pp. 4-13.
- Farber, Henry S. & Max H. Bazerman. 1986. "The General Basis of Arbitrator Behavior: An Empirical Analysis of Conventional and Final-Offer Arbitration." Econometrica, Vol. 54, No. 4, pp. 1503-1528.
- Farber, Henry S. & Harry C. Katz. 1979. "Interest Arbitration, Outcomes and the Incentive to Bargain." Industrial and Labor Relations Review Vol. 33, No. , pp. 55-63.
- Gordon, Michael, Neal Schmitt and Walter G. Schneider. 1984. "Laboratory Research on Bargaining and Negotiations: An Evaluation." Industrial Relations, Vol. 23, No. 2, pp. 218-233.
- Halm, Francis P. 1985. "Athletic Conference: The Question of Comparability." Wisconsin Public Employer Relations Association.

- Locke, Edwin A. 1986. "Generalizing From Laboratory to Field: Ecological Validity to Abstraction of Essential Elements?" Generalizing from Laboratory to Field Settings, Lexington Books, D.C. Heath: Lexington, Massachusetts, pp. 1-10.
- Olson, Craig A. 1988. "Dispute Resolution in the Public Sector." In B. Aaron, J. Najita, J. Stern, eds., Public Sector Bargaining. Bureau of National Affairs.
- Olson, Craig A. & Paul Jarley. 1991. "Arbitrator Decisions in Wisconsin Teacher Wage Disputes." Industrial and Labor Relations Review, Vol. 44, No. 3, pp. 536-547.
- Wisconsin Employment Relations Commission. Undated. "Digest and Report of Arbitration Awards Issued Under the Municipal Arbitration Law and the Municipal Interest Arbitration Law and Section 111.70(4)(jm) During Fiscal Year 1988-89." Madison, Wisconsin.
- Wisconsin Legislative Council. 1985. "Analyses of Employer and Employee Experience Under Wisconsin's Mediation-Arbitration Law." Madison, Wisconsin. October 15, 1985.