Can Subjects Meaningfully Learn Actual Voting Powers?: Pattern Recognition and Feedback Information (revised 01/20 in 2022)

K. Ogawa[1], Y. Osaki[2], T. Kawamura[3], H. Takahashi[4], S. Taguchi[5], Y. Fujii[6], N. Watanabe[7]

[1]Kansai U., [2]Waseda U., [3]Tezukayama U., [4]Hiroshima City U.
[5]Doshisha U. [6]Meiji U. [7]Keio U.

Jan. 22, 2022 at Waseda U.
24th Conference on Exp. Soc. Sci.

1. Introduction: Research Question

- **Our Concern**: Who can "meaningfully" learn the latent feature of weighted voting under what information provision?
  - Gelman et al. (2004): Empirical voting powers are different from the theoretical predictions (Shapley-Schubik, 1954; Banzhaf, 1965).
  - Felsenthal and Machover (1998): It would be difficult for people to see the underlying relationship between the actual voting powers and the nominal voting weights.

Why?: We should use weighted voting better, because it is a popular collective decision-making system.

1. Introduction: How to answer?

When people generalize what they have learned in a situation to a similar but different one, this higher order concept of learning is called meaningful learning (Rick and Weber, 2010).

- 2-armed bandit experiment: subjects choose one of two weighted voting games repeatedly and their payoffs were determined stochastically according to a payoff-generating function that was hidden from subjects (Guerci et al. 2017).
- subjects' ability for pattern recognition of payoffs: measured by Raven's APM test (Raven score).
- feedback information: Full-fb, Partial-fb, and No-fb of their payoffs immediately after their choice.

1. Introduction: Results

We affirmatively confirmed the following hypotheses.

- H1: Immediate feedback information about subjects' payoffs confused their inference on the relationship between nominal voting weights and actual payoffs so that they took the win-stay-lose-shift strategy (Nowak-Sigmund, 1993).
- H2: Withholding feed-back information promoted meaningful leaning at the sites where subjects had significantly higher Raven scores on average.
- H3: Subjects who have experienced easier binary choice problems in early periods for their learning process meaningfully learn the underlying structure of weighted voting.

Our experimental sites: Kansai U, Osaka SU, Doshisha, and Hiroshima CU.

2. Experimental Design: Binary Choice Problem

- Subjects choose one of two weighted voting games.
  - $[q; v_1, v_2, v_3, v_4]$; q is the quota, v_i is the voting weight.
  - Each subject acts as Member 1, who has v_1.
- clearly informed that the other three members are all fictitious.
- one binary choice problem for the first 40 periods, and a similar but different one in the following 20 periods; A $\rightarrow$ B, B $\rightarrow$ A, C $\rightarrow$ D, D $\rightarrow$C
- payoff generating function: Deagan-Packel index
  - The correct answer is Choice 2.

Table: Binary choice problems and expected payoffs for Member 1

Problem	Choice 1	(expected payoff)	Choice 2	(expected payoff)
A	$[14; \mathbf{5}, 3, 7, 7]$	$(120 \times 2/3 \times 1/3)$	$[14; \mathbf{5}, 4, 6, 7]$	$(120 \times 3/4 \times 1/3)$
B	$[6; \mathbf{1}, 2, 3, 4]$	$(120 \times 1/3 \times 1/3)$	$[6; \mathbf{1}, 1, 4, 4]$	$(120 \times 2/3 \times 1/3)$
C	$[14; \mathbf{3}, 5, 6, 8]$	$(120 \times 2/3 \times 1/3)$	$[14; \mathbf{3}, 6, 6, 7]$	$(120 \times 3/4 \times 1/3)$
D	$[9; \mathbf{1}, 3, 5, 6]$	$(120 \times 1/3 \times 1/3)$	$[9; \mathbf{1}, 2, 6, 6]$	$(120 \times 2/3 \times 1/3)$

- three feedback treatments: (1) no feedback (No-fb), (2) partial feedback (Part-fb), and (3) full feedback (Full-fb).
- a 30-second time limit for the choice stage, and a 10-second limit for the feedback stage
  - no choice within 30 seconds → zero point
  - "Wait" message is given in feedback stage for No-fb.
- 1 point is converted 1 JPY.
- zTree (Fischbacher, 2007) is used.

Subject's Monitor

Please choose one out of the following two committees (Choice 1 or Choice 2). Each committee decides a distribution of 120 points among four members. You are Member 1.

In both committees, 22 votes are apportioned to those members and you have 5 votes.

Any proposals of point distributions need 14 votes in favor to be adopted.

$$\text{Choice 1} \quad [14; \mathbf{5}, 3, 7, 7], \quad \text{Choice 2} \quad [14; \mathbf{5}, 4, 6, 7]$$

When subjects choose Choice 2 and MWC $(5, 6, 7)$ appears, they see, for instance, the following results on their monitor, regardless of any treatments.

You chose the following committee.

$$\text{Choice 2: } [14; \mathbf{5}, 4, 6, 7].$$

Next, in the full-feedback treatment, subjects see

The committee decided to distribute 120 points this time as follows.
You obtained 40 points this time.

$$(\mathbf{40}, 0, 40, 40)$$

on their monitors. In the partial-feedback treatment, the payoff distribution is not shown, but rather the following note is shown on their monitors:

You have obtained 40 points this time.

In the no-feedback treatment, the payoff distribution is not shown and simply

Please wait for a while.

is shown on the subjects' monitors.

- Raven test: measuring subjects' ability of pattern recognition, which fits our bandit experiment.
  - In each question of the test, eight patterns are drawn, and the subject selects a pattern that matches those visual patterns from the options.
  - Our subjects are asked to learn the pattern of payoffs
- We used 16 questions excerpted from the Raven's APM test.
  - The APM version is composed of 48 questions in total.
  - Subjects answered 16 questions within 10 minutes after completing the binary choice problems.
  - Colored Progressive Matrices (CPM), Standard Progressive Matrices (SPM), and Advanced Progressive Matrices (APM), in ascending order of difficulty.
    - Gill and Prowse (2016) and Basteck and Mantovani (2018) used the SPM version, while Proto et al. (2019) used the APM version. Guerci et al. (2017) used 16 questions of the Raven APM test.

2. Experimental Design: Session Information

- March 2, 2018 to October 17, 2019
- Kansai University, Osaka Sangyo University, Doshisha University, Hiroshima City University
- undergraduate students who do not know indices of voting power. In total, 816 subjects participated.
- the average amount paid as a reward was 2534 JPY (1 USD was about 110 JPY at that time).
- rewards paid for participation (not incentive payment) was different at Kansai U-A and Osaka SU (Group A) from those of Kansai U-B, Doshisha, and Hiroshima CU (Group B).
- The experimenter is the same for Group A.
  A text-to-speech software was used for Group B.

Table: subjects' attribute information (p-val for Fisher exact test)

site	# of subj.	male	female	p-val	econ	sci-eng	others
Kansai U-A	240	128	112		21	53	166
Osaka SU	120	99	21	$<$**0.001**	47	38	35
Kansai U-B	197	98	99		24	41	132
Doshisha	135	76	56	0.828	21	0	114
Hiroshima CU	124	49	75	**0.047**	0	24	100

Table: Raven scores of subjects: basic statistics

site	# of subj.	mean	std.dev.	p-value	min	max
Kansai U-A	240	11.208	2.170		3	16
Osaka SU	120	10.625	3.041	**0.027**	3	16
Kansai U-B	197	11.518	2.398		2	15
Doshisha	135	11.578	2.300	0.890	5	16
Hiroshima CU	124	10.976	2.441	**0.038**	3	15

Table: Raven scores of general public: basic statistics

test site	# of subj.	mean	std. dev.	min	max
EEL at Kansai U	1,057	7.986	3.326	0	16

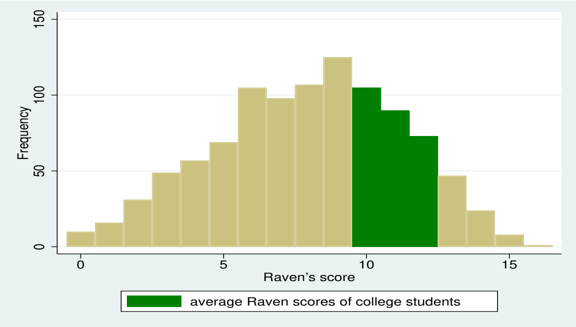

Figure: Histogram of Raven scores of non-student general public

The Raven scores of our subjects are, on average, significantly higher than those of non-student general public.

2. Definitions and Analysis

Let FR_k^i denote the relative frequency of periods in which subject i chose the correct answer within the k-th block of 5 consecutive periods, that is, from $5(k-1)+1$ to $5k$. The change in the relative frequencies that subject i chose the correct answer between the l-th block and the m-th block is defined as

$$\Delta\mathrm{FR}_{l,m}^i = \mathrm{FR}_l^i - \mathrm{FR}_m^i,$$

where $l > m$. Let $\Delta\mathrm{FR}_{l,m}$ denote a vector whose i-th component is $\Delta\mathrm{FR}_{l,m}^i$.

Definition 1: For each binary choice problem, we consider that subjects learned the correct answers, if (1) $\Delta\mathrm{FR}_{8,1} > 0$, (2) the WSLS strategy is not observed in the 8th block of 5 consecutive periods for the partial-feedback and full-feedback treatments, and (3) the rate of correct answers is at least 60% in the 8th block, regardless of feedback treatments.

Result 1: For the partial feedback treatment, subjects' learning was observed in Problem B at Kansai U-A and Osaka SU. For no-feedback treatment, subjects' learning was observed in Problems B and D at Kansai U-A.

Result 2: At Kansai U-B and Doshisha, no learning by subjects was observed, but at Hiroshima CU, it was observed in Problem B for the full-feedback treatment.

Table: *p*-values for one-tailed SR test, $\Delta\mathrm{FR}_{8,1}$, Group A

	Kansai U-A			Osaka SU		
	No-fb	Part-fb	Full-fb	No-fb	Part-fb	Full-fb
A	0.227	0.387	0.212	0.688	0.125	0.363
B	0.048*	0.001*	0.001	0.855	0.008*	0.500
C	0.500	0.212	0.227	0.813	0.500	0.109
D	0.038*	<0.001	0.059	0.227	0.500	0.172

Table: *p*-values for one-tailed SR test, $\Delta\mathrm{FR}_{8,1}$, Group B

	Kansai U-B			Doshisha			Hiroshima CU		
	No-fb	Part-fb	Full-fb	No-fb	Part-fb	Full-fb	No-fb	Part-fb	Full-fb
A	0.066	0.668	0.115	0.808	0.808	0.151	0.395	0.315	0.105
B	0.402	<0.001	0.007	0.090	<0.001	0.240	0.011	0.227	0.006*

The asterisk (and blue numerics) indicates that learning was observed.

Table: Frequency of changing alternatives in B8: Kansai U-A (two-sided Fisher exact test). n =20 for each Problem.

Part				Full			
		0 point	40 points			0 point	40 points
A	freq	28	72	A	freq	23	77
	switch	14	19		switch	12	13
	ratio	0.500	0.264		ratio	0.522	0.169
	p-value		0.033		p-value		0.010
B	freq	45	55	B	freq	47	53
	switch	13	10		switch	18	5
	ratio	0.289	0.182		ratio	0.383	0.094
	p-value		0.238		p-value		$<$0.001
C	freq	34	66	C	freq	24	76
	switch	14	21		switch	10	21
	ratio	0.412	0.318		ratio	0.417	0.276
	p-value		0.382		p-value		0.213
D	freq	37	62	D	freq	46	54
	switch	21	46		switch	21	13
	ratio	0.324	0.742		ratio	0.457	0.241
	p-value		$<$0.001		p-value		0.034

Table: Frequency of changing alternatives in B8: Osaka SU (two-sided Fisher exact test). For each binary choice problem, the number of subjects is ten ($n = 10$).

Part				Full			
		0 point	40 points			0 point	40 points
A	freq	18	32	A	freq	11	37
	switch	7	4		switch	5	7
	ratio	0.389	0.125		ratio	0.455	0.189
	p-value		0.041		p-value		0.113
B	freq	21	29	B	freq	18	32
	switch	6	6		switch	8	8
	ratio	0.286	0.207		ratio	0.444	0.250
	p-value		0.738		p-value		0.211
C	freq	13	37	C	freq	17	33
	switch	6	2		switch	8	13
	ratio	0.462	0.054		ratio	0.471	0.394
	p-value		0.002		p-value		0.764
D	freq	27	23	D	freq	17	33
	switch	12	9		switch	6	15
	ratio	0.444	0.391		ratio	0.353	0.455
	p-value		0.779		p-value		0.557

Table: Frequency of changing alternatives in B8: Kansai U-B (two-sided Fisher exact test). The number in the parentheses indicates the number of subjects.

Part				Full			
		0 point	40 points			0 point	40 points
A	freq	45	90	A	freq	46	99
(27)	switch	17	30	(29)	switch	21	24
	ratio	0.378	0.333		ratio	0.457	0.242
	p-value		0.702		p-value		0.012
B	freq	54	76	B	freq	87	123
(26)	switch	25	8	(43)	switch	46	22
	ratio	0.463	0.105		ratio	0.529	0.179
	p-value		<0.001		p-value		<0.001

Table: Frequency of changing alternatives in B8: Doshisha (two-sided Fisher exact test). The number in the parentheses indicates the number of subjects.

Part				Full			
		0 point	40 points			0 point	40 points
A	freq	37	103	A	freq	25	75
(28)	switch	11	38	(20)	switch	10	20
	ratio	0.297	0.369		ratio	0.400	0.267
	p-value		0.547		p-value		0.218
B	freq	41	59	B	freq	46	52
(20)	switch	17	12	(20)	switch	19	14
	ratio	0.415	0.203		ratio	0.413	0.269
	p-value		0.027		p-value		0.142

Table: Frequency of changing alternatives in B8: Hiroshima CU (two-sided Fisher exact test). The number in the parentheses indicates the number of subjects.

Par				Full			
		0 point	40 points			0 point	40 points
A	freq	3	85	A	freq	29	76
(23)	switch	11	34	(21)	switch	15	18
	ratio	0.367	0.400		ratio	0.517	0.237
	p-value		0.830		p-value		0.009
B	freq	48	49	B	freq	35	65
(20)	switch	22	13	(20)	switch	11	14
	ratio	0.458	0.265		ratio	0.314	0.215
	p-value		0.059		p-value		0.335

Table: Average rate of correct answer.

	FR_1	No-fb	Part-fb	Full-fb	FR_8	No-fb	Part-fb	Full-fb
Kansai U-A	A	0.530	0.630	0.680	A	0.590	0.660	0.760
	B	0.380	0.450	0.440	B	0.600	0.710	0.790
	C	0.690	0.530	0.560	C	0.730	0.640	0.650
	D	0.430	0.510	0.550	D	0.620	0.770	0.750
Osaka SU	A	0.700	0.720	0.580	A	0.660	0.820	0.660
	B	0.460	0.440	0.620	B	0.380	0.820	0.760
	C	0.660	0.700	0.600	C	0.680	0.740	0.720
	D	0.480	0.440	0.400	D	0.560	0.480	0.620
Kansai U-B	A	0.680	0.593	0.641	A	0.733	0.556	0.683
	B	0.452	0.492	0.493	B	0.519	0.800	0.707
Doshisha	A	0.622	0.614	0.670	A	0.585	0.536	0.750
	B	0.490	0.500	0.490	B	0.620	0.770	0.630
Hiroshima CU	A	0.560	0.626	0.571	A	0.570	0.617	0.667
	B	0.420	0.510	0.430	B	0.580	0.610	0.710

Definition 2: For each binary choice problem, we consider that subjects who have learned in the binary choice problem meaningfully learned the underlying structure of weighted voting games if (a) the elements of FR_9 of experienced subjects is, on average, significantly larger than those of FR_1 of inexperienced subjects, (b) the WSLS strategy is not observed in the 9th and 12th blocks for partial-feedback or full-feedback treatments, and (c) the rates of correct answers are at least 60% in the 9th and 12th blocks, regardless of feedback treatments.

- If subjects were confident in what they had learned up to the 40th period, they would not engage in the WSLS strategy in the 9th block.
- If the WSLS strategy was observed in the 12th block, then the conviction should have fluctuated.

Result 3: Meaningful leaning was not observed at Osaka SU for any binary choice problems in any treatments, while it was observed at Kansai U-A in Problem D for no-feedback treatment (average rages of correct answer is 0.60 in B9 and 0.66 in B12, respectively).

Result 4: Meaningful leaning was not observed at Kansai U-B, Doshisha, and Hiroshima CU for any binary choice problems in any treatments.

Table: Average rate of correct answer: Kansai U-A, Osaka SU

	Kansai U-A			Osaka SU		
Problem A	No-fb	Part-fb	Full-fb	No-fb	Part-fb	Full-fb
Inexperienced (FR_1)	0.530	0.630	0.680	0.700	0.720	0.580
Experienced (FR_9)	0.670	0.500	0.560	0.620	0.480	0.380
p-value	0.089	0.316	0.322	0.682	0.166	0.313
Experienced (FR_{12})	0.760	0.450	0.590	0.680	0.500	0.560
Problem B	No-fb	Part-fb	Full-fb	No-fb	Part-fb	Full-fb
Inexperienced (FR_1)	0.380	0.450	0.440	0.460	0.440	0.620
Experienced (FR_9)	0.500	0.540	0.490	0.360	0.440	0.500
p-value	0.243	0.260	0.466	0.328	0.971	0.323
Experienced (FR_{12})	0.620	0.770	0.620	0.440	0.740	0.660
Problem C	No-fb	Part-fb	Full-fb	No-fb	Part-fb	Full-fb
Inexperienced (FR_1)	0.690	0.530	0.560	0.660	0.700	0.600
Experienced (FR_9)	0.620	0.640	0.730	0.740	0.620	0.540
p-value	0.777	0.206	0.034	0.498	0.449	0.625
Experienced (FR_{12})	0.650	0.620	0.230	0.720	0.640	0.630
Problem D	No-fb	Part-fb	Full-fb	No-fb	Part-fb	Full-fb
Inexperienced (FR_1)	0.430	0.510	0.550	0.480	0.380	0.400
Experienced (FR_9)	0.620	0.610	0.510	0.200	0.440	0.600
p-value	0.031	0.182	0.560	0.024	0.456	0.103
Experienced (FR_{12})	0.660	0.670	0.720	0.280	0.560	0.580

Table: average rate of correct answer: Kansai U-B, Doshisha, Hiroshima CU

	Kansai U-B			Doshisha			Hiroshima CU		
A	No-fb	Part-fb	Full-fb	No-fb	Part-fb	Full-fb	No-fb	Part-fb	Full-fb
(FR_1)	0.680	0.593	0.641	0.622	0.614	0.670	0.560	0.626	0.571
(FR_9)	0.689	0.685	0.586	0.690	0.610	0.560	0.590	0.410	0.560
p-value	0.840	0.136	0.632	0.305	0.885	0.258	0.569	<0.001	0.891
(FR_{12})	0.634	0.810	0.676	0.690	0.600	0.640	0.650	0.550	0.670

B	No-fb	Part-fb	Full-fb	No-fb	Part-fb	Full-fb	No-fb	Part-fb	Full-fb
(FR_1)	0.452	0.492	0.493	0.490	0.500	0.490	0.420	0.510	0.430
(FR_9)	0.538	0.563	0.469	0.541	0.636	0.460	0.520	0.548	0.571
p-value	0.231	0.270	0.761	0.581	0.039	0.517	0.235	0.475	0.069
(FR_{12})	0.501	0.732	0.791	0.541	0.836	0.640	0.620	0.739	0.676

- Meaningful learning in a weaker sense (without (2) and (3) for learning, and without (b) and (c) for meaningful learning) was observed at Kansai U-A, Doshisha as well as Osaka University and the University of Tsukuba (Guerci et al., 2017; Watanabe 2018).
- The Raven scores of the subjects at Osaka U and Tsukuba (12.3 and 12.5) are, on average, significantly higher than those of subjects at Kansai U-A., Kansai U-B, Doshisha, OSU, and HCU (11.2, 11.5, 11.5, 10.6, and 10.9, respectively).

Result 5: In this experiment, meaningful leaning in a weaker sense of the underlying structure of weighted voting was observed at schools where subjects' ability for pattern recognition of their payoffs is relatively high.

Subjects' ability for pattern recognition is related to their ability for meaningful learning in weighted voting.

3. Conclusion

All the following hypotheses were affirmatively confirmed.

Hypothesis 1: Immediate payoff-related feedback information confuses subjects and induces them to chose the WSLS strategy.

Hypothesis 2: Meaningful learning of the underlying structure of weighted voting is observed at experimental sites where subjects have, on average, relatively higher ability for pattern recognition of their payoffs.

- Hypotheses 1 and 2 were affirmatively confirmed. $\Rightarrow$ It would be difficult for non-student general public to infer the underlying structure of weighted voting.
- Then, how should we provided them with information in their learning process? As noted in the Introduction, weighted voting is a popular collective decision-making system. We should use it better.

Problems A and C are easy: a committee has two "large" voters who can form an MWC by themselves, while the other does not. Problems B and D are difficult: there is no such a clear difference between the two committees.

- Result 3: Meaningful learning was observed in Problem D.
- We applied the runs test to Problems C and D and counted the number of choosing wrong options in those binary choice problems.

Hypothesis 3: Subjects who have experienced easy binary choice problems in early periods meaningfully learn the underlying structure of weighted voting, but they fail to meaningfully learn it when they have experienced difficult binary choice problems in those early periods.