J. Econ. Theory 216 (2024) 105791

Contents lists available at ScienceDirect

Journal of Economic Theory

journal homepage: www.elsevier.com/locate/jet

Making predictions based on data: Holistic and atomistic procedures ☆

Jacob Glazer[a], Ariel Rubinstein[b,*]

[a] Coller Faculty of Management, Tel Aviv University and Department of Economics, The University of Warwick, United Kingdom of Great Britain and Northern Ireland

[b] School of Economics, Tel Aviv University and Department of Economics, New York University, United States of America

ARTICLE INFO

JEL classification:
D01

Keywords:
Experimental economics
Non-Bayesian beliefs
Data-based predictions

ABSTRACT

Subjects were asked to predict the choice made by a hypothetical individual, based on a small sample of his past choices in identical situations. Each of the alternatives facing the individual has a number of components. When the data is presented explicitly as a distribution of past choices, most subjects use the *holistic* procedure according to which the prediction is the most common choice made in the past. Nevertheless, a significant minority of subjects use *atomistic* procedures, which relate to the various components individually. When the distribution is not explicitly presented but can be derived from the data, subjects apply both atomistic procedures and holistic rules of thumb.

1. Introduction

In order to slide into the paper, consider the following question:

W1: Imagine an individual (simulated by a computer) who starts each morning by surfing the web. The individual visits each of the three websites A, B and C exactly once. There are links from each site to the other two. The order in which the individual chooses to visit the three sites can vary. The following table presents the number of days on which each of the routes was chosen during the last 21 days:

	W1
ABC	0
ACB	0
BAC	6
BCA	7
CAB	8
CBA	0

What is your best guess as to the individual's route on the 22nd day?

☆ We are grateful to Aron Tobias for a thorough review of the paper and to Ayala Arad for her advice. We also thank Michael Crystal, Michael Eldar, Kfir Eliaz, Ignacio Esponda, Gerd Gigerenzer, Ori Heffetz, Amnon Maltz, Ayush Pant, Dotan Persitz, Kirill Pogorelskiy, Andrew Schotter, Emanuel Vespa and Xueying Zhao for their comments.

* Corresponding author.
E-mail address: rariel@tauex.tau.ac.il (A. Rubinstein).

https://doi.org/10.1016/j.jet.2023.105791
Received 28 May 2023; final version received 28 December 2023; Accepted 29 December 2023

Available online 9 January 2024

In the question above, you are being asked to make a *point prediction* about the future action of another individual and the only information you are given is some statistics about a *small* number of relevant observations of that individual's past behavior. Such situations appear to be common in real life. For example:
1) Defending against a penalty kick in soccer is a difficult task. A goalkeeper tries to predict the characteristics of the kick based on the penalty kicker's previous kicks, including its direction (left, center or right) and its height (high or low). He can predict the most frequently observed combination of characteristics or he can predict each characteristic separately.
2) A paparachi wishes to predict the route of a celebrity he is following from one night club to another, based on the celebrity's past routes. He can either adopt a holistic view, and predict the most frequent route taken by the celebrity in the past, or he can adopt an atomistic view, start from the night club the celebrity visited first most frequently in the past.
3) A manager periodically evaluates his employee. On each occasion, he examines the employee's performance in two out of a number of possible categories. The employee has information about the categories his boss has chosen in the past and tries to predict the two he will use in the next evaluation. His prediction can be the most frequently paired categories in the past, or the two categories most frequently chosen by the manager in the past.

The situations studied here have the same structure as W1 above: Individual A wishes to predict individual B's choice. A only observes some data about B's past choices in the same situation. A succeeds if his prediction is correct and fails if it is not.

The conventional approach to analyse such a situation is to assume that A has in mind a probability belief about B's behavior and predicts the most likely alternative B will choose given the data. However, it seems that in the real world such a procedure is rarely used to predict behavior.[1]

The novelty of our study is that it attempts to understand the procedures people actually use in such a situation based on a number of experiments. Each of them begins with a short description of a choice problem, followed by information on an individual's past choices in identical situations. The subjects were then asked to provide their "best guess" of the individual's next choice. We confined ourselves to binary scenarios in which the predictor can only succeed or fail. This allows us to avoid becoming entangled in theories of decision making under uncertainty.

Each alternative has a number of components. This enables us to distinguish between *holistic* procedures in which a subject views each alternative as a whole and *atomistic* procedures in which he considers each component separately, whether sequentially or simultaneously.

Thus, in W1, using the most frequently chosen route (CAB) as the prediction is the prime example of a holistic procedure. Starting the route from the site most commonly chosen first (B) and continuing from there to the most frequently chosen site following B (resulting in the prediction BCA) is a prime example of an atomistic procedure.

Note that neither W1 nor any other of the experiments provides any details about the process by which the individual makes his choices. The holistic procedure of choosing the most frequently chosen sequence in the past might reflect the assumption that the individual follows a probabilistic process consisting of identical independent randomizations. An atomistic procedure might reflect an assumption that the individual follows some probabilistic process in which he randomly draws the first component of an alternative and then applies another conditional independent randomization to complete the predicted route. In Section 4, we briefly comment on such rationales for different choice procedures.

In all versions of the experiment, after submitting his prediction, the subject was asked to provide an explanation for his prediction. The explanation was used to classify the procedure he used. The results are reported after excluding subjects who: (i) did not provide an explanation; (ii) provided an explanation for a choice other than the one he actually made; or (iii) provided a nonsensical explanation (such as "bla, bla, bla"). The classification was carried out independently by each of us and by a research assistant and the few disagreements were than discussed and resolved.

In Section 2, we report on the main set of experiments. In addition to W1, the set contains three other experiments, each of which differs from W1 in one aspect. In W2, we slightly change the distribution of past choices. In W3, we change the background story so that instead of choosing a route the individual is choosing a pair of flower types. In W4, we return to the choice of routes except that the distribution of past choices is given only implicitly rather than explicitly.

The experiments were carried out at the Centre for Behavioral and Experimental Social Science at the University of East Anglia (UEA) and at the Center for Experimental Social Science at New York University.[2] Students who were previously registered with the labs were invited to participate in a short online experiment and each participant was assigned randomly to one of the experiments. Subjects were incentivized to give the right answer (the one predicted by a computer simulation).[3] A subject who gave the right answer received $10 at NYU and a chance of one out of ten to receive £20 at UEA.[4]

[1] This impression is confirmed by our experimental findings: there was no case in which a subject explained his answer to a question like W1 as being the outcome of updating his subjective beliefs.

[2] Prior to the reported experiments, we conducted a series of non-incentivized experiments by means of the site gametheory.tau.ac.il, which is used for pedagogical purposes in game theory courses in numerous countries. The results are reported in https://arielrubinstein.tau.ac.il/papers/GR2021OLD.pdf and the interested reader can verify that there are no significant differences between those results and the ones reported here. For example, there are no significant difference between the choices of BAC, BCA and CAB in W1 and those made in the prelimenary experiments (question V2 in the old version): $\chi^2(2, N = 166) = 1.706$, $p = .426$. The use of Holistic and Atomistic procedures also do not differ: $\chi^2(1, N = 146) = 1.232$, $p = .267$.

[3] Each experiment started with the statement: "Imagine an individual (simulated by a computer) who…".

[4] For example, at UEA, each version of the experiment ended with the statement: "One out of every ten participants in this experiment will be randomly selected and awarded £20 if his or her answer is correct.".

Table 1
W1 and W2 – data and results.

	W1		W2	
	Data	Results	Data	Results
ABC	0	0% [0:]⟨0⟩	0	2% [2:]⟨0⟩
ACB	0	1% [1:1T^{min}]⟨3⟩	0	0% [0:]⟨1⟩
BAC	6	26% [22: 13T^{max}, 4H^{min}]⟨5⟩	5	24% [24: 11T^{max}, 9H^{min}]⟨1⟩
BCA	7	21% [18: 12T^{max}]⟨6⟩	6	18% [18: 13T^{max}]⟨1⟩
CAB	8	48% [41: 41H^{max}]⟨4⟩	8	54% [55: 51H^{max}]⟨6⟩
CBA	0	5% [4:]⟨2⟩	0	3% [3:]⟨3⟩
n =		[86]⟨20 = 19%⟩		[102]⟨12 = 12%⟩

Following are the main insights gained from the results:
(i) Given *explicit* statistics about past choices, most people predict that in a similar situation the individual will make the choice most frequently made in the past, although a significant minority of subjects use atomistic procedures. This is confirmed in W1, W2 and W3. Note that our goal is *not* to estimate the "actual" distribution of the procedures, but only to identify the most commonly used procedures.
(ii) A comparison of W1 to W2 indicates that the results are robust to small modifications of the distribution of past choices.
(iii) The background story of W3 (choosing a combination of flowers rather than a route of websites) seems to trigger a more holistic approach. Indeed, holistic procedures are more commonly used in W3.
(iv) A comparison of W1 to W4 indicates that when the statistics on past choices are only *implicitly* provided, people often apply various rules of thumb, which are usually atomistic.[5]

In Section 3 we report on an additional set of three (non-incentivized) experiments in which routes among websites are chosen. In one of them, the data on past choices is presented explicitely while in the other two they are presented only implicitly. In this set of experiments, insights (i) and (iv) are recomfirmed.

In Section 4 we comment on the probabilistic rationale underlying atomistic procedures used in W1 and on the rates of success of some of the rules of thumb observed in W4.

We believe that our results can motivate the construction of economic models in which agents make such predictions. The finding that individuals tend to use systematic non-standard procedures may lead to theoretical models in which Bayesian agents are replaced by more realistic types of agents. Section 5 presents a simple model of equilibrium in which individuals apply a variety of procedures - of the types analysed in this paper - to predict the actions of others.

2. The main set of experiments

In this section, we report on the main set of (incentivized) experiments.

2.1. Predicting the route between websites based on an explicit distribution (W1 and W2)

Returning to W1, an alternative is a permutation of the three websites. A holistic procedure treats each alternative as a whole, while an atomistic procedure starts by predicting a particular website (usually the first one) and then uses some rule to complete the predicted route.

The difference between holistic and atomistic approaches is highlighted in this experiment. The most popular holistic procedure is H^{max} whose prediction is the most frequently observed route (CAB). The most popular atomistic procedure is T^{max} which predicts a route starting from the site that appears most frequently as the first one (B) and continuing on to either C (since conditional on B being the first site, C is the most likely second site[6]) or A (since A is the most frequently chosen second site[7]).

A small number of subjects followed the holistic procedure H^{min} whose prediction is the least frequent alternative among those chosen at least once in the past (BAC[8]). They probably believe that people tend to diversify their choices or alternatively they fell victim to the "gambler's fallacy" (Tversky and Kahneman (1971)). Atomistic procedures that activate a minimal (rather than maximal) criterion for choosing the first site in the sequence were not observed.

In order to check the robustness of the findings, we ran a similar experiment (W2) in which there is a larger gap between the most frequently made choice and the other two.

Table 1 presents the results for W1 and W2. In the last row of the table, the number of subjects who gave a valid answer appears in [...] while the number and proportion of omitted subjects appear in ⟨...⟩. In the "Results" columns the percentage indicates the proportion of the non-excluded subjects who chose the alternative corresponding to that row. The first element in [...] is the number

[5] For a survey of decision-making heuristics, see Gigerenzer and Gaissmaier (2011).
[6] BCA: "The person goes to the B website first on 13 days. When they go to a B website they are more likely (only slightly though) to go to C website afterwards.".
[7] BAC: "She went on B first on 13 days/21 so B first is the most likely option for her. And then A Second is most likely as she went on A Second 14/21 days. BAC is the only option where she went with the first two most likely options so I picked that".
[8] BAC:"Out of the ones they have chosen before, this has been chosen only 6 times so would go there next time.".

Table 2
W3 - Results.

	Data	W3-Results
GO	0	0% [0]
GP	6	47% [48: 28T^{max}, 17H^{min}]⟨3⟩
GR	7	12% [12:6H]⟨2⟩
OP	8	39% [40:40H^{max}]⟨0⟩
OR	0	2% [2]⟨4⟩
PR	0	0% [0]
n=		[102]⟨9=8%⟩

of subjects who gave the answer corresponding to that entry and it is followed by a list of the *main* procedures used to justify the choice, together with the numbers of subjects who applied each of them. Finally, the number of subjects excluded in each entry appears in ⟨...⟩. For example, the entry corresponding to the choice of BAC in W1 indicates that 26% of the subjects chose BAC in W1; that this group consisted of 22 subjects, of whom 13 used T^{max} and 4 used H^{min}; and that 5 answers of BAC were omitted.

The results in W1 and W2 are quite similar both in terms of the distribution of predictions[9] and the use of holistic and atomistic procedures.[10] About half of the subjects used H^{max}, about 7% used H^{min}, and one third used atomistic procedures. Response time is useful in interpreting the prediction made and, in particular, in distinguishing between instinctive and contemplative responses (see, for example, Rubinstein (2007, 2013)).[11] However, given the relatively small number of subjects, deriving conclusions from response time has limited power. Nevertheless, note that in W1 there is a large difference in median response time between the subjects who chose BAC or BCA (155 seconds) and those who chose CAB (75 seconds). This is likely due to the fact that the execution of the atomistic T^{max} procedure requires more effort than the holistic H^{max} procedure.

2.2. Predicting the composition of a Bouquet of flowers based on an explicit distribution (W3)

W3 differs from W1 only in that it has a different background story. Rather than choosing the order in which three websites are visited, the two colors in a bouquet are chosen. In both scenarios, the distribution of past choices includes six options, three of them never made and three others were made 8, 7 and 6 times, respectively. Furthermore, the natural atomistic procedures in both scenarios lead to the choices of the less frequently chosen option among the three.

W3: A flower store sells 4 types of flowers: G, O, P and R. Imagine a regular customer (simulated by a computer) who comes to the store every Friday and buys a bouquet consisting of two types of flowers. The following table specifies how many times the customer bought each pair of flowers during the last 21 weeks:

	W3
GO	0
GP	6
GR	7
OP	8
OR	0
PR	0

What is your best guess as to the pair of flowers the customer will choose next Friday?

We again observe the use of two main holistic procedures (see Table 2). The more popular of the two is H^{max} in which the most frequently observed combination of flowers (OP) is predicted. A significant proportion of subjects followed H^{min} in which the least frequently chosen combination from among those that were observed at least once in the past (GP[12]) is predicted. A few subjects justified the choice of GR by means of holistic procedures.[13] The only atomistic procedure observed was T^{max} in which the bouquet consisting of the two most frequently chosen colors (GP[14]) is predicted.

In this context, minimization might be an expression of diversification. Thus, the fact that H^{min} is observed while T^{min} (i.e. choosing the bouquet consisting of the two least frequently chosen colors) is not may be related to Simonson (1990)'s claim that "people tend to choose more diversity when the choices are bracketed broadly than when they are bracketed narrowly."

Our initial conjecture was that the bouquet story nudges subjects toward holistic procedures since combining colors has value in itself. Indeed, the distributions of predictions in W1 and W3 differ significantly.[15] However, our conjecture was not confirmed: in

[9] The chi-square statistic is χ^2 (2, N = 178) = 0.696, p = .706.
[10] The chi-square statistic is χ^2 (1, N = 155) = 0.785, p = .376.
[11] The time between when a subject received the question and when he submitted his answer.
[12] GP: "He has done GP the least out of the 21 times so he will alternate back to it again".
[13] GR: "Unlikey that till will purchase a combination that they have never purchased and GR seems like a midpoint choice".
[14] GP: "The customer bought G and P the most from the 21 days, so they are the most likely".
[15] The chi-square statistic is $\chi^2(2, N = 178) = 8.974, p = .011$.

Table 3
W4 compared to W1.

	W1	W4-Results
ABC	0	1% [1:] ⟨1⟩
ACB	1%	0% [0 :] ⟨1⟩
BAC	26%	70% [69: $59T^{max}$, $2H^{sum}$, $3H$] ⟨5⟩
BCA	21%	8% [8:$6T$] ⟨4⟩
CAB	48%	16% [16: $10T$, $2H^{max}$] ⟨5⟩
CBA	5%	5% [5 :] ⟨0⟩
n =	[86]	[99] ⟨16 = 14%⟩

both W1 and W3, around two thirds of the subjects who gave a clear explanation mentioned a holistic procedure while one third mentioned an atomistic one.[16]

2.3. Predicting the route between websites based on an implicit distribution (W4)

The background story of this experiment is the same as that of W1, except that rather than providing the distribution of routes chosen in the past, the subjects were given a table with the number of times that each site was visited first, second and third. Specifically, the following data were provided:

W4: The following table presents the number of days on which the individual visited each site first, second and third during the last 21 days.

	1st	2nd	3rd
A	0	14	7
B	13	0	8
C	8	7	6

What is your best guess as to the individual's route on the 22^{nd} day?

Such a table has only 4 degrees of freedom (since each row and each column must add up to the total number of observations) in contrast to the 5 degrees of freedom in the set of choice distributions. Therefore, such a table often corresponds to more than one distribution of choices. However, in the case of W4, the unique distribution of choices which is consistent with the table is the one presented in W1.[17]

Table 3 presents the results. Although the data in W4 is uniquely consistent with the distribution of choices in W1, the difference between the results is striking: 70% of the subjects chose BAC in W4 as opposed to only 26% in W1.[18] The difference in the distributions of holistic and atomistic procedures is also striking: in W4, almost all the subjects used atomistic procedures while in W1 two thirds of the subjects used holistic procedures.[19] The most popular procedure is the atomistic T^{max} in which the first site is predicted to be the one visited first most often in the past and the second is predicted to be the one visited second most often between the remaining two sites (BAC[20]). Only a few subjects gave a clear holistic explanation: Two (out of 99) subjects[21] followed H^{max}, that is, they correctly derived the underlying distribution of choices and chose CAB. Two subjects followed H^{sum} and predicted the path (a_1, a_2, a_3) with the largest sum of numbers in the three entries $(a_1, 1)$, $(a_2, 2)$ and $(a_3, 3)$ (BAC).

In contrast to the experiments with an explicit distribution of choices, in which about 10% applied H^{min}, none of the subjects in this experiment applied that procedure. This raises the possibility that when the data is presented implicitly, subjects tend to focus on processing it rather than on the diversification of choices that motivates H^{min}.

3. The graph experiments (Z1, Z2, Z3)

In this section, we report on an additional set of experiments which were conducted only by means of the pedagogical site gametheory.tau.ac.il and without monetary incentives. The subjects were all students who had taken or were taking a game theory course. No money was awarded for the right answer, although a few subjects were drawn randomly to receive $40 as a reward for participation (regardless of their answers).

The background story in the graph experiments is similar to that of W1 (i.e. an individual chooses a route for visiting sites on the web) except that:
(a) the set of feasible routes is restricted to a given directed graph; and

16 The chi-square statistic is $\chi^2(1, N = 161) = 0.438, p = .508$.

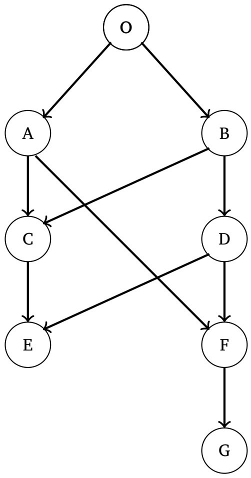
17 To see this without solving a set of equations, notice that whenever the individual started from C he must have continued to A and thus CAB occurs 8 times. Therefore, when the individual started from B he must have continued to A on 6 occasions and to C on 7 occasions.

18 There is a significant difference between the distributions of conjectures: The chi-square statistic is $\chi^2(2, N = 174) = 56.585, p < 0.0001$.

19 The chi-square statistic is $\chi^2(2, N = 174) = 38.441, p < 0.0001$.

20 BAC: "B has the highest likelihood to be the 1st, so I put B in first, between A and C, A has the higher probability to be 2 nd, so order is BAC.".

21 For the purpose of comparison, 7 out of 77 subjects in the pilot experiment concluded analytically that CAB is the most frequently chosen route in the past.



ACE	4
AFG	13
BCE	5
BDE	6
BDFG	1

Fig. 1. Z1 – data.

(b) the subjects were also provided with partial information about the individual's route on the day for which they were asked to make the prediction.

The three versions in this set of experiments start with the text:
"An individual surfs the web every day. He always starts from O and finishes at either the website E or the website G. A link from one website to another is indicated in the graph by an arrow. Last month the individual was tracked for 29 days." This background story was followed by a graph presenting the feasibility constraints, together with data about the individual's last 29 choices. The subject was then told: "Today it was observed that the individual ended up at E. What is your guess as to the individual's route today?"

As in the other experiments, subjects who did not provide an explanation of their prediction, who mistakenly justified a different prediction or who provided an explanation that made no sense were excluded. Here, we also excluded the 10% of subjects who chose a path that does not end at E.

3.1. Explicit distribution of past choices

In the basic version of this set of experiments, the data consisted of the explicit distribution of past choices (see Fig. 1).

Almost all of the subjects (80%) chose BDE and most justified their choice using H^{max}. This is unsurprising given that in Z1 there is no tension between the holistic procedures and the atomistic procedures (unlike in W1). Only 2 subjects used H^{min} (ACE) and very few gave atomistic explanations.

3.2. Implicit distributions of past choices

In Z2 and Z3, the data was provided implicitly in two different ways, each of them consistent only with the distribution of choices in Z1 (see Fig. 2). In Z2, subjects were provided with the number of times the individual used each link together with the explanation: "The number next to an arrow is the number of times the person used the corresponding link." In Z3, subjects were provided with the number of times the individual visited each of the sites, along with the explanation: "The number next to each website is the number of times the individual visited this website on his 29 daily routes."

In Z2, two holistic and three atomistic procedures were identified:
H^{max}: Find the underlying distribution of choices and choose the most frequent path ending at E (BDE).
H^{sum}: Choose the path ending at E with the maximal number of transitions (ACE[22]).
T^{max}: Proceed from the origin by choosing the mst frequently used link, as long as that does not rule out ending at E (ACE[23]).
T^{rep}: Choose the "representative" first node, that is, the one which, according to the data, leads to E most frequently (BCE or BDE[24]).
T^{bi}: Work backwards from E by choosing the most frequently used link (BCE[25]).

[22] ACE:"If you sum the amount of times he used each website in every row OACE-30, OBCE-26,OBDE-25, so I thought it was more likely that the route would be OACE.".

[23] ACE: "He chose A more days so the chance he will do it again is higher and from A it is the only way.".

[24] BDE: "Out of 12 times he reaches to B he got 11 times to E, and just 4 times out of 17 he got from A to E. Then 7 times to D (instead of 5 times to C) and then E. better odds.".

[25] BCE: "Out of the 15 times ending up at E, 9 times he got there by C and 6 times by D. Out of 9 times ending up at C, 5 times he got there by B, 4 times by A. Therefore the most probable route was that from O he went to B, then C and then E.".

Fig. 2. Z2 (left) and Z3 (right) - data.

Table 4
Z – results.

	Data	Z1 -distribution	Z2 - transitions	Z3 - visits
ACE	4	10% [8]⟨6⟩	25% [22: 4H^{sum}, 13T^{max}]⟨8⟩	54% [57: 15H^{sum}, 32T^{max}, 4T^{bi}]⟨33⟩
AFG	13	0% [0]⟨12⟩	0% ⟨12⟩	0% ⟨11⟩
BCE	5	10% [8]⟨13⟩	14% [12: 6T^{bi}, 3T^{rep}]⟨11⟩	17% [18: 12T^{rep}, 3T^{bi}]⟨18⟩
BDE	6	79% [61]⟨6⟩	61% [54: 32H^{max}, 18T^{rep}]⟨23⟩	29% [30: 25H^{max}, 3T^{rep}]⟨10⟩
BDFG	1	0% [0]⟨2⟩	0% ⟨4⟩	0% ⟨5⟩
n=		[77]⟨39 =34%⟩	[88]⟨58=40%⟩	[105]⟨77=42%⟩

In Z3, we observed procedures analogous to those in Z2: holistic – H^{max} (BDE) and H^{sum} which predicts the path ending at E with the largest sum of visits (ACE[26]); atomistic – T^{max} (ACE[27]), T^{rep} (BCE[28]) and T^{bi} (ACE[29]).

The results are presented in Table 4. While in Z1 almost all subjects used H^{max}, a majority of subjects used atomistic procedures in Z2 (56%) and in Z3 (57%). This finding is in line with the results reported in Section 2 regarding the difference in responses between W1 and W4.

4. Evaluating the procedures

4.1. On the probabilistic rationale underlying atomistic procedures

One way to explain the use of a procedure in W1 is by assuming that the subject believes that the data is a realization of some probabilistic process whose parameters are unknown to him. Given the data, he estimates the parameters using a maximum likelihood calculation. He then chooses the most likely sequence given the estimated parameters of the stochastic process. Of course, different probabilistic models may yield different predictions.

According to one probabilistic model, there exists a probability measure over the set of (holistic) alternatives, and the data is a realization of a number of independent draws. The maximum likelihood calculation estimates the probability of each alternative to be equal to its frequency in the data.[30] Given this model, the rational prediction is the most frequent sequence in the data.

According to a different model, the data is created by a process built on μ, a probability measure over the set of sites $\{A, B, C\}$. The first element in the sequence is the realization of μ while the second is the realization of μ conditional on the two remaining elements. Under this probabilistic model, the maximum likelihood estimator may assign the highest probability to an element that is never observed first. For example, given the data $\{ACB, BCA\}$, the maximum likelihood estimator is the maximizer of the function

[26] ACE: “If I add up the amount of times the surfer visited in each site for each route, the route I chose has the biggest number of visits so it’s most likely.”.

[27] ACE: “I picked the one he is more likely to pick statistically, (17/29 >12/29), and then the road was known (the only way to get to E from A is ACE).”.

[28] BCE: “If he goes to A it is more likely he’ll go to F. And because we know he went to E I guess he went to B first. It seems more likely he will end up to E from B. Because from C and D he can go to E still.”.

[29] ACE: “We know the individual ended at E, which means he got there either from C or D, C being more frequent. He can get to C from A or B, with A being more frequent. So rolling back we can say route ACE is the likely path.”.

[30] Arad and Gayer (2012) present an experiment in which subjects appear to evaluate the proportion of a color in an urn of colored balls according to the proportion of the color in a sample of past draws with replacement.

$[\frac{\mu(A)\mu(C)}{1-\mu(A)}][\frac{\mu(B)\mu(C)}{1-\mu(B)}]$ which is $(\mu(A), \mu(B), \mu(C)) = (1 - \frac{1}{\sqrt{2}}, 1 - \frac{1}{\sqrt{2}}, \sqrt{2} - 1) \sim (0.29, 0.29, 0.41)$.[31] Thus, the most likely sequence given the data is CAB or CBA even though C is never observed first in the data.

When applying the second probabilistic model to the data in W1, the maximum likelihood estimator assigns the probabilities (0.26, 0.39, 0.35) to A,B,C, respectively, implying that BCA is the most likely sequence, even though it is not the most frequently chosen. Thus, this probabilistic model does not generate the prediction based on T^{max} and is not a rationalization of the atomistic procedures observed in W1.

4.2. Rules of thumb

In W4, in which the distribution of choices was not explicitly provided, some of the observed procedures can be thought of as rules of thumb for choosing the most frequently observed alternative. A rule of thumb is evaluated by calculating its success rate given the underlying probabilistic assumptions regarding the circumstances in which it is used. We consider two probabilistic models:
S1: The frequencies of the alternatives are drawn from a uniform distribution over the simplex.
S2: The frequencies of the 6 alternatives are derived from 1001 independent draws from the uniform distribution over the alternatives.

The success of the rules of thumb under S1 was calculated[32] using Wolfram Mathematica,[33] and under S2 it was estimated by running a simulation with one million iterations and for each draw checking whether the sequence selected by the rule of thumb is the maximum of the drawn distribution. It was found that the success rates of H^{sum} given the two probability models are $7/8 = 0.875$ and 0.867, respectively. In the case of the atomistic procedure T^{max}, the rates are $3473/4320 \approx 0.804$

5. The potential for constructing new bounded rationality models

What can a theoretical economist do with the results? A large literature on "modeling bounded rationality" builds and analyses theoretical microeconomic models in which the standard assumptions about agents' behavior are replaced by more "realistic" procedural assumptions. Here are some examples: Piccione and Rubinstein (2003),[34] Eyster and Piccione (2013),[35] Esponda and Pouzo (2016),[36] Kfir and Spiegler (2020)[37] and Piccione and Rubinstein (2022).[38]

Similarly, we believe that one can create and analyse microeconomic models in which economic agents use the procedures observed in our experiments in order to make prediction. To illustrate the potential of this approach, consider a simple multi-generational world in which a generation consists of 21 agents. As in W1, each agent chooses the order of three objects: A, B and C and would like his choice to be identical to that made by another randomly chosen agent in the same generation. Therefore, he needs to predict the behavior of that anonymous agent. Assume that he can only base his decision on the choices made by the 21 agents in the previous generation and uses one of the following three procedures identified experimentally in W1:
- 8 agents applied the holistic procedure H^{max}: choose the most commonly made choice.
- 7 agents apply the atomistic procedure T^1: start with the object chosen first most often and choose the second element in the sequence to be the object which is chosen second most often from among the past choices that are consistent with the selection of the first element.
- 6 agents apply the atomistic procedure T^2: start with the object chosen first most often and choose the second element in the sequence to be the element most frequently chosen second from among all past choices.

Obviously, there exists a steady state with all agents choosing the same sequence. But there are also steady states with heterogenous behavior, such that all H^{max} agents choose CAB, the T^1 agents choose BAC and the T^2 agents choose BCA. Thus, once we allow for a variety of realistic prediction procedures, new non-trivial equilibria can emerge.

6. Related literature

6.1. Narrow bracketing

The distinction between holistic and atomistic procedures is related to that between wide and narrow choice bracketing. When a subject needs to choose one alternative from $\{A, B\}$ and another from $\{C, D\}$, narrow bracketing means that he makes two separate

[31] Setting the derivative of $log[\alpha\gamma/(1-\alpha)][\beta\gamma/(1-\beta)]$ to 0 gives $\frac{1}{\alpha} + \frac{1}{1-\alpha} = \frac{1}{\beta} + \frac{1}{1-\beta} = \frac{2}{\gamma}$ which implies that $\frac{1}{\alpha} + \frac{1}{1-\alpha} = \frac{1}{\alpha(1-\alpha)} = \frac{2}{1-2\alpha}$, i.e. $1 - 4\alpha + 2\alpha^2 = 0$.

[32] Aron Tobias suggested and carried out this calculation using Wolfram Mathematica.

[33] Using the fact that the frequency vector $(X_1, X_2, X_3, X_4, X_5, X_6)/\sum_{i=1,..,6} x_i$, where (X_i) is a vector of independent and exponentially distributed random variables with mean 1, is uniformly distributed over the 5-dimensional simplex.

[34] Piccione and Rubinstein (2003) model agents who differ in their ability to recognize temporal patterns of prices and demonstrate the existence of equilibria in which prices fluctuate in a pattern that is independent of the fundamentals.

[35] Eyster and Piccione (2013) analyse "a dynamic, competitive market, where traders lack structural knowledge and use different incomplete theories, all of which give statistically correct beliefs about next period's market price of the long-term asset.".

[36] Esponda and Pouzo (2016) suggest an equilibrium framework for a game in which a player is specified by "a (possibly misspecified) subjective model, which describes the set of feasible beliefs over payoff-relevant consequences as a function of actions.".

[37] In Eliaz and Spiegler (2020), equilibrium involves agents with two different narratives where a narrative is "a causal model that maps actions into consequences, weaving a selection of other random variables into the story.".

[38] Piccione and Rubinstein (2022) analyse a jury model in which a juror forms beliefs about the guilt of the accused given his personal signal and the event that the jurer is pivotal. It is shown that when a positive proportion of the jurors use a non-Bayseian method the resulting equilibrium is "disastrous".

choices, whereas wide bracketing means that he makes a single choice from the set $\{AC, AD, BC, BD\}$. Read et al. (1999) demonstrated that people may not bracket widely even when it is beneficial for them to do so.[39]

Note that the narrow/wide bracketing literature relates to situations in which a decision maker makes several separate choices, each yielding a separate prize. In contrast, we relate to situations where a decision maker makes one choice consisting of several components and his payoff depends on the entire combination of components.

Another somewhat related literature investigates procedures used to choose between two alternatives where each has two components (as in a lottery that yields $\$x$ with probability p). In this context, we observe holistic procedures, such as comparing the expectations of the two lotteries, alongside atomistic procedures that ignore a dimension in which the values are "similar" and making a decision by comparing the values in the other dimension (see, for example, Rubinstein (1988, 2013)).

6.2. Story building

The Z-experiments which involved a directed graph were inspired by Glazer and Rubinstein (2021). In that purely theoretical paper, we modeled a *story builder* as a procedure for extending partial information about visits to some of the nodes on the graph into a complete and coherent *story* (a path from the origin to a terminal node). Several examples of story builders were provided, some of which are holistic while others are atomistic. The completed story can be thought of as an answer to the question subjects were presented with in the current experiments. Note however that the two frameworks are very different. In Glazer and Rubinstein (2021), the story builder did not have information about past choices, which he does have in the current framework. There we focused on a consistency property (referred to as *stickiness*) of the story builder across instances in which he has varying information about a current choice while in the current experiments a story is constructed based on past events.

6.3. Framing

The difference in results between the cases in which data is provided explicitly (W1, W2, W3 and Z1) and those in which it is provided implicitly (W4, Z2 and Z3) can be viewed as the result of a framing effect. However, this framing effect differs from those discussed in the literature, where a minor difference in the text leads to a major difference in behavior (see Tversky and Kahneman (1986)). In all of the experiments, most subjects were probably interested in finding the most frequent alternative, but in W4, Z2 and Z3 the task of doing so was not straightforward. As a result, subjects tended to use rules of thumb which yield different behavior than in the cases where the distribution of past choices was given explicitely.

A related framing effect is suggested by Enke and Zimmerman (2019). Some of their subjects were asked to estimate a numerical state of nature based on four uncorrelated numerical signals, A, B, C and D, where the optimal estimate is their average and indeed most of the subjects chose the average. Other subjects were asked to estimate the state of nature based on $\alpha = A$, $\beta = A+B$, $\gamma = A+C$, $\delta = A + D$. The optimal estimate in this case is the average of α, $\beta - \alpha$, $\gamma - \alpha$, and $\delta - \alpha$; nevertheless, many of the subjects still averaged over α, β, γ and δ, thus ignoring the correlation between them.

6.4. Intuitive psychology

The paper is somewhat related to "intuitive psychology" which looks at whether "People are capable of learning other people's preferences by observing the choices they make" (Jern et al. (2017)). Unlike the "revealed preferences" literature in economics whose goal is to estimate an individual's preferences by applying various econometric techniques to data on his past choices, intuitive psychology studies how "real people" infer an individual's preferences from his past behavior. Analogously, we are not looking for a way to optimally predict an individual's next choice based on his own past behavior, which is probably the task of AI. Rather, we are interested in the ways that people actually make point predictions about another individual's next choice based on data about his past ones.

CRediT authorship contribution statement

Jacob Glazer: Writing – review & editing, Writing – original draft, Validation, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Ariel Rubinstein:** Writing – review & editing, Writing – original draft, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization.

Declaration of competing interest

The authors have no relevant source of financial support to disclose. IRB approval was obtained.

Data availability

Data will be made available on request.

[39] For example, in the case that A is a sure gain of \$240, B is a lottery that yields a gain of \$1000 with probability 0.25, C is a sure loss of \$750 and D is a lottery that yields a loss of \$1000 with probability 0.75, most people choose AD even though it is dominated by BC.

References

Arad, Ayala, Gayer, Gabi, 2012. Imprecise datasets as a source of ambiguity: a model and experimental evidence. Manag. Sci. 58, 188–202.
Eliaz, Kfir, Spiegler, Ran, 2020. A model of competing narratives. Am. Econ. Rev. 110, 3786–3816.
Enke, Benjamin, Zimmerman, Florian, 2019. Correlation neglect in belief formation. Rev. Econ. Stud. 86, 313–332.
Esponda, Ignacio, Pouzo, Demian, 2016. Berk-Nash equilibrium: a framework for modeling agents with misspecified models. Econometrica 84 (3), 1093–1130.
Eyster, Erik, Piccione, Michele, 2013. An approach to asset pricing under incomplete and diverse perceptions. Econometrica 81 (4), 1483–1506.
Gigerenzer, Gerd, Gaissmaier, Wolfgang, 2011. Heuristic decision making. Annu. Rev. Psychol. 62, 451–482.
Glazer, Jacob, Rubinstein, Ariel, 2021. Story builders. J. Econ. Theory 193, 105211.
Jern, Alan, Lucas, Christopher G., Kemp, Charles, 2017. People learn other people's preferences through inverse decision-making. Cognition 168, 46–64.
Piccione, Michele, Rubinstein, Ariel, 2003. Modeling the economic interaction of agents with diverse abilities to recognize equilibrium patterns. J. Eur. Econ. Assoc. 1 (1), 212–223.
Piccione, Michele, Rubinstein, Ariel, 2022. Real People Aggregating Signals: An Experiment and a Short Story. Memo.
Read, Daniel, Loewenstein, George, Rabin, Matthew, 1999. Choice bracketing. J. Risk Uncertain. 19, 171–197.
Rubinstein, Ariel, 1988. Similarity and decision-making under risk. J. Econ. Theory 46, 145–153.
Rubinstein, Ariel, 2007. Instinctive and cognitive reasoning: a study of response times. Econ. J. 117, 1243–1259.
Rubinstein, Ariel, 2013. Response time and decision making: a "free" experimental study. Judgm. Decis. Mak. 8, 540–551.
Simonson, Itamar, 1990. The effect of purchase quantity and timing on variety seeking behaviour. J. Mark. Res. 27, 150–162.
Tversky, Amos, Kahneman, Daniel, 1971. Belief in the law of small numbers. Psychol. Bull. 76, 105–110.
Tversky, Amos, Kahneman, Daniel, 1986. Rational choice and the framing of decisions. J. Bus. 59, S251–S278.