Similarity Nash Equilibria in Statistical Games*

Rossella Argenziano† and Itzhak Gilboa‡

May 2022

Abstract

A *statistical game* is a game in which strategic interaction is mediated via a binary outcome y, coupled with a prediction problem where a characteristic x of the game may be used to predict its outcome y based on past values of (x, y). In *Similarity Nash Equilibria*, players combine statistical and strategic reasoning, using an estimate of y as a coordination device. They predict y by its similarity-weighted frequency, and learn the optimal notion of similarity from the data. We prove that the model captures the importance of precedents and the endogenous formation of sunspots.

1 Introduction

1.1 A Motivating Example

The Soviet bloc started collapsing with Poland, which was the first country in the Warsaw Pact to break free from the rule of the USSR. Once this was allowed by the USSR, practically all its satellites in Eastern Europe underwent democratic revolutions, culminating in the fall of the Berlin Wall in 1989. The

*Argenziano gratefully acknowledges financial support through Leverhulme Research Fellowship RF-2021-103\7. Gilboa gratefully acknowledges support from the Investissements d'Avenir ANR -11- IDEX-0003 / Labex ECODEC No. ANR - 11-LABX-0047 as well as ISF Grants 1077/17 and 1443/20, the AXA Chair for Decision Sciences at HEC, the Foerder Institute at Tel-Aviv University and the Sapir Center for Economic Developmemt.

†Department of Economics, University of Essex. r_argenziano@essex.ac.uk

‡HEC, Paris and Tel-Aviv University. tzachigilboa@gmail.com

single precedent of Poland generated a "domino effect." This paper suggests a belief formation process that explains how a single precedent can have such a dramatic effect even in the absence of informational spillovers and strategic dependency among games.

Revolution attempts are typically modeled as coordination games: the expected utility derived from taking part in an uprising increases in the probability of its success, which in turn increases in the number of participants[1]. For a citizen trying to decide whether to join such an attempt, it is crucial to predict the outcome of the uprising. A natural piece of information to use for such a prediction is the outcome of past revolutions in similar contexts.[2] We suggest that the importance of the successful revolution in Poland didn't lie only in changing the relative frequency of successful revolutions, but also in changing the notion of *which* past revolution attempts were similar to current ones, hence relevant to predict their outcomes.

Specifically, the case of Poland was the first revolution attempt after the "Glasnost" policy was declared and implemented by the USSR. Pre-Glasnost attempts in Hungary in 1956 and in Czechoslovakia in 1968 had failed. In 1989, one might well wonder, has Glasnost made a difference? Is it a new era, where older cases of revolution attempts are no longer relevant to predict the outcome of a new one, or is it "Business as usual", and Glasnost doesn't change much more than does, say, a leader's proper name, leaving pre-Glasnost failed attempts relevant for prediction?

If the revolution attempt in Poland were to fail as did previous ones, it would seem that the variable "post-Glasnost" does not matter for prediction: with or without it, revolution attempts fail. As a result, when a person wonders what is the "right" way of judging similarity between past cases, she would likely be led to the conclusion that the variable "post-Glasnost" should be ignored, and that, consequently, the statistics are zero successes out of 3 revolution attempts. By contrast, because the revolution attempt in Poland succeeded, it had a double effect on the statistics. First, it increased the frequency of successful revolutions from 0:2 to 1:3. While $\frac{1}{3}$ is larger than 0, it still leads to pessimistic predictions about successes of future attempts.

[1] See, for example, Edmond, 2013.

[2] Steiner and Stewart, 2008, Argenziano and Gilboa, 2012, and Halaburda, Jullien, and Yehezkel, 2020 provide models in which similarity-weighted frequencies of past cases are used to form beliefs in coordination games.

However, if people also learn how to judge similarity, the single case of Poland leads them to the conclusion that "post-Glasnost" is an important variable. Indeed, the frequency of successes post-Glasnost, 1:1, differs dramatically from the pre-Glasnost frequency, 0:2 . Once this is taken into account, pre-Glasnost events are not as relevant for prediction as they used to be. If we consider the somewhat extreme view that post-Glasnost attempts constitute a class apart, the relevant empirical frequency of success becomes 1:1 rather than 1:3. Correspondingly, other countries in the Soviet Bloc could be encouraged by this single precedent, and soon it wasn't single any more.

1.2 Statistics and Equilibrium Selection

The example above illustrates the main ideas of the paper: if players share a common memory of similar games played by others, they can use this history to predict the outcome of the current game, hence to choose their optimal action. When considering past games, players need to make a relevance judgement: which cases are similar to the current one, in the sense that they are relevant to predict its outcome? We argue that players learn the optimal notion of similarity from history itself. Learning the similarity function from the data is referred to as "second-order induction".

To capture this reasoning in a model, we follow three steps. First, we associate a statistical problem to a binary coordination game. Second, we propose a solution concept that combines statistical and strategic reasoning. Finally, we specify our solution concept by proposing second-order induction as the form of statistical reasoning in which the players engage. We then prove that this simple model captures phenomena such as the importance of a single precedent and the endogenous emergence of sunspots.

Binary Statistical Coordination Games First, we introduce the notion of a binary statistical coordination game of regime change. The term refers to a binary coordination game accompanied by a statistical problem in which a variable y (the *outcome* of the game) is predicted based on an observed *characteristic* x and on past values of both x and y. The statistical problem interacts with the game in two ways: first, the value of y is determined by the players' strategy choices (and, possibly, by the current value of x); second, it affects the payoffs of the game. We assume that a player's utility depends only on her own strategy and on the values of (x, y) in the

current period. That is, what matters to a player is not strategic uncertainty per se, but the uncertainty about the outcome of the game. In this sense, the current value of y is a "strategically-sufficient statistic" for the game.

In our motivating example, each player chooses whether to join the revolution attempt or not. The characteristic x denotes the current state of the polity, and, specifically, whether it occurs before or after Glasnost was declared. The outcome y indicates the success or failure of the attempt. It depends on the players' choices (with the probability of success increasing in the number of players who join the revolution) and affects the payoffs of the two strategies. Neither the characteristics x not the outcomes y of past revolutions affect current payoffs. In this paper we restrict attention to binary variables x, y, which suffices to convey the main points.

Statistical and Strategic Reasoning Next, we propose that, when confronted with a statistical game, players combine statistical and strategic reasoning. To select her optimal choice, a player needs to make a prediction about the outcome y. Pure statistical reasoning would estimate y based on the observed current value of x and on past values of both x and y, ignoring the fact y will be determined by the players' chosen strategies. Pure strategic reasoning, on the other hand, when commonly known, would focus on equilibria of the game, and infer an estimate of y from the equilibrium strategy of all players. Strategic reasoning would thus ignore past values of the variables (x, y), which are payoff-irrelevant (as well as the current value of x if it is also payoff-irrelevant).

We propose a solution concept that combines both modes of reasoning, and that is compatible with many possible assumptions about rationality and higher order beliefs in rationality. In coordination games of regime change, there are typically two pure strategy Nash equilibria. In our motivating example, in one equilibrium citizens participate in the revolution, which therefore succeeds with high probability, and in the other one they do not, and it likely fails. We assume that players start with a statistical estimate of y based on past values (x, y) and on the current value of x, and choose a best-response to it. As a result, they play one of the two equilibria. The estimate of y thus acts as an equilibrium selection device: it singles out the equilibrium that can be justified by both strategic reasoning and pure statistical reasoning.

Second-Order Induction Finally, to complete the characterization of

our solution concept, we propose second-order induction as the statistical method used by players to estimate y. Statistics and machine learning offer a wide range of estimation and learning techniques and, in principle, each of these could be used as a way to define coordination devices.[3] We seek a method that can also serve as a reasonable model of the way most people think about their strategic choice, as in the example of the revolution games.

We start from the simplest prediction method, namely, estimating probabilities by empirical frequencies in similar cases in the past. This begs the question, which cases are deemed similar? In particular: will x be used to predict y or will x be ignored? In other words, will x act as a coordination device? Ignoring x would mean estimating the probability that y be 1 by the overall (unconditional) frequency of $y = 1$ in the past; by contrast, taking x into account would estimate it by the (conditional) empirical frequency of y in the sub-database in which x had the same value currently observed. In this paper, we assume that players learn the optimal estimation method. I.e., they choose the method that would have performed best had it been used in the past. This is a special case of the "empirically optimal similarity" as in Gilboa, Lieberman, and Schmeidler (2006) and Argenziano and Gilboa (2019). We label the equilibrium played by players forming an estimate of y based on second-order induction "Similarity Nash Equilibrium" (SNE).

The Results We prove that Similarity Nash Equilibria capture several phenomena having to do with equilibrium selection. First, the concept explains the importance of precedents, and provides an account of a mechanism by which a single success sets a domino effect into motion. Second, the process by which agents learn the similarity function from the data can also explain why some conspicuous but immaterial signals affect the play of the game and others do not. Specifically, the model describes the difference between successful and unsuccessful currency redenominations, showing when the seemingly-irrelevant currency denomination might become a determinant of similarity, and thereby change equilibrium selection, and when it will likely be ignored. Third, the results show that changing the similarity function becomes harder with experience. Finally, we provide an asymptotic result, showing that a "sunspot" may or may not emerge when the process is repeated. In our

[3] One may embed the game in a reasoning game, where each player first chooses a method of reasoning, and then plays a best response to the estimate that this method generates. If the original game is a coordination game, so will be the reasoning game.

model, equilibrium selection is generically unique in each period, but external shocks would determine whether it converges to be a signal-dependent selection (sunspot equilibrium) or a signal-independent one.

The rest of the paper is organized as follows. In Section 2 we present the formal definition of binary statistical coordination games and of "Similarity Nash Equilibrium". Section 3 presents our results. Section 4 discusses related literature, while Section 5 concludes with a discussion.

2 Model

2.1 Statistical Games

By the term "binary outcome game" we refer to a game in which each player has two possible actions, and her payoff is a function of her own action and a binary outcome that depends on all players' actions. Formally, a *binary outcome game* is a triple $G = (H, u, f)$ where:

(i) $H = [0, 1]$ is a continuum of players;

(ii) $u : \{0, 1\} \times \{0, 1\} \to \mathbb{R}$ is a player's payoff function, depending on her action and the outcome $y \in \{0, 1\}$;

(iii) $f : [0, 1] \to [0, 1]$ is a continuous function determining the distribution of the outcome as a function of the distribution of the players' actions.

The game $G = (H, u, f)$ defines a standard game played by members of H, as follows.

• Stage 1: All players take simultaneous actions: player $h \in H$ selects an action $a^h \in \{0, 1\}$, determining $a = \left(a^h\right)_{h \in H} \in \{0, 1\}^H$;

We assume that the set of players choosing each action is Lebesgue measurable.

• Stage 2: Nature selects a value for the outcome $y \in \{0, 1\}$ according to the distribution

$$\Pr\left(y = 1 \,|a\right) = f\left(\alpha\right)$$

where α is the proportion of players (in H) that chose $a^h = 1$.

• The game ends and player h's payoff is given by $u\left(a^h, y\right)$.

Note that the game is symmetric across players: there is a single function u for all players, and the function f depends only on the proportion of players choosing each action.

We assume that, in addition, players observe the realization of a *characteristic* x, that might or might not be payoff-relevant, and have access to data about past realizations of both y and x.[4] We restrict attention to the case in which x (like y) is a single binary variable. Formally, we define a *binary statistical problem of size* $(t-1)$ as $B_t = \left((x_i, y_i)_{i<t}, x_t\right)$ where, for each $i < t$, $x_i, y_i \in \{0,1\}$ are past realizations, and, at time t, the value x_t is observed.

Given a binary outcome game $G = (H, u, f)$ and a binary statistical problem $B_t = \left((x_i, y_i)_{i<t}, x_t\right)$, we think of (G, B_t) as a (binary) *statistical game*.[5] A statistical game differs from a standard game in two ways. First, it is augmented by a statistical problem $B_t = \left((x_i, y_i)_{i<t}, x_t\right)$. This problem is implicitly assumed to be commonly known to all players, as are the sets of players, their strategies, etc.[6] Past values of x and y are payoff-irrelevant but can serve as a coordination device. Second, the current values of x and y, (x_t, y_t), summarize the strategic aspect of the game. A player is assumed to know x_t, and if she also knew what y_t is about to be, she could ignore the strategy choices of the other players. Although y_t is stochastic and its distribution depends on all players' choices, its realization can be thought of as a "strategically-sufficient statistic" for the game G.

We are interested in the selection of equilibria in coordination games. Formally, a *binary outcome coordination game* is a binary outcome game $G = (H, u, f)$ with the normalized payoff matrix

$u\left(a^h, y_t\right)$	$y_t = 1$	$y_t = 0$
$a^h = 1$	1	0
$a^h = 0$	d	c

where (i) $0 < c, d < 1$, (ii) f is an increasing function, and (iii) $f(0) = \varepsilon$ and $f(1) = 1 - \varepsilon$ for some $\varepsilon \in (0, \bar{\varepsilon})$, with $\bar{\varepsilon} < \frac{1-d}{1-d+c}, \frac{c}{1-d+c}$.

[4] The general definition is silent on whether these past values of x and y were related to a game played at the time. In particular, it is possible that y reflects actions that were not a matter of choice, for example, if no other options were available at the time.

[5] A statistical game is therefore defined in the context of a given x_t. It follows that for a different value of x_t we can have a different game (or no game at all). In particular, the definition allows for the possibility that x_t is payoff-relevant.

[6] We implicitly assume that all the players encode information in the same way and that they agree on the meaning of statements such as "$x_i^j = 0$" or "$y_i = 1$". If, for instance, different players think of a given case as a "success" ($y_i = 1$) and others – as a "failure" ($y_i = 0$), without a 1-1 mapping between the different languages they use, we cannot assume a common process of statistical learning. See Sugden (1995) who proposes a theory of labeling in the context of coordination games.

Condition (i) guarantees that $a^h = y_t$ is the best response to y_t, condition (ii) guarantees that the game has strategic complementarities, and condition (iii) guarantees that the game has two strict Nash equilibria in which all players play $a^h = 0$ and $a^h = 1$, respectively. The corresponding statistical game will be referred to as a *binary statistical coordination game*

In section 3.4 we will consider sequences of statistical games, (G_t, B_t) where the games $G_t = (H_t, u, f)$ are identical, but each is played by a different set of players (to be precise, we assume that H_t are pairwise disjoint and the payoff function u is the same for the two possible realizations of x_t). The statistical problems are related, with B_t being the continuation of B_{t-1}, so that the game G_t has a longer history of past $(x_i, y_i)_{i<t}$ to consult than does G_{t-1}.

As in repeated games, sequences of statistical games allow players to use history as a coordination device. But, given that each player participates in only one statistical game, they do not have any long-run strategic considerations.

2.2 Similarity Nash Equilibria

How does a player $h \in H$ choose her action in game G? There are at least two approaches to the player's problem. The first relies on the fact that the player's payoff does not depend on the others' choices beyond the realization of y_t. Thus, the player can ask herself what y_t is likely to be, given x_t and previous values $(x_i, y_i)_{i<t}$, and directly best-respond to her estimate of the outcome. We refer to this as "statistical reasoning". The second approach, that we label "strategic reasoning", requires that the player take into account not only the dependence of her payoff on y_t, but also the dependence of the latter on all the players' actions, thus focusing on Nash equilibria of the game.

Formally, suppose players use pure statistical reasoning. Denote by

$$\bar{y}_t = \bar{y}_t\left((x_i, y_i)_{i<t}, x_t\right)$$

the players' statistical estimate (to be specified shortly) of the probability that $y_t = 1$, given x_t and previous values $(x_i, y_i)_{i<t}$. A player using pure statistical reasoning would play an action $a^h \in \{0, 1\}$ that maximizes

$$\bar{y}_t\left((x_i, y_i)_{i<t}, x_t\right) u\left(a^h, 1\right) + \left[1 - \bar{y}_t\left((x_i, y_i)_{i<t}, x_t\right)\right] u\left(a^h, 0\right)$$

as if $\Pr(y_t = 1)$ did not depend on her action a^h or the action of any other player. In a binary coordination game, if all players use pure statistical reasoning and compute the same statistical estimate of y_t, the resulting profile of actions $a^* = \left(a^h\right)_{h \in H}$ is a Nash equilibrium of G. Therefore, a^* is also compatible with the assumption that players use pure strategic reasoning, ignoring the past. Our solution concept proposes to use the external observer's statistical analysis as an equilibrium selection, or coordination device, that selects the outcome of the game compatible with both pure statistical reasoning and pure strategic reasoning.

To complete the characterization of our solution concept, we need to specify how players form the statistical estimate $\bar{y}_t$. The most fundamental method to estimate y_t from the commonly known history of past games would be its empirical frequency:

$$\overline{y}_t = \frac{1}{t-1} \sum_{i<t} y_i.$$

However, in line with Hume's (1748) dictum, "from causes $[x]$ which appear similar, we expect similar effects $[y]$", we should ask ourselves, are all past games "similar" to the current one, i.e., relevant to predict its outcome? Or should one only take into account periods i in which $x_i = x_t$? In other words, should one look at the overall empirical frequency of y or only at the conditional one? More formally, if players predict y_t by a similarity-weighted average[7]

$$\overline{y}_t^s = \frac{\sum_{i<t} s(x_i, x_t) y_i}{\sum_{i<t} s(x_i, x_t)}$$

would players use the similarity defined by $s_0(x_i, x_t) = 1$ for all x_i, x_t or by $s_x(x_i, x_t) = \mathbf{1}_{\{x_i = x_t\}}$?[8]

Psychological evidence suggests that people learn the notion of similarity between data points from the database itself.[9] We therefore assume that players choose the similarity function that, had it been used to predict the existing data points, where each is estimated based on the others, would have

[7] For cases where $\sum_{i<t} s(x_i, x_t) = 0$, we define $\overline{y}_t^s = 0.5$.

[8] Observe that we only consider two similarity functions here. One could allow for a variety of other functions, for example, letting $s(1,0) = s(0,1) = \alpha$ for $\alpha \in (0,1)$ while retaining the normalization $s(1,1) = s(0,0) = 1$.

[9] See Nosofsky (1984, 1986, 1991).

performed best.[10]

Formally, we use a leave-one-out cross-validation technique.[11] For a similarity function s, and $i < t$, define

$$\widehat{y}_i^s = \frac{\sum_{r \neq i} s(x_r, x_i) y_r}{\sum_{r \neq i} s(x_r, x_i)}$$

and consider the sum of squared errors,

$$SSE(s) = \sum_{i=1}^{t-1} (\widehat{y}_i^s - y_i)^2$$

We assume that players estimate y_t by $\overline{y}_t^s$ using an *empirically optimal similarity*, which we define as a similarity function (between s_0 and s_x) that minimizes the SSE. In case of a tie we assume that s_0 is selected, as it is simpler, using fewer variables in the similarity judgment.[12]

Using the similarity function s_x allows one to make distinct predictions $\overline{y}_t$ for two sub-databases (depending on the value of x_t). Intuitively, this additional freedom should result in a lower SSE overall. However, with a relatively small database, the freedom to select $\overline{y}_t$ comes at a cost: some observations may be relatively "isolated" in their sub-database, implying a loss in accuracy.[13]

We conclude this section by formally defining our solution concept: we define *Similarity Nash Equilibria (SNE)* of the statistical game (G, B_t) to be any action profile $\widetilde{a}$ such that for each player $h \in H$, the following two

[10] See Argenziano and Gilboa (2019) for similar definitions in a continuous model.

[11] The leave-one-out cross validation technique is widely used in machine learning and in statistics. We use it here as an idealized model of the way people learn which similarity function is the most appropriate to use in making predictions.

[12] The preference for fewer variables is similar to the simplicity criteria implicit in the adjusted R^2, Lasso, the Akaike Information Criterion etc. Standard arguments for the preference for simplcity apply here. In particular, using fewer variables results in lower memory and computation costs. The similarity s_0 has the additional advantage over s_x of having fewer cases of an empty database. However, the choice of a tie-breaking rule is immaterial for the results that follow.

[13] While we only consider here one dimension, the basic logic is identical to that of "the curse of dimensionality".

conditions hold

$$\begin{aligned}\widetilde{a}^h &\in \arg\max_{a^h}\left[f\left(\widetilde{\alpha}\right)u\left(a^h,1\right)+\left[1-f\left(\widetilde{\alpha}\right)\right]u\left(a^h,0\right)\right]\\ \widetilde{a}^h &\in \arg\max_{a^h}\left[\bar{y}_t\left(\left(x_i,y_i\right)_{i<t},x_t\right)u\left(a^h,1\right)+\left[1-\bar{y}_t\left(\left(x_i,y_i\right)_{i<t},x_t\right)\right]u\left(a^h,0\right)\right]\end{aligned}$$

where $\widetilde{\alpha}$ denotes the measure of players playing action 1 in action profile $\widetilde{a}$.

3 Results

It will be convenient to use the following notation: there are $(t-1)$ points in the database, and they are divided into four types, according to the values of x and of y. Let the number of cases of each type be given by the following case-frequency matrix:

# of cases	$x=0$	$x=1$
$y=0$	L	l
$y=1$	W	w

In the motivating example of subsection 1.1, let $y=1$ (or zero) denote the success (or failure) of a revolution attempt (w for "win", and l for "lose"), while $x=1$ (or zero) – whether or not it occurred post-Glasnost. Consider citizens in Hungary in 1989. They lived in a post-Glasnost world, i.e., $x=1$. After the successful revolution in Poland, they observed two failed revolutions pre-Glasnost, and a successful one post-Glasnost: $(L,W,l,w)=(2,0,0,1)$. At this point, if they had ignored x they would have predicted failure as the most likely outcome of a revolution attempt (with a relative frequency of 2/3) and therefore, for reasonable choices of c,d,f, would have found it optimal not to take part in one. Instead, by taking into account x, they would have considered only the case of Poland as relevant for their predictions, expected a success, and therefore participated in the attempt. Second-order induction is consistent with the fact that Glasnost was indeed considered relevant for predictions and a revolution was therefore attempted (successfully) in Hungary. Ignoring x yields $SSE\left(s_0\right)=1.5$ while taking it into account reduces the sum of squared errors to $SSE\left(s_x\right)=0.25$. Thus, the single case of a successful revolution made the variable "post-Glasnost" informative enough to enter the

similarity judgment. Note that, had the case of Poland ended in a failure, $SSE(s_0) = 0$ would hold and the empirically optimal similarity would ignore the post-Glasnost variable.

In the rest of this section we will focus on larger databases, assuming that there is a non-trivial history in which $x = 0$. Specifically, we assume throughout that $L, W > 2$. This assumption means that (i) history contains a non-trivial number of cases overall, and that (ii) the prediction of the outcome y is a non-trivial task: there are a few (at least three) cases with $y = 0$ as well as with $y = 1$.

3.1 A New Value

We start by looking at SNE of statistical games for which there's a non-trivial history of cases with different outcomes but the characteristic x had a constant value $x = 0$ in all of them: $L, W > 2$ and $l = w = 0$. Consider classical examples of coordination games such as a revolutionary attempt, a bank run, or a currency attack. Suppose that, in a sequence of such games, $x = 1$ is observed for the first time: a new political leader appears, or a new policy is announced. History includes cases with various outcomes of analogous attempts to attack a government, a bank, or a currency. Some succeeded, some failed. But in all these cases, the new leader or policy was not in place (x was constantly equal to zero). As a result, x doesn't have any predictive power in the existing database, hence the first time that $x = 1$ appears, it is ignored.[14] The natural question then is: what will it take for players to start paying attention to it? Starting from a clean slate, what does it take for a new leader or policy to be taken seriously, to be considered something that separates history into two periods: a past regime, which is not relevant anymore, and a new regime which contains cases relevant for predicting the outcome of the current game?

Our first two results answer this question. Proposition 1 says that even a single case is sufficient to convince players that they are under a new regime, if and only if the observed outcome y is the one which had been less frequently

[14] Observe that, since all past cases have $x = 0$, the characteristic does not affect their similarity to each other. Thus, one obtains exactly the same in-sample predictions whether one considers the variable x or not. This means that $SSE(s_0) = SSE(s_x)$. However, the similarity function s_x cannot be used for out-of-sample prediction as it defines an empty database. As mentioned above, the tie-breaking rule favors s_0.

observed in the past. This result is rather intuitive: in order to be noticed, one needs to be different.

Proposition 1 *Let $L, W > 2$. If $(l, w) = (1, 0)$, any*[15] *SNE is selected by s_x if $L < W$ and by s_0 otherwise. Symmetrically, if $(l, w) = (0, 1)$, any SNE is selected by s_x if $L > W$ and by s_0 otherwise.*

Thus a new feature (leader, policy, etc.) that results in the modal outcome will not be considered relevant for prediction. However, if it is consistently the case that $x = 1$ is associated with a particular value of y, we would expect players to "notice" this regularity by taking x into account in the similarity judgement. The following result corroborates this intuition and shows that "consistently" need not be more than twice, provided that there are no counter-examples:

Proposition 2 *Let $L, W > 2$. If either ($l > 1$ and $w = 0$) or ($l = 0$ and $w > 1$), then any SNE is selected by s_x .*

The importance of this proposition lies in the comparison of case-based and rule-based reasoning: while our model does not equip players with the language in which general rules can be stated, learned, or acted upon, the empirically-optimal similarity function can mimic this type of reasoning. If it so happens that the associative rule "If $x_i = 1$ then $y_i = b$" (for $b \in \{0, 1\}$) is valid in the database, the players will notice this regularity: the empirically optimal similarity function will be s_x and in any SNE of the game, if $x = 1$, players will expect $y = b$ and play $a^h = b$. By contrast, if $x = 0$, they will expect y to be equal to the average value of y in the past cases with $x_i = 0$ and play accordingly.

As an example of Proposition 1, consider a central bank which redenominates its currency in an attempt to restrain inflation. Inflation is an equilibrium phenomenon: an economic agent who expects others to raise prices of goods and services would be wise to do so herself. Thus, one can think of the inflation game as a price-setting game with multiple equilibria, and redenomination as an attempt to switch from a hyperinflation equilibrium to a low

[15] Recall that for each similarity function the corresponding Nash equilibria are generically unique. In our setup there is always a unique empirically-optimal similarity function (either s_0 or s_x), and non-uniqueness can only follow from ties.

inflation equilibrium[16]. If x denotes the new currency, then $x_i = 0$ throughout all cases in history ($i < t$), and setting $x_t = 1$ is an attempt to signal a new regime, and to coordinate on the non-inflationary equilibrium. Will economic agents use x in their belief formation, or will they dismiss the redenomination as a "cosmetic change" and believe that inflation will continue to run high? Proposition 1 suggests that the answer depends on the first period: if, in this period, inflation is low – namely, y takes the value that was less frequent in the past – the characteristic will be used for prediction and a new, low-inflation equilibrium can be reached. By contrast, if in the first period the inflation rate continues to be high, the redenomination will be judged irrelevant. Israel switched from a Lira to a Shekel (worth 10 Liras) in 1980 and then to a New Shekel (worth 1,000 Shekels) in 1985. In 1980 the change was not accompanied by fiscal policy changes, and inflation spiraled into hyperinflation. By contrast, the change in 1985 was accompanied by budget cuts, and inflation was curbed in the following years. These two examples seem to corroborate the intuition behind Proposition 1: a change of currency is a payoff-irrelevant but perceptually-conspicuous difference that might change the equilibrium selected; whether it succeeds in doing so depends on the realization of a payoff-relevant variable (y). In these examples psychological considerations suggest potential sunspots; but rational learning of the optimal similarity function implies that economic outcomes will determine which sunspots are used for coordination and which get ignored.

3.2 The Power of a Single Precedent

Suppose now that after a non-trivial history ($L, W > 2$) of cases with $x = 0$, a new leader appeared, $x = 1$, and established herself as relevant for prediction either through a series of consistent outcomes, as in Proposition 2, or through a single, "surprising" outcome, as in Proposition 1. The next proposition asks what would it take for the new leader to *lose* her role as a coordination device. Would a single inconsistency, a single precedent with the opposite outcome, be enough for the players to stop paying attention to the characteristic x?

[16] See Mosley (2005): "...redenominations often occur after economic crises, as governments attempt to convince citizens and markets that hyperinflation is a thing of the past. In some cases, the timing is correct, in that redenomination caps off high levels of inflation. In other cases, governments are not able to reign in inflation immediately after redenomination, and they may make multiple efforts....".

The result is rather intuitive: a single precedent can make a characteristic irrelevant for prediction if the number of consistent outcomes of the opposite sign that have established its relevance is not too large.

Proposition 3 *Let $L, W > 2$. If either ($l = 1$ and $0 < w \leq \left\lfloor \frac{W}{L} \right\rfloor + 1$) or, symmetrically, ($w = 1$ and $0 < l \leq \left\lfloor \frac{L}{W} \right\rfloor + 1$), then any SNE is selected by s_0.*

Consider the first statement (the second is symmetric): if relevance for prediction had been established with a single surprising outcome, i.e., if $W < L$ and $w = 1$, a single case ($l = 1$) makes the characteristic irrelevant again. Similarly, it makes it irrelevant if relevance had been established with multiple, but not too many, outcomes of the type most frequent in the past, i.e., if $W > L$ and $1 < w \leq \left\lfloor \frac{W}{L} \right\rfloor + 1$. Finally, note that, if $W > L$ and $w = 1$, we already know by Proposition 1 that the empirical similarity is s_0 for $l = 0$, and Proposition 3 shows that this is the case also for $l = 1$: if in the first case in which the new leader was in office the outcome of the game was the one most frequent in the past, the new leader does not become a coordination device, and that is still true even if a second case ends up having the opposite outcome.

3.3 The General Case

We now turn to the more general case, where a new leader ($x = 1$) appeared after a non-trivial history with $L, W > 2$, and outcomes of both types have been observed: $l, w > 0$. We ask what it will take for players to take into account the change in leadership when they form their beliefs. The basic intuition is simple: if the ratio w/l is close to W/L, the change of leadership will seem immaterial and players will ignore it when forming beliefs: the empirically optimal similarity is s_0. If, however, the relative frequency of $y = 1$ in the sub-database corresponding to $x = 1$ is very different from that corresponding to $x = 0$, players will be convinced that they are under a "new regime" and the empirically optimal similarity will be s_x.

Proposition 4 starts from a scenario in which the sub-database with $x = 1$ has, up to integrality constraint, the same ratio of cases with $y = 0$ and $y = 1$ as the sub-database with $x = 0$. In this case x is irrelevant for predicting y (part (i) of the Proposition 4). Suppose that we now increase w. We find that this improves the performance of the similarity function that takes x into

account, up to a point where it outperforms the similarity function that does not (part(ii)). As could be expected, the minimum $w^* > \frac{lW}{L}$ for which this inequality holds increases in the number of cases with the opposite outcome, l (part (iii)). Moreover, up to details of integrality constraints, the number of additional cases needed to get to this minimum ($w^* - \frac{lW}{L}$) is also non-decreasing in l (part (iv)).

Formally, let $[\ ] : R \to Z$ be the nearest integer function, selecting the ceiling in case of a tie. (That is, for all $x \in R$ and $z \in Z$, we have $[x] = z$ if $x = z + \varepsilon$ and $\varepsilon \in [-0.5, 0.5)$.) We prove the following:

Proposition 4 *Let L, W, l, w be any four integers such that $L, W > 2$, $l > 0$, and $w = \left[\frac{lW}{L}\right] \geq 0$. The following hold:*

(i) For databases (L, W, l, w) and $(L, W, l, w + 1)$, the unique SNE is the one selected by s_0.

(ii) There exists an integer $w^*(L, W, l) \geq w+2$ such that, for every $q \geq w$, the unique SNE is the one selected by s_0 for $q < w^*(L, W, l)$ and by s_x for $q \geq w^*(L, W, l)$. (Clearly, if such an integer exists it is unique.)

(iii) $w^*(L, W, l)$ is non-decreasing in l.

(iv) If W/L is an integer, $(w^*(L, W, l) - w)$ is non-decreasing in l.

Thus, our model captures the fact that it is harder to re-establish relevance than to establish it at the outset. Suppose that a new leader whose identity is characterized by $x = 1$ wishes to associate herself with successes, that is, to make others predict that $y = 1$ when $x = 1$. Let us assume that, in the past, successes were less frequent than failures ($W < L$) so that if the leader does not single herself out, players will expect failures and such beliefs will be self-fulfilling. On this background, Proposition 1 guarantees that starting off with a single success ($w = 1$, $l = 0$) suffices to establish relevance of x and thereby to place the leader in a class apart. In the sub-database defined by $x = 1$, only the less frequent outcome $y = 1$ has been observed and thus the leader is associated with success.

However, if it so happens that one starts out with a failure ($w = 0$, $l = 1$) the task will be harder: by Proposition 1, the leader's identity won't be considered relevant after the initial failure and parts (i) and (ii) of Proposition 4 show that for the leader to be noticed, and associated with successes, at least two or three successes will be needed (depending on how unusual successes were in the past). More generally, for any number of adverse outcomes $l > 0$

there is a sufficiently large number of successes w that would eventually make x a coordination device followed by the players (part (ii)), but the number of successes required (part (iii)), and even the *additional* number of such successes (part (iv)) weakly increase (up to integrality constraints in part (iv)). One does get a second chance to make a first impression, but it becomes costlier.

3.4 Sequences of Statistical Games

We consider now a sequence of binary statistical coordination games, and assume that the only relevant statistics are the past plays of these games. The payoffs are

$u\left(a^h, y_t\right)$	$y_t = 1$	$y_t = 0$
$a^h = 1$	1	0
$a^h = 0$	d	c

and we assume that c and d are independent of x_t. This assumption simplifies the computations, though similar results would hold without it. More importantly, this assumption allows us to study the pure coordination role of x: should we find a convergence to playing at period t an equilibrium that depends on x_t, we will think of x_t as a sunspot, that is, a coordination device that does not affect payoffs at all. Our question is, therefore, will x become a sunspot that determines equilibrium selection, or will it be ignored?

We assume that, for each t, x_t is exogenously determined according to an i.i.d. process. Specifically,

$$x_t = \begin{cases} 0 & \text{with probability } 1-\beta \\ 1 & \text{with probability } \beta \end{cases}$$

independently of $(x_i, y_i)_{i<t}$ with $\beta \in (0,1)$. At time t, after x_t is realized, all players in H_t observe x_t as well as the history that can be summarized by

# of cases	$x = 0$	$x = 1$
$y = 0$	L_t	l_t
$y = 1$	W_t	w_t

where $L_t + l_t + W_t + w_t = t - 1$. We will be interested in the limit of the

relative frequencies, and observe that, with probability 1,

$$\begin{aligned} \frac{L_t + W_t}{t-1} &\rightarrow {}_{t\to\infty} 1-\beta \\ \frac{l_t + w_t}{t-1} &\rightarrow {}_{t\to\infty} \beta \end{aligned}$$

– so that the question is how often each equilibrium will be played when $x_t = 0$ and when $x_t = 1$.

Consider the following four candidates for limit frequency matrices:

I	$x=0$	$x=1$	II	$x=0$	$x=1$
$y=0$	$(1-\varepsilon)(1-\beta)$	$(1-\varepsilon)\beta$	$y=0$	$(1-\varepsilon)(1-\beta)$	$\varepsilon\beta$
$y=1$	$\varepsilon(1-\beta)$	$\varepsilon\beta$	$y=1$	$\varepsilon(1-\beta)$	$(1-\varepsilon)\beta$

III	$x=0$	$x=1$	IV	$x=0$	$x=1$
$y=0$	$\varepsilon(1-\beta)$	$(1-\varepsilon)\beta$	$y=0$	$\varepsilon(1-\beta)$	$\varepsilon\beta$
$y=1$	$(1-\varepsilon)(1-\beta)$	$\varepsilon\beta$	$y=1$	$(1-\varepsilon)(1-\beta)$	$(1-\varepsilon)\beta$

where in matrices I and IV history suggests that the equilibrium does not depend on x and in matrices II and III – that it does. We can now state

Proposition 5 *Under the assumptions above,*

(A) the relative frequencies $(L_t, l_t, W_t, w_t)/(t-1)$ *converge to one of the matrices* I, II, III, IV *with probability* 1*. Moreover, each of the matrices above is the limit with positive probability;*

(B) the optimal similarity converges to a limit with probability 1*: it is* s_x *from some* t *onwards, if the relative frequencies converge to matrix* II *or* III*, and it is* s_0 *from some* t *onwards if the relative frequencies converge to matrix* I *or* IV*.*

Thus, we find that for all games that satisfy our assumptions, the variable x may or may not determine equilibrium selection in the limit. Clearly, in case the limit is one of the matrices II or III, the ratios L_t/W_t and l_t/w_t become very different (one approaching $\frac{\varepsilon}{1-\varepsilon}$ and the other – $\frac{1-\varepsilon}{\varepsilon}$), so that the empirically optimal similarity function is bound to be s_x. By contrast, if the limits are matrices I or IV, both ratios converge to the same limit point ($\frac{\varepsilon}{1-\varepsilon}$ or $\frac{1-\varepsilon}{\varepsilon}$). This means that the optimal choice of the players is determined to be the same, and thus the same equilibrium is selected for both values of x.

However, this does not yet imply that the empirically optimal similarity is s_0: it is possible that the rate of convergence of the two ratios is different and that, along the way to the limit, they are sufficiently different so as to make s_x the optimal similarity. Part (B) of the Proposition states that this is not the case. Thus, we obtain the following result: the process may converge to a limit in which there is a sunspot x, so that the players coordinate on $a^h = x$ or on $a^h = 1 - x$. It is also possible that the players ignore the sunspot and play $a^h = 0$ or $a^h = 1$ independently of x. In the latter case, the empirically optimal similarity will indeed reflect the fact that the players make the same choices whether $x = 0$ or $x = 1$.

4 Related Literature

Statistical games are reminiscent of "Aggregative Games" (Selten, 1970) and of "Congestion Games" (Rosenthal, 1973, Schmeidler, 1973) in that a player's payoff depends only on a summary statistic of the others' choices. In the former, strategies are real numbers and the statistic is their sum. In the latter, there are typically finitely many strategies and the statistic is the relative frequencies of choice. In both, each player finds the others interchangeable. Similarly, in statistical games each players should only bother about the prediction of y, and the others' choices only matter to the extent that they affect y. The definition of statistical games brings the summary statistic y to the fore, allowing for a variety of ways in which it is determined by players' choices, encapsulated in the function f.[17] Moreover, statistical games are equipped with a history of past observations of x and y, which has no counterpart in the standard models of aggregative or congestion games.

Statistical games are similar to Correlated Equilibria (Aumann, 1974) in that we assume that Nature sends a signal to each player before the game is played. However, in our context the signal is commonly known. Thus, the correlation device x (coupled with the database $(x_i, y_i)_{i<t}$) selects an equilibrium but does not allow non-equilibrium plays. In this sense our correlating signal, x, brings to mind "sunspots" (Cass and Shell, 1983). In particular, if one imposes the additional assumption that in a statistical game x is payoff-

[17] Note that, if we were to allow y to assume values in larger spaces, aggregative games and congestion games could be embedded in our model (by allowing y to be real-valued, or a point in a corresponding simplex, respectively).

irrelevant, it does function, like sunspots, as a mere public correlation device. Viewed thus, our suggestion to use second-order induction to find the similarity function can be considered a theory of sunspot selection.

Kets and Sandroni (2021) suggest a notion of equilibrium selection that is based on impulses, which are attended to by introspection and used as coordination devices. Similarity Nash Equilibria bear resemblance to their equilibrium selection process, in particular by using some non-strategic hunch that is also compatible with strategic reasoning. However, Similarity Nash Equilibria focus on statistical learning rather than on cultural impulses, and the notion of second-order induction suggests which signals will be used for equilibrium selection and which might be ignored.

When considered as a method of equilibrium selection in coordination games, statistical games cannot fail to remind one of "Global Games" (Carlsson and van Damme, 1993). Like Global Games, our approach attempts to embed the game in context in order to predict equilibrium selection. However, in Global Games equilibria are chosen ex ante, simultaneously for all games, whereas in statistical games they are chosen sequentially, highlighting the role of statistical learning. Global Games rely on some uncertainty about the game played, while a statistical game is commonly known among its players, and the variable x only serves as a coordination device.

In a 2x2 symmetric coordination game, Similarity Nash Equilibria are related to risk-dominant equilibria (Harsanyi and Selten, 1988). Specifically, assume that there is no history to be considered ($t = 1$) and that players use an initial guess of $P(y = 1) = 0.5$. When players best respond to this guess, they will select the risk-dominant equilibrium. Indeed, even when a history $(x_i, y_i)_{i<t}$ is available, the players may choose to ignore it, use $P(y = 1) = 0.5$ as a starting point and select the risk-dominant equilibrium. By contrast, Similarity Nash Equilibria assume that the initial statistical estimate is a function of history, where the values of (x, y) are used for weighted averaging, as well as for determining the weights in the averaging formula.

As mentioned above, one can also view Similarity Nash Equilibria as a possible formalization of Schelling's (1960) focal points: estimating y based on its past values, and finding the equilibrium that consists of best responses to this estimate can be viewed as a procedure to determine focality. In the simplest case, assume that a game is played repeatedly and that a given equilibrium is played in the vast majority of past observations. It then stands

to reason that a statistical prediction function would estimate a value of y that gives rise to the same equilibrium played in the past. In this sense, SNEs capture "statistical focality". We view this analysis as complementary to Sugden (1995), who focuses on the labelling in pure coordination games.

Similarity-weighted relative frequencies are formally equivalent to kernel estimation of probabilities (Akaike, 1954, Rosenblatt, 1956, Parzen, 1962; see Silverman, 1986) and they are also reminiscent of exemplar learning in psychology (Shepard, 1957, 1987, Medin and Schaffer, 1978, Nosofsky, 1984, 1988). The formula has also been axiomatized in Billot, Gilboa, Samet, and Schmeidler (2005) (if y takes at least three values), and in Gilboa, Lieberman, and Schmeidler (2006) (for the case of two values discussed here).

As previously mentioned, Steiner and Stewart (2008), Argenziano and Gilboa (2012), Halaburda, Jullien, and Yehezkel (2020) deal with Nash equilibria selected by appropriately defined similarity functions. As opposed to this literature, in this paper we do not assume that a similarity function is given a priori, but that it is learned from the data itself. This notion of "second order induction" (in the terms of Gilboa, Lieberman, and Schmeidler, 2006 and Argenziano and Gilboa, 2019) appeared both in the statistical literature (Hardlë and Marron, 1985) and in the psychological one (Nosofsky, 2011).

5 Discussion

For binary statistical coordination games, at least one SNE exists for any database B_t, and it is consistent with a gamut of assumptions on the players' higher-level reasoning. For example, they may all be strategic and be aware of the fact that statistics only serves as a coordination device, or they may all be statistical and ignore the fact that other players are optimizing relative to their beliefs, too. Moreover, there could be a fraction $\eta \in (0,1)$ of statistical players, and the strategic players might or might not be aware of this fact. Since each Nash-Equilibrium (NE) action is a best response to a range of beliefs, as long as the different modes of reasoning concur on the same best response, equilibrium behavior may result even in case of disagreement on beliefs. For example, a statistical player may think that the probability of $y = 1$ is just high enough to choose action 1, whereas a strategic player may

think that $y = 1$ will occur with a (say, higher) probability $f(1)$. Yet, their best response is the same.

Alternatively, suppose that each player in a binary statistical coordination game is capable of Level-K reasoning for a given $K \leq \infty$ (see Nagel, 1995, Stahl and Wilson, 1995). There may be players at level $K = 0$, who are incapable of strategic reasoning, and they estimate $P(y_t = 1)$ by $\bar{y}_t$ and respond optimally to this estimate. There are others who are at level $K = 1$, and compute $\bar{y}_t$ as well as the best response to this estimate, and believe that this best response would be the choice made by all the other players, and so forth. Eventually we may find also players of Level-∞ reasoning, who can compute equilibria. These players may also be sophisticated enough to have beliefs over the distribution of levels of reasoning in the population. Because an SNE consists (by definition) of strategies that are best response to the initial guess, $\bar{y}_t$, and to themselves, all levels of reasoning would lead to the same choices, namely the equilibrium strategies. Similarly, even if all the players are in fact capable of Level-∞ reasoning, but this fact is not common knowledge among them, we might be led to an SNE again. Thus, SNE are rather robust to assumptions about rationality and common belief thereof in binary statistical coordination games.

By contrast, in more general coordination games SNE might fail to exist. For example, consider a modified version of the coordination game described in Section 2. Suppose that there is a continuum of heterogeneous players where player h's payoff is given by

$$\begin{array}{ccc} u^h\left(a^h, y_t\right) & y_t = 1 & y_t = 0 \\ a^h = 1 & 1 + \varepsilon^h & 0 \\ a^h = 0 & 0 & 1 - \varepsilon^h \end{array}$$

and $\varepsilon^h \sim U(-1, 1)$, so that her best response is to join the revolution attempt if and only if she thinks that the probability of success is at least $\frac{1-\varepsilon^h}{2} \sim U(0, 1)$. For any initial belief $\Pr(y_t = 1) = p_0 \in (0, 1)$, the best response would be to join the revolution for a fraction $\alpha_0 = p_0$ of the population and not to join it for the remaining fraction. If, for example, $f(\alpha) = \alpha^2$, no SNE exists. One may generalize the statistical-strategic reasoning process and the notion of SNE, allowing for an iterative process of best-response to initial beliefs. In the example above, the best response to the initial belief

$\Pr(y_t = 1) = p_0 \in (0,1)$ would be to join the revolution for a fraction $\alpha_0 = p_0$ of the population. This in turn would generate beliefs $p_1 = f(p_0) = p_0^2 < p_0$, to which the best response would be to join the revolution for an analogous fraction p_1 of the population. Such an iterative process would converge to an equilibrium with $\alpha = 0$ for any initial belief $p \in (0,1)$. Note that an iterative process of best responses is at the heart of equilibrium selection in Global Games (Carlsson and van Damme, 1993). Thus, an extension of our equilibrium selection to iterative best responses can simultaneously generalize Global Games (by allowing different games) and our analysis above.

Another class of games where SNEs need not exist are statistical games where the two modes of reasoning lead to conflicting best responses. For example, consider a simple binary congestion game with a continuum of identical players with payoff

$$\begin{array}{cc} & u\left(a^h, y_t\right) \\ a^h = 0 & y_t \\ a^h = 1 & 1 - y_t \end{array}$$

in which the distribution of a continuous outcome $y \in [0,1]$ is determined by the fraction α of players choosing action 1, and $\mathbb{E}(y) = \alpha$. The game has a unique symmetric NE, in which all players choose the mixed strategy $(0.5, 0.5)$. A statistical player's best response to belief $\bar{y}$ is $a^h = 1$ if $\bar{y} \leq 0.5$ and $a^h = 0$ if $\bar{y} \geq 0.5$. Therefore, if at least some players use statistical reasoning, SNE almost never exists. More precisely, it exists only for databases B_t that generate a belief $\bar{y} = 0.5$.

Finally, another example of conflict between the two modes of reasoning and non-existence of SNE can arise in Centipede Games. In these games, strategic reasoning leads to the unique equilibrium outcome, in which the first player stops the game ("play Down"). On the other hand, statistical reasoning can lead a player to continue the game ("play Across"), if a given history of centipede games played by other players leads her to believe with high enough probability that the next player will also do so for at least one more stage.

Appendix: Proofs

For the following proofs, it is useful to define $\Delta(L,W,l,w) \equiv SSE(s_x) - SSE(s_0)$, where $\Delta(L,W,l,w) > 0$ implies that the variable x should not be included in the empirically optimal similarity function, whereas $\Delta(L,W,l,w) < 0$ implies that it should. Clearly, $\Delta(L,W,l,w) = \Delta(W,L,w,l)$ and $\Delta(L,W,l,w) = \Delta(l,w,L,W)$, as the SSE calculations do not change if we switch between 0 and 1 either for a predictor x or for the predicted variable y[18].

Proof of Proposition 1:

We need to show that

(i) If $L < W$, $\Delta(L,W,1,0) < 0$ and $\Delta(L,W,0,1) > 0$;

(ii) If $L > W$, $\Delta(L,W,1,0) > 0$ and $\Delta(L,W,0,1) < 0$;

(iii) $\Delta(L,L,1,0), \Delta(L,L,0,1) > 0$.

We first show that $\Delta(L,W,1,0)$ is positive for $L \geq W$ and negative for $L < W$. By symmetry, this implies that $\Delta(L,W,0,1)$ is positive for $L \leq W$ and negative for $L > W$, together completing the proof.

The SSE's are given by $SSE(s_0) = W\left(1 - \frac{W-1}{L+W}\right)^2 + (L+1)\left(-\frac{W}{L+W}\right)^2$ and $SSE(s_x) = W\left(1 - \frac{W-1}{L+W-1}\right)^2 + L\left(-\frac{W}{L+W-1}\right)^2 + 0.25$ (where the sub-database for which $x = 1$ yields $SSE = \frac{1}{4}$).

It follows that $\Delta(L,W,1,0)$ is equal to:

$$\frac{L^4 + L^3(4W-2) + L^2(2W^2+2W+1) + L(2W - 4W^3 + 6W^2) - 3W^4 + 2W^3 + 5W^2 - 4W}{4(L+W-1)^2(L+W)^2} \tag{1}$$

The denominator of expression (1) is positive. Let $a(L,W)$ denote the numerator. First, we observe that $a(L,L) = 4L(2L^2 + 2L - 1) > 0$.This establishes Part (iii), and will also be a useful benchmark for Parts (i) and (ii). Indeed, to prove that $a(L,W) > 0$ (and thus that $\Delta(L,W,1,0) > 0$) for $L > W$, we will consider the partial derivative of $a(L,W)$ relative to its first argument, and show that it is positive for $L \geq W$. (Clearly, $a(L,W)$ is a polynomial in its two arguments, and it is well-defined and smooth for all real values of (L,W).) To see this, observe that $\frac{\partial a(L,W)}{\partial L}$ is equal to:

$$4L^3 + (12W-6)L^2 + \left(4W^2 + 4W + 2\right)L + \left(-4W^3 + 6W^2 + 2W\right). \tag{2}$$

[18] Whenever needed, we use partial derivatives to derive inequalities. In doing so we obviously extend the definition of the function $\Delta(L,W,l,w)$ to all non-negative real numbers (L,W,l,w) by the function's algebraic formula, whenever well-defined.

Since $W > 2$ implies $12W - 6 > 0$, the only negative term in (2) is $-4W^3$. However, for $L \geq W$ it is true that $4LW^2 - 4W^3 \geq 0$ and thus, for $L \geq W$ we have $\frac{\partial a(L,W)}{\partial L} > 0$. Because, for $L \geq W$, $a(L, W)$ is strictly increasing in L and $a(L, L) > 0$, we also have $a(L, W) > 0$ for $L > W$.

We now turn to the case $L < W$, where expression (2) might be negative (and, indeed, will become negative if L is held fixed and $K \to \infty$.) Again the strategy of the proof is to use direct evaluation at a benchmark and partial derivative arguments beyond, though a few special cases will require attention. The benchmark we use is the case $W = L + 1$. Here direct calculations yield $a(L, L + 1) = -4L(2L^2 - 1) < 0$.

This time we consider the partial derivative of $a(L, W)$ w.r.t. to its second argument, and would like to establish that it is negative. If it were, increasing K from $(L + 1)$ further up would only result in lower values of $a(L, W)$, and therefore the negativity of $a(L, W)$ (and of $\Delta(L, W, 1, 0)$) for $L < W$ would be established.

Consider, then,

$$\begin{aligned}
\frac{\partial a(L,W)}{\partial W} &= 4L^3 + 4L^2W + 2L^2 - 12LW^2 + 12LW + 2L - 12W^3 + 6W^2 + 10W - 4 \\
&= 4L^3 + (4W + 2)L^2 + \left(12W - 12W^2 + 2\right)L + \left(6W^2 - 12W^3 + 10W - 4\right) \\
&< 4W^3 + (4W + 2)W^2 + 12W^2 + 2W - 12LW^2 + 6W^2 - 12W^3 + 10W - 4 \\
&< 4W^3 + (4W + 2)W^2 + 12W^2 + 2W + 6W^2 - 12W^3 + 10W - 4 \\
&= -4\left(-3W - 5W^2 + W^3 + 1\right) \qquad (3)
\end{aligned}$$

where the first inequality follows from the fact that $L < W$ and the second from the fact that $L, W > 0$.

We now observe that expression (3) is negative for $W \geq 6$, and thus the partial derivative $\frac{\partial a(L,W)}{\partial W}$ is indeed negative for all $W \geq 6$, $L < W$. Coupled with the fact that $a(L, L + 1) < 0$, we obtain $a(L, W) < 0$ for all $W \geq 6$ (and $2 < L < W$).

We now wish to show that $a(L, W) < 0$ holds also for lower values of W. However, as $W > L > 2$ only a few pairs of values (L, W) are possible: $(3, 4)$,$(3, 5)$,$(4, 5)$. Direct calculation shows that $a(L, W)$ is negative for all these pairs. Specifically, $a(3, 4) = -204$, $a(3, 5) = -1,424$, and $a(4, 5) = -496$. This concludes the proof of Parts (i) and (ii). $\square$

Proof of Proposition 2:

Let there be given $l > 1$. We wish to prove that for any $L, W > 2$, $\Delta(L, W, l, 0) < 0$ (where the case $l = 0$, $w > 1$ is obviously symmetric).

The SSE's are given by $SSE(s_0) = (L+l)\left(-\frac{W}{l+L+W-1}\right)^2 + W\left(1 - \frac{W-1}{l+L+W-1}\right)^2$ and $SSE(s_x) = L\left(-\frac{W}{L+W-1}\right)^2 + W\left(1 - \frac{W-1}{L+W-1}\right)^2$ (where the sub-database for which $x^j = 1$ yields $SSE = 0$). Straightforward calculation yields

$$\Delta(L, W, l, 0) = -Wl\frac{\left(L(W-2) + (W-1)^2\right)l + (L+W-1)(L(W-2) + W(W-1))}{(L+W-1)^2(l+L+W-1)^2}$$

which is clearly negative. □

For convenience, we prove Proposition 4 before Proposition 3.

Proof of Proposition 4

First, observe that if $w = \left[\frac{lW}{L}\right] = 0$, then the first result in part (i), namely that for databases $(L, W, l, 0)$ the unique SNE is the one selected by s_0, follows directly from Proposition 1. To prove the rest of the Proposition, it will be convenient to extend the definition of Δ to real-valued arguments and use calculus. We will only resort to (first- and second- order) partial derivatives with respect to the last two arguments. Note that for positive integers L, W, l, w, the SSE formulae are

$$SSE(s_0) = (L+l)\frac{(W+w)^2}{(L+W+l+w-1)^2} + (L+l)^2\frac{W+w}{(L+W+l+w-1)^2} \quad (4)$$

$$SSE(s_x) = LW\frac{L+W}{(L+W-1)^2} + lw\frac{l+w}{(l+w-1)^2}. \quad (5)$$

It is therefore natural to define, for positive integers L, W, and any $l, w \in \mathbb{R}$ such that $l + w \neq 1 - (L+W)$ and $w \neq 1 - l$,

$$\begin{aligned}\Delta(L, W, l, w) = {} & LW\frac{L+W}{(L+W-1)^2} + lw\frac{l+w}{(l+w-1)^2} \\ & - (L+l)\frac{(W+w)^2}{(L+W+l+w-1)^2} - (L+l)^2\frac{W+w}{(L+W+l+w-1)^2}\end{aligned}$$

Clearly, the function Δ is a rational function in its four arguments, and apart from these points of singularity, it is well-defined and smooth. Note that we are interested in l, w that are positive integers, hence $l, w \geq 1$. In

particular, $l+w \geq 2$ while $1-(L+W) < -3$ and $w \geq 1$ while $1-l \leq 0$, so that none of the two singular points of Δ is within or even on the boundary of the range of values that is of interest to the statement of the proposition, apart from the special case discussed in the first paragraph of this proof. However, these points will prove useful in analyzing the function.

Next, because our focus is on the behavior of Δ as we change its fourth argument, starting from the critical point $w = \frac{lW}{L}$, it will simplify notation if we shift the fourth variable to center it around that point. Formally, let $\omega \in R$ and define a function $b : \mathbb{Z}_+^2 \times \mathbb{R}^2 \to \mathbb{R}$ by $b(L,W,l,\omega) = \Delta\left(L,W,l,\frac{lW}{L}+\omega\right)$.

The statements in the Proposition are about the value of the $\Delta(\cdot)$ function evaluated at points where the third argument is a positive integer and the fourth argument is an integer larger or equal than $\left\lceil \frac{lW}{L} \right\rceil$. It is therefore useful to notice that for any positive integers L, W, l, and integer z we can write

$$\Delta\left(L,W,l,\left\lceil\frac{lW}{L}\right\rceil+z\right) = \Delta\left(L,W,l,\frac{lW}{L}+\varepsilon+z\right) = b(L,W,l,z+\varepsilon) \quad (6)$$

where $\varepsilon = \left\lceil \frac{lW}{L} \right\rceil - \frac{lW}{L}$. Note that $\varepsilon \in [-0.5, 0]$ if $\left\lceil \frac{lW}{L} \right\rceil = \left\lfloor \frac{lW}{L} \right\rfloor$ and $\varepsilon \in [0, 0.5)$ if $\left\lceil \frac{lW}{L} \right\rceil = \left\lceil \frac{lW}{L} \right\rceil$.

We prove the proposition as follows:

(1) We first show that $b(L,W,l,\omega)$ is strictly decreasing in ω for $\omega \geq 1$ (Lemma 1);

(2) Next, we prove that $b(L,W,l,\omega)$ has a limit as $\omega \to \infty$ and that it is a negative number (Lemma 2);

(3) Direct calculation shows that $b(L,W,l,1.5) > 0$, and from this we conclude that, as a function of ω, $b(L,W,l,\omega)$ has a unique root larger than 1.5 (Lemma 3);

(4) We prove that $b(L,W,l,\omega) > 0$ for $\omega \in [-0.5, 1.5]$ if $\left\lceil \frac{lW}{L} \right\rceil \geq 1$, and for $\omega \in [0.5, 1.5]$ if $\left\lceil \frac{lW}{L} \right\rceil = 0$ (Lemma 4);

(5) Next, we show that $\frac{\partial b(L,W,l,\omega)}{\partial l} > 0$ for $\omega \geq 2$ (Lemma 5);

(6) We then show that, for all $l' > l > 1$, $\widetilde{w} > \frac{l'W}{L}$, if $\Delta(L,W,l,\widetilde{w}) \geq 0$ then $\Delta(L,W,l',\widetilde{w}) \geq 0$ (Lemma 6).

Before we proceed to formally state and prove these lemmas, let us explain why they prove the result:

Part (i) follows from (4). For $\left\lceil \frac{lW}{L} \right\rceil \geq 1$ we need to show that (for all $L, W > 2$, $l > 0$), we have $\Delta(L,W,l,w), \Delta(L,W,l,w+1) > 0$. In

terms of the function b, $\Delta(L,W,l,w) = b(L,W,l,\varepsilon)$ and $\Delta(L,W,l,w+1) = b(L,W,l,\varepsilon+1)$. Thus we have to show that $b(L,W,l,\varepsilon), b(L,W,l,\varepsilon+1) > 0$ where $\varepsilon = \left[\frac{lW}{L}\right] - \frac{lW}{L} \in [-0.5, 0.5)$. Clearly, this follows from Lemma 4. Similarly, for $\left[\frac{lW}{L}\right] = 0$ we need to show that (for all $L, W > 2$, $l > 0$), we have $\Delta(L,W,l,w+1) = b(L,W,l,\varepsilon+1) > 0$, where $\varepsilon = \left[\frac{lW}{L}\right] - \frac{lW}{L} \in [-0.5, 0.5)$. Clearly, this also follows from Lemma 4.

Part (ii) follows from (1) and (3) because b is a smooth function of ω in the range $\omega \geq 1$.

Part (iii) follows from (6): If l' is such that $\left[\frac{l'W}{L}\right] \geq w^*(L,W,l) - 2$, the claim follows from the fact that $w^*(L,W,l') \geq \left[\frac{l'W}{L}\right] + 2$. Thus we focus on the case $\left[\frac{l'W}{L}\right] < w^*(L,W,l) - 2$.

Using part (i) and the definition of w^*, $\Delta(L,W,l,q) \geq 0$ for any integer q such that $0 \leq q \leq w^*(L,W,l) - 1$. Claim (6) implies that for the same values of q, $\Delta(L,W,l',q) \geq 0$. It follows that the smallest integer w'' ($w'' > \left[\frac{l'W}{L}\right]$) for which $\Delta(L,W,l',w'')$ becomes negative is greater or equal than $w^*(L,W,l)$ and thus $w^*(L,W,l') \geq w^*(L,W,l)$.

Finally, for Part (iv), assume that W/L is an integer, and consider integers $l' > l > 1$. Let $w = \left[\frac{lW}{L}\right]$ and $w' = \left[\frac{l'W}{L}\right]$, that is, $w = \frac{lW}{L}$ and $w' = \frac{l'W}{L}$ as these are integers. Lemma 5 implies that, if $b(L,W,l,\omega) = \Delta(L,W,l,w+\omega) > 0$ for $\omega \geq 2$, then $b(L,W,l',\omega) = \Delta(L,W,l',w'+\omega) > 0$ (for the same ω). It follows that the smallest integer $\omega > 1$ for which $\Delta(L,W,l',w'+\omega)$ becomes negative is bigger than that for which $\Delta(L,W,l,w+\omega)$ becomes negative, thus $w^*(L,W,l') - w' \geq w^*(L,W,l) - w$.

We start by providing the explicit formula for $b(L,W,l,\omega)$:

$$b(L,W,l,\omega) = \frac{LW(L+W)}{(L+W-1)^2} + \frac{l(lW+L\omega)[l(L+W)+L\omega]}{[lW+L(l+\omega-1)]^2} \qquad (7)$$
$$-\frac{(l+L)(lW+LW+L\omega)(lL+L^2+lW+LW+L\omega)}{(-L+lL+L^2+lW+LW+L\omega)^2}$$

This is a rational function in ω, with two vertical asymptotes where either the denominator of the first term or the denominator of the third term in 7

vanishes. We denote these singular points by $\underline{\omega}$ and $\overline{\omega}$, respectively:

$$\begin{aligned}\overline{\omega} &= 1-\frac{l\,(L+W)}{L}=1-l-\frac{lW}{L}<0\\ \underline{\omega} &= 1-\frac{(l+L)\,(L+W)}{L}<\overline{\omega}\end{aligned}$$

Thus, for $\omega>\overline{\omega}$, $b\,(L,W,l,\omega)$ is a smooth function.

We can now establish:

Lemma 1 *$b\,(L,W,l,\omega)$ is strictly decreasing in ω for $\omega\geq 1$.*

Proof: Differentiate $b\,(L,W,l,\omega)$ with respect to ω:

$$\begin{aligned}&\frac{\partial b\,(L,W,l,\omega)}{\partial\omega}=\frac{(2L(l+L)(lW+L(W+\omega))(l(L+W)+L(L+W+\omega)))}{(L^2+lW+L(-1+l+W+\omega))^3}\\ &-\frac{(L(l+L)(l(L+2W)+L(L+2(W+\omega))))}{(L^2+lW+L(-1+l+W+\omega))^2}\\ &+\frac{(lL^2(-2lW+l^2(L+W)+lL(-1+\omega)-2L\omega))}{(lW+L(-1+l+\omega))^3}\end{aligned}$$

The above expression can be rewritten as

$$-\frac{L^3\left[z_0(L,W,l)+z_1(L,W,l)\omega+z_2(L,W,l)\omega^2+z_3(L,W,l)\omega^3+z_4(L,W,l)\omega^4\right]}{(lW+L(l+\omega-1))^3(L^2+lW+L(l+W+\omega-1))^3} \tag{8}$$

where we define $z_0(L,W,l), z_1(L,W,l), z_2(L,W,l), z_3(L,W,l), z_4(L,W,l)$ as:

$$\begin{aligned}z_0(L,W,l) &= -2l^4(L-W)(L+W)^3-l^2L^2(L+W)^2(6+L(2L-9)-2W^2)\\ &\quad -2l^3L(L+W)^2(L(2L-3)-2W^2)+L^4\left[2W-L(L+W-1)\right]\\ &\quad +lL^3\left[L(2+3(L-2)L)+4W+6(L-2)LW+3(+L-2)W^2\right]\end{aligned}$$

$$z_1(L,W,l)=L\left\{\begin{array}{c}L^3\left[(2(l-1)^4+4(l-1)^3L+(3-4l+2l^2)L^2\right]\\ +W\left[\begin{array}{c}6(l-1)l(2-l+l^2)L^2+6(2l-1)(1-l+l^2)L^3\\ +3(1-2l+2l^2)L^4+6lL(l+L)(1+l^2+lL)W\\ +2l(l+L)(2l+l^2+L+lL)W^2\end{array}\right]\end{array}\right\}$$

$$z_2(L,W,l)=3L^2\left\{2l^3W^2+L\left[\begin{array}{c}(-2+4l-4l^2+2l^3)L+L^2\left[2-4l+3l^2+(l-1)L\right]\\ +[(4l(1-l+l^2)+2L+l(6l-4)L+(2l-1)L^2]W\\ +(3l^2+lL)W^2\end{array}\right]\right\}$$

$$z_3(L,W,l) = L^3\left[L^3 + 2l(3l-2)W + L^2(-4+6l+W) + L(6-8l+6l^2-2W+6lW)\right]$$

$$z_4(L,W,l) = L^4(-2+2l+L)$$

First, notice that L^3 and the denominator of expression (8) are strictly positive, hence the sign of (8) is equal to the opposite sign of the polynomial in ω on its numerator. Second, notice that $z_1(L,W,l)$, $z_2(L,W,l)$, $z_3(L,W,l)$, and $z_4(L,W,l)$ are strictly positive for all admissible values of $\{L,W,l\}$. It follows that the derivative of the polynomial in ω on the numerator of (8) is strictly positive for positive values of ω. Hence, if we can show that the polynomial is positive for some positive value of ω, then it is positive for all larger values of ω as well. Finally, we evaluate the polynomial at $\omega = 1$ and show that it is positive.

$$\begin{aligned} & z_0(L,W,l) + z_1(L,W,l)\,(1) + z_2(L,W,l)\,(1) + z_3(L,W,l)\,(1) + z_4(L,W,l)\,(1) \\ = \;& 2l(l+L)(L+W)^3[L^2 + l^2W + lL(2+W)] > 0 \end{aligned}$$

This allows us to conclude that $\frac{\partial b(L,W,l,\omega)}{\partial \omega} < 0$ for all $\omega \geq 1$. □□

Lemma 2 $\exists \lim_{\omega\to\infty} b\,(L,W,l,\omega) < 0$.

Proof:

$$\lim_{\omega\to\infty} b\,(L,W,l,\omega) = \frac{LW(L+W)}{(L+W-1)^2} + l - l - L = \frac{-L(L-1)^2 - (L-2)LW}{(L+W-1)^2} < 0. \;\square$$

Lemma 3 *$b\,(L,W,l,\omega)$ has exactly one root in $\omega \in (1.5,\infty)$.*

Proof: We know that the singular points of b are negative. This means that for $\omega \geq 0$, $b\,(L,W,l,\omega)$ is a smooth function. Further, algebraic calculations[19] show that $b\,(L,W,l,1.5) > 0$ for all $L,W > 2$, $l > 0$. Since we established that $b\,(L,W,l,\omega)$ is negative for ω large enough, it has to have a root at some $\omega > 1.5$. Further, it is unique because b is strictly decreasing in ω over this range. □

Lemma 4 *$b\,(L,W,l,\omega) > 0$ for $\omega \in [-0.5,1.5]$ if $\left[\frac{lW}{L}\right] \geq 1$, and for $\omega \in [0.5,1.5]$ if $\left[\frac{lW}{L}\right] = 0$.*

[19] See online appendix part (a).

Proof: We need to consider two cases.

Case 1: $l = 1$

In this case, the vertical asymptotes are at $\underline{w} = -\frac{W}{L} - (W + L)$ and $\overline{\omega} = -\frac{W}{L}$ so for $\omega \geq -\frac{W}{L}$ the function is smooth. Algebraic calculations[20] show that for $l = 1$ and for all $L, W > 2$, $\frac{\partial b(L,W,l,\omega)}{\partial \omega}$ is strictly negative for all $\omega \geq -\frac{W}{L}$. This, together with the fact that $b(L, W, l, 1.5) > 0$, proves that $b(L, W, l, \omega) > 0$ for $\omega \in (-\frac{W}{L}, 1, 5]$. If $\left[\frac{lW}{L}\right] \geq 1$, the fact that $-\frac{W}{L} < 0.5$ proves that $b(L, W, l, \omega) > 0$ for $\omega \in [-0.5, 1.5]$. Similarly, if $\left[\frac{lW}{L}\right] = 0$ the fact that $-\frac{W}{L} < 0$ proves that $b(L, W, l, \omega) > 0$ for $\omega \in [0.5, 1.5]$.

Case 2: $l > 1$

Algebraic calculations[21] show that $b(L, W, l, -0.5) > 0$ for all $l > 1$, $L, W > 2$ such that $\left[\frac{lW}{L}\right] \geq 1$, and that $b(L, W, l, 0.5) > 0$ for all $l > 1$, $L, W > 2$ such that $\left[\frac{lW}{L}\right] = 0$. Consider first the case $\left[\frac{lW}{L}\right] \geq 1$. To study the sign of $b(L, W, l, \omega)$ for $\omega \in [-0.5, 1.5]$ we observe that it is positive at $\omega = -0.5$ and at $\omega = 1.5$, and that it is continuous on the interval. Thus, to prove that it is positive throughout the interval it suffices to show that it has no roots in it.

Observe that $b(L, W, l, \omega)$ is a rational function in ω with a fourth degree polynomial (in ω) in its numerator. Every root of b is a root of this polynomial, and thus b can have at most four real roots. We claim that it has at least one real root in each of the following intervals:

(a) $(\underline{\omega}, \overline{\omega})$, (b) $(\overline{\omega}, -0.5)$, (c) $(1.5, \infty)$.

To see that there is a root in (a), observe that

$$\lim_{\omega \to_+ \overline{\omega}} b(L, W, l, \omega) = \lim_{\omega \to_- \overline{\omega}} b(L, W, l, \omega)$$

$$= \frac{LW(L+W)}{(L+W-1)^2} - \frac{L^2 l(l-1)}{0} - \frac{L(L+l)(L+LW-Ll)(L+W+1)}{L^2(L+W)^2} = -\infty$$

$$\lim_{\omega \to_+ \underline{\omega}} b(L, W, l, \omega) = \frac{LW(L+W)}{(L+W-1)^2} + \frac{l[-L(L+W+l-1)][-L(L+W-1)]}{L^2(L+W)^2}$$

$$- \frac{-L^2[l(L+2l-1)+l(l-1)]}{0^+} = +\infty$$

Thus, b, which is continuous over $(\underline{\omega}, \overline{\omega})$, goes from $+\infty$ to $-\infty$ and has to cross 0 over the interval.

[20] See online appendix part (c).

[21] See online appendix part (b).

As for interval (b), observe, again, that $\lim_{\omega \to +\overline{\omega}} b(L,W,l,\omega) = -\infty$ and that $b(L,W,l,-0.5) > 0$. Finally, it was established in Lemma 3 that there is a root in (c).

We can now consider the interval of interest, $[-0.5, 1.5]$. We know that b is positive at the two endpoints. If it were non-positive at some point over this interval, the numerator of b would have to have two roots in the interval – either two distinct roots or a multiple one. In either case, we would have a total of five real roots for a polynomial of degree 4, which is impossible, and thus we conclude that b is strictly positive throughout $[-0.5, 1.5]$.

Next, consider the case $\left[\frac{lW}{L}\right] = 0$. We need to study the sign of $b(L,W,l,\omega)$ for $\omega \in [0.5, 1.5]$. The proof is analogous the the one for the previous case. In particular, it has been shown that $b(L,W,l,\omega) > 0$ at the two endpoints of the interval and continuous over the interval. Moreover, $b(\cdot)$ has at least one real root in each of the following intervals: (a) $(\underline{\omega}, \overline{\omega})$, (b)' $(\overline{\omega}, 0.5)$, (c) $(1.5, \infty)$. Since the numerator of $b(\cdot)$ can have at most four real roots, there are no roots in the interval $\omega \in [0.5, 1.5]$ and the function is positive over the whole interval.□

Lemma 5 *$b(L,W,l,\omega)$ is strictly increasing in l for $\omega \geq 2$.*

Proof: The derivative of $b(L,W,l,\omega)$ w.r.t. l is:

$$L^3 \frac{\zeta_0(L,W,\omega) + \zeta_1(L,W,\omega)\, l + \zeta_2(L,W,\omega)\, l^2 + \zeta_3(L,W,\omega)\, l^3}{(-L + lL + lW + L\omega)^3(-L + lL + L^2 + lW + LW + L\omega)^3} \tag{9}$$

where $\zeta_0(L,W,\omega)$, $\zeta_1(L,W,\omega)$, $\zeta_2(L,W,\omega)$, $\zeta_3(L,W,\omega)$ are defined as:

$$\zeta_0(L,W,\omega) = L^3(\omega - 1)\begin{pmatrix} L^3\omega^2 + W(4(\omega-1)^2\omega + W^2(2\omega-1) + 3W(1-3\omega+2\omega^2)) \\ +L^2(3(\omega-1)\omega^2 + W(2\omega(1+\omega)-1)) \\ +L\begin{pmatrix} 2(\omega-1)^2\omega(1+\omega) + W^2(\omega(4+\omega)-2) \\ +3W(1+\omega(-3+\omega+\omega^2)) \end{pmatrix} \end{pmatrix}$$

$$\zeta_1(L,W,\omega) = L^2\begin{pmatrix} W^2(12W(\omega-1)^2 + W^2(2\omega-3) + 6(\omega-1)^2(2\omega-1)) \\ +L^4(\omega-2)\omega + 3L^2(2(\omega-1)^2\omega^2 + 4W(\omega-1)^2(1+\omega) + W^2(\omega^2-3)) \\ +LW(-6 + 6W(\omega-1)^2(4+\omega) + 6\omega(4-4\omega+\omega^3) + W^2(-9+\omega(4+\omega))) \\ +L^3(6(\omega-1)^2\omega + W(-3+\omega(3\omega-4))) \end{pmatrix}$$

$$\zeta_2(L,W,\omega) = 3L\left(L+W\right)^2 \begin{pmatrix} L\left(L\left(\omega-2\right)+2\left(\omega-1\right)^2\right)\omega \\ +W^2\left(2\omega-3\right)+W\left(4\left(\omega-1\right)^2+L\left(\omega^2-3\right)\right) \end{pmatrix}$$

$$\zeta_3(L,W,\omega) = 2(L+W)^3(L(\omega-2)\omega+W(2\omega-3))$$

First, notice that L^3 and the denominator of expression (9) are strictly positive. Second, notice that $\zeta_0\left(L,W,\omega\right)$, $\zeta_1\left(L,W,\omega\right)$, $\zeta_2\left(L,W,\omega\right)$, $\zeta_3\left(L,W,\omega\right)$ are strictly positive for all admissible values of $\{L,W\}$ and $\omega \geq 2$. Since l is an integer, it follows that the polynomial in l on the numerator of (9) is strictly positive for $\omega \geq 2$. This allows us to conclude that $\frac{\partial b(L,W,l,\omega)}{\partial \omega} > 0$ for all $\omega \geq 2$. □

Lemma 6 *For all* $l' > l > 1$, $\widetilde{w} > \frac{l'W}{L}$, *if* $\Delta\left(L,W,l,\widetilde{w}\right) \geq 0$ *then* $\Delta\left(L,W,l',\widetilde{w}\right) \geq 0$.

Proof: If $\widetilde{w} = \left[\frac{l'W}{L}\right]$ or $\widetilde{w} = \left[\frac{l'W}{L}\right]+1$, the conclusion $\Delta\left(L,W,l',\widetilde{w}\right) \geq 0$ follows from either Part (i) or Proposition 1.

Assume, then, that $\widetilde{w} \geq \left[\frac{l'W}{L}\right]+2 \geq \left[\frac{lW}{L}\right]+2$. Recall that $w = \left[\frac{lW}{L}\right]$ with $\varepsilon = \left[\frac{lW}{L}\right] - \frac{lW}{L}$ and denote $w' = \left[\frac{l'W}{L}\right], \varepsilon' = \left[\frac{l'W}{L}\right] - \frac{l'W}{L}$. Next, let $\omega = \widetilde{w} - w$ and $\omega' = \widetilde{w} - w'$. Thus, $\widetilde{w} = w+\omega = lW/L+\varepsilon+\omega = w'+\omega' = l'W/L+\varepsilon'+\omega'$.

Clearly, as $l' > l$, we have $w' \geq w$ and therefore $\varepsilon'+\omega' \leq \varepsilon+\omega$. Note that $\omega, \omega' \geq 2$ and thus $\omega+\varepsilon, \omega'+\varepsilon' \geq 1$.

We assume that $\Delta\left(L,W,l,\widetilde{w}\right) = \Delta\left(L,W,l,w+\omega\right) = b\left(L,W,l,\omega+\varepsilon\right) \geq 0$ and need to show $\Delta\left(L,W,l',\widetilde{w}\right) = \Delta\left(L,W,l',w'+\omega'\right) = b\left(L,W,l',\omega'+\varepsilon'\right) \geq 0$. Indeed, $b\left(L,W,l,\omega+\varepsilon\right) \geq 0$, coupled with Lemma 5, implies that $b\left(L,W,l',\omega+\varepsilon\right) \geq 0$. Further, as $\omega'+\varepsilon' \leq \omega+\varepsilon$, Lemma 1 (with $\omega+\varepsilon, \omega'+\varepsilon' \geq 1$) implies that $b\left(L,W,l',\omega'+\varepsilon'\right) \geq 0$, which completes the proof of the lemma. □□

Proof of Proposition 3

The proof relies on the analysis used to prove Proposition 4. Here, we prove only the first statement. The second holds by symmetry of the Δ function.

Let us denote by $\overline{w}$ the closest integer to $\frac{W}{L}$ ($= \frac{lW}{L}$ because we deal with the case $l = 1$), that is, $\overline{w} = \left[\frac{W}{L}\right]$.

We need to show that, for every $0 < w \leq \left\lfloor \frac{W}{L} \right\rfloor + 1$, $\Delta\left(L,W,1,w\right) > 0$.

In (6) we had

$$\Delta\left(L,W,l,\left[\frac{lW}{L}\right]+z\right) = \Delta\left(L,W,l,\frac{lW}{L}+\varepsilon+z\right) = b\left(L,W,l,z+\varepsilon\right)$$

which, by setting $l = 1$, becomes

$$\Delta\left(L, W, 1, \left[\frac{W}{L}\right] + z\right) = \Delta\left(L, W, 1, \frac{W}{L} + \varepsilon + z\right) = b\left(L, W, 1, z + \varepsilon\right)$$

For $0 < w \leq \left\lfloor \frac{W}{L} \right\rfloor + 1$, denoting $z = w - \overline{w}$ we have $w = \overline{w} + z = \left[\frac{W}{L}\right] + z$. We can then write

$$\Delta\left(L, W, 1, w\right) = \Delta\left(L, W, 1, \left[\frac{W}{L}\right] + z\right) = \Delta\left(L, W, 1, \frac{W}{L} + \varepsilon + z\right) = b\left(L, W, 1, z + \varepsilon\right)$$

where $\varepsilon = \left[\frac{W}{L}\right] - \frac{W}{L} \in [-0.5, 0.5)$ and $z \in \{1 - \left[\frac{W}{L}\right], ..., 1\}$ if $\left[\frac{W}{L}\right] = \left\lfloor \frac{W}{L} \right\rfloor$ and $z \in \{1 - \left[\frac{W}{L}\right], ..., 0\}$ if $\left[\frac{W}{L}\right] = \left\lfloor \frac{W}{L} \right\rfloor + 1$.

Denoting the fourth argument of b by $\omega = z + \varepsilon$, we observe that, because $z \geq 1 - \left[\frac{W}{L}\right]$, $\omega \geq 1 - \frac{W}{L}$. Further, as $z \leq 1$ and $\varepsilon < 0.5$, $\omega < 1.5$. Thus, it suffices to show that $b\left(L, W, 1, \omega\right) > 0$ for $\omega \in \left[-\frac{W}{L} + 1, 1.5\right]$. We know that $b\left(L, W, 1, \omega\right)$ is continuous and differentiable for $\omega > -\frac{W}{L}$, that $\frac{\partial b(W,L,1,\omega)}{\partial \omega} < 0$ for all $\omega \geq -\frac{W}{L}$, and that $b\left(L, W, 1, 1.5\right) > 0$. Therefore, $b\left(L, W, 1, \omega\right) > 0$ for all $\omega \in \left[-\frac{W}{L} + 1, 1.5\right]$. This concludes the proof. □

Proof of Proposition 5

Consider the vector $\left(\frac{W}{W+L}, \frac{w}{w+l}\right)$ in the square $[0, 1]^2$ as depicted in Figure 1. If $\left(\frac{W}{W+L}, \frac{w}{w+l}\right)$ is near the diagonal, the optimal empirical similarity is s_0, and it is s_x when $\frac{W}{W+L}$ and $\frac{w}{w+l}$ differ significantly (where the exact bound depends both on where they are on the diagonal and on t, which is not graphically represented in the Figure). We wish to focus on the optimal choice for a player who has observed $(W, L, w, l) = (W_t, L_t, w_t, l_t)$. Assume that the player were to use the similarity s_x and therefore to compute empirical frequencies of $y_\tau = 1$ $(\tau < t)$ separately for $x_t = 0$ and $x_t = 1$. The choice $a^h = 1$ is optimal for $x_t = 0$ iff $\frac{W}{W+L} \geq \frac{c}{1+c-d}$, and for $x_t = 1$ $a^h = 1$ is optimal for $x_t = 0$ iff $\frac{w}{w+l} \geq \frac{c}{1+c-d}$.

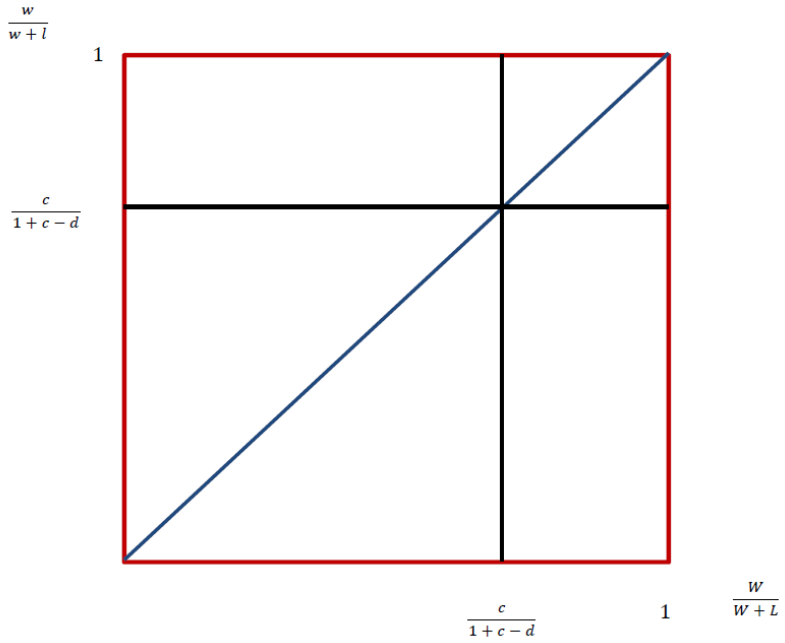

Figure 1

Define auxiliary random variables A_t, D_t as follows:

$$\begin{aligned} A_t &= (1-d)W_t - cL_t \\ D_t &= (1-d)w_t - cl_t \end{aligned}$$

$A_t/(W_t + L_t)$ is the difference between the expected payoff of $a^h = 1$ and of $a^h = 0$ for a player who believes that $y_t = 1$ with probability $\frac{W_t}{W_t+L_t}$. Indeed, $A_t \geq 0$ iff $\frac{W_t}{W_t+L_t} \geq \frac{c}{1+c-d}$. Similarly, D_t is the corresponding difference for the probability $\frac{W_t}{W_t+L_t}$.

Consider first the fictitious auxiliary process $(A_t, D_t)_t$ that would correspond to the assumption that the players always use s_x, that is, that they compute empirical similarities for $x_t = 0$ and for $x_t = 1$ separately regardless of (W_t, L_t, w_t, l_t). In this case, we would have a two-dimensional biased random walk, where, for each t, with probability β only D_t changes its value, and we would have

$$A_{t+1} = A_t$$
$$D_{t+1} = \begin{cases} D_t + (1-d) & 1-\varepsilon \\ D_t - c & \varepsilon \end{cases} \qquad if \quad D_t \geq 0$$
$$D_{t+1} = \begin{cases} D_t + (1-d) & \varepsilon \\ D_t - c & 1-\varepsilon \end{cases} \qquad if \quad D_t < 0$$

and with probability $(1-\beta)$ – only A_t changes and we would have

$$A_{t+1} = \begin{cases} A_t + (1-d) & 1-\varepsilon \\ A_t - c & \varepsilon \end{cases} \qquad if \quad A_t \geq 0$$
$$A_{t+1} = \begin{cases} A_t + (1-d) & \varepsilon \\ A_t - c & 1-\varepsilon \end{cases} \qquad if \quad A_t < 0$$
$$D_{t+1} = D_t.$$

Thus, if we condition A_t on the periods in which it is active ($x_t = 0$), it is a Markov process, where, on the non-negative reals it is the sum of i.i.d. variables

$$z_t = \begin{cases} 1-d & 1-\varepsilon \\ -c & \varepsilon \end{cases}$$

which have strictly positive expectation, and on the negative reals it is the sum of

$$v_t = \begin{cases} 1-d & \varepsilon \\ -c & 1-\varepsilon \end{cases}$$

By standard arguments,

(i) With probability 1 both processes will change values infinitely often, and with fixed relative frequencies of $(1-\beta, \beta)$;

(ii) With probability 1 each process will cross 0 only finitely many times, converging to ∞ or to $-\infty$, each with positive probability.

We now consider the actual process, in which the players do not optimize relative to or to $\frac{w_t}{w_t+l_t}$, but relative to one of these (depending on x_t) or relative to $\frac{W_t+w_t}{W_t+w_t+L_t+l_t}$, where the latter choice depends on the similarity function that obtains the lower SSE (i.e., on $\Delta(L_t, W_t, l_t, w_t)$). Fix $\delta > 0$ (which we will later shrink to zero). Consider a band of width 2δ around the diagonal, as depicted in Figure 2.

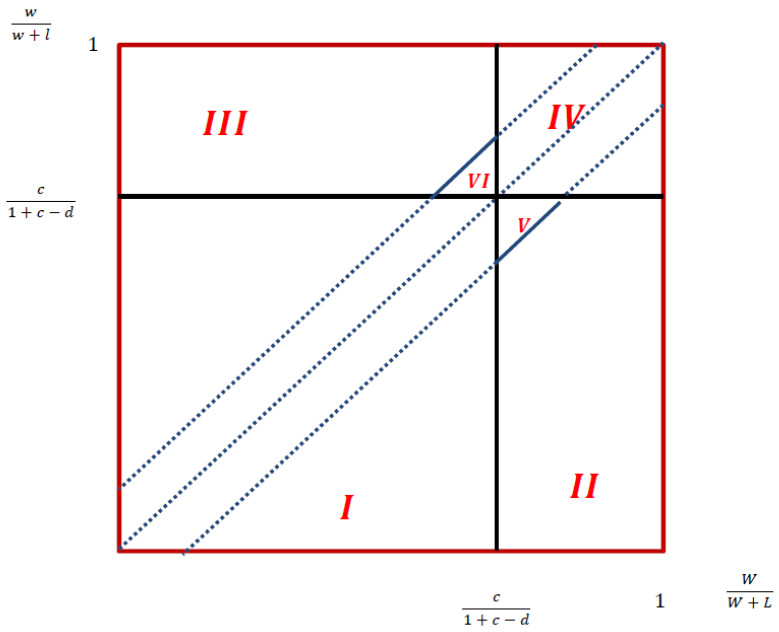

Figure 2

In the region marked at I, where (dropping the t subscript) $\frac{W}{W+L}, \frac{w}{w+l} < \frac{c}{1+c-d}$, we have $A, D < 0$. Notice that this region is defined independently of the relation between $\frac{W}{W+L}$ and $\frac{w}{w+l}$ and of δ. The optimal choice for the players in this region is $a^h = 0$ irrespective of the similarity function, as $\frac{W+w}{W+w+L+l}$ is a weighted average of $\frac{W}{W+L}$ and $\frac{w}{w+l}$. As long as $A_t, D_t < 0$, the process thus behaves as the auxiliary process. Importantly, for any $A_t, D_t < 0$ the process has a positive probability, bounded away from 0, of remaining negative ($A_\tau, D_\tau < 0$ for all $\tau \geq t$) and this is true also of the actual process. In this event, $\left(\frac{W_t}{W_t+L_t}, \frac{w_t}{w_t+l_t}\right) \to (\varepsilon, \varepsilon)$. In a completely symmetric way, the process can only leave region IV finitely many times with probability 1: either it leaves it forever from some t on, or stays there forever, with $\left(\frac{W_t}{W_t+L_t}, \frac{w_t}{w_t+l_t}\right) \to (1-\varepsilon, 1-\varepsilon)$.

Next, consider regions II_δ and III_δ. They are defined by the optimal choice for the players in each subhistory, as well as by a distance from the diagonal. Specifically, region II_δ would correspond to $\frac{W_t}{W_t+L_t} > \frac{c}{1+c-d} > \frac{w_t}{w_t+l_t}$ and $\frac{W_t}{W_t+L_t} > \frac{w_t}{w_t+l_t} + \delta$. For large enough t, the latter inequality implies that $\Delta(L_t, W_t, l_t, w_t) < 0$ and that the optimal empirical similarity is s_x. Therefore, in this region the process again behaves as the auxiliary process.

This implies that, again, with probability 1 the process will leave region II_δ only finitely many times; it will either leave it forever or stay there from some point on, with $\left(\frac{W_t}{W_t+L_t}, \frac{w_t}{w_t+l_t}\right) \to (1-\varepsilon, \varepsilon)$. Symmetrical arguments apply to region III_δ.

Consider now a converging sequence of δ's, say, $\delta_k = \delta/2^k$. The intersection of all these regions ($\cap V_{\delta_k}$ as well as $\cap VI_{\delta_k}$) is empty and thus, with probability 1, the process will leave them forever at some point. This establishes Part A of the Proposition (including the claim that each of the four limit matrices can be obtained with positive probability).

We now turn to Part (B). It is immediate that, should the process converge to matrices II or III, the optimal similarity will be s_x. Consider, for example, convergence to the matrix IV. We know that both $\frac{W_t}{W_t+L_t}$ and $\frac{w_t}{w_t+l_t}$ converge to $1-\varepsilon$, but can they differ from each other, along the way to the limit, so justify s_x as the optimal empirical similarity function?

To analyze this case, we consider the SSE formulae 4 and 5. We can approximate (W_t, L_t, w_t, l_t) by $((1-\beta)(1-\varepsilon)t, (1-\beta)\varepsilon t, \beta(1-\varepsilon)t, \beta\varepsilon t)$ and observe that

$$SSE(s_0) \simeq \frac{\varepsilon(1-\varepsilon)^2 t^3}{(t-1)^2}$$

and

$$SSE(s_x) \simeq \frac{\varepsilon(1-\varepsilon)^2 \beta^3 t^3}{(\beta t-1)^2} + \frac{\varepsilon(1-\varepsilon)^2(1-\beta)^3 t^3}{((1-\beta)t-1)^2}$$

So that

$$SSE(s_0) \simeq \varepsilon(1-\varepsilon)^2 \frac{1}{1-\frac{2}{t}+\frac{1}{t^2}}$$

and

$$SSE(s_x) \simeq \varepsilon(1-\varepsilon)^2 \left[\frac{1}{1-\frac{2}{\beta t}+\frac{1}{\beta^2 t^2}} + \frac{1}{1-\frac{2}{(1-\beta)t}+\frac{1}{(1-\beta)^2 t^2}}\right]$$

It follows that $SSE(s_0) < SSE(s_x)$.

Thus we establish the intuitive result that, when the play of the game is identical, at the limit there is no sunspot.□

References

[1] Akaike, H. (1954), "An Approximation to the Density Function", *Annals of the Institute of Statistical Mathematics*, **6**: 127-132.

[2] Argenziano, R. and I. Gilboa (2012), "History as a Coordination Device", *Theory and Decision*, **73**: 501-512.

[3] Argenziano, R. and I. Gilboa (2019), "Learning What is Similar: Precedents and Equilibrium Selection", *PNAS,* **116**.

[4] Aumann, R. (1974), "Subjectivity and Correlation in Randomized Strategies", *Journal of Mathematical Economics*, **1**: 67-96.

[5] Billot, A., I. Gilboa, D. Samet, and D. Schmeidler (2005), "Probabilities as Similarity-Weighted Frequencies", *Econometrica*, **73**: 1125-1136.

[6] Carlsson, H. and Van Damme, E. (1993), "Global Games and Equilibrium Selection", *Econometrica*, **61**: 989-1018.

[7] Cass, D. and K. Shell (1983), "Do Sunspots Matter?", *Journal of Political Economy*, **91**: 193-228.

[8] Edmond, C. (2013) "Information Manipulation, Coordination, and Regime Change", *Review of Economic Studies*, **80**: 1422-1458.

[9] Gilboa, I., O. Lieberman, and D. Schmeidler (2006), "Empirical Similarity", *Review of Economics and Statistics*, **88**: 433-444.

[10] Halaburda, H., Jullien, B., & Yehezkel, Y. (2020). Dynamic competition with network externalities: how history matters. *The RAND Journal of Economics*, 51(1), 3-31.

[11] Hardlë, W. and J. S. Marron (1985), "Optimal Bandwidth Selection in Non-Parametric Regression Function Estimation", *The Annals of Statistics*, **13**: 1465-1481.

[12] Harsanyi, J. C. and R. Selten (1988), *A General Theory of Equilibrium Selection in Games*. Cambridge: MIT Press.

[13] Hume, D. (1748), *An Enquiry Concerning Human Understanding*. Oxford: Clarendon Press.

[14] Kets, W. and A. Sandroni (2021), "A Theory of Strategic Uncertainty and Cultural Diversity", *Review of Economic Studies*, **88**: 287-333.

[15] Medin, D. L. and M. M. Schaffer (1978), "Context Theory of Classification Learning", *Psychological Review*, **85**: 207-238.

[16] Mosley, L. (2005), "Dropping Zeros, Gaining Credibility? Currency Redenomination in Developing Nations." *mimeo.*

[17] Nagel, R. (1995), "Unraveling in Guessing Games: An Experimental Study", *American Economic Review*, **85**: 1313-1326.

[18] Nosofsky, R. M. (1984), "Choice, Similarity, and the Context Theory of Classification", *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **10**: 104-114.

[19] Nosofsky, R. M. (1986), "Attention, similarity, and the identification–categorization relationship." *Journal of experimental psychology: General* **115**: 39-57.

[20] Nosofsky, R. M. (1988), "Exemplar-Based Accounts of Relations Between Classification, Recognition, and Typicality", *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **14**: 700-708.

[21] Nosofsky, R. M. (1991), "Tests of an exemplar model for relating perceptual classification and recognition memory." *Journal of experimental psychology: human perception and performance,* **17**: 3-27.

[22] Nosofsky, R. M. (2011), "The Generalized Context Model: An Exemplar Model of Classification", in *Formal Approaches in Categorization*, Cambridge University Press, New York, Chapter 2, 18-39.

[23] Parzen, E. (1962), "On the Estimation of a Probability Density Function and the Mode", *Annals of Mathematical Statistics*, **33**: 1065-1076.

[24] Rosenblatt, M. (1956), "Remarks on Some Nonparametric Estimates of a Density Function", *Annals of Mathematical Statistics*, **27**: 832-837.

[25] Rosenthal, R. W. (1973), "A Class of Games Possessing Pure-Strategy Nash Equilibria", *International Journal of Game Theory*, **2**: 65–67.

[26] Schelling, T. C. (1960), *The Strategy of Conflict*. Cambridge: Harvard University Press

[27] Schmeidler, D. (1973), "Equilibrium Points of Nonatomic Games", *Journal of Statistical Physics*, **7**: 295-300.

[28] Selten, R. (1970), *Preispolitik der Mehrproduktenunternehmung in der Statischen Theorie* (First ed.). Springer Verlag, Berlin.

[29] Shepard, R. N. (1957), "Stimulus and Response Generalization: A Stochastic Model Relating Generalization to Distance in Psychological Space", *Psychometrika*, **22**: 325-345

[30] Shepard RN (1987), "Toward a universal law of generalization for psychological science", *Science* **237**:1317–1323.

[31] Silverman, B. W. (1986), *Density Estimation for Statistics and Data Analysis*. London and New York: Chapman and Hall.

[32] Stahl, D. O. and P. W. Wilson (1995), "On Players' Models of Other Players: Theory and Experimental Evidence", *Games and Economic Behavior*, **10**: 213-254.

[33] Steiner, J., and C. Stewart, C. (2008), "Contagion through Learning", *Theoretical Economics*, **3**: 431-458.

[34] Sugden, R. (1995), "A Theory of Focal Points", *The Economic Journal*, **105**: 533–550.

Online Appendix for
"Similarity Nash Equilibria in Statistical Games"
by Rossella Argenziano and Itzhak Gilboa

a) Calculation of $b(L,W,l,1.5)>0$.

We evaluate the function $b(L,W,l,\omega)$ at $\omega=1.5$ and find the following expression:

$$\frac{L*g(L,W,l)}{(-1+L+W)^2(L+2lL+2lW)^2(L+2lL+2L^2+2lW+2LW)^2}$$

where $g(L,W,l)$ can be expressed as a polynomial in W :

$$\begin{aligned}
& g(L,W,l) \\
= \; & (32l^4+64l^3L+32l^2L^2)W^5 \\
& +16l\left[2L^3+l^3(8L-1)+2l^2L(1+8L)+lL^2(5+8L)\right]W^4 \\
& +4lL\left[12l^3(4L-1)+3lL^2(21+16L)+L^2(4+27L)+8l^2(-1+3L+12L^2)\right]W^3 \\
& +2L^2\left[\begin{array}{c}-3(L-2)L^2+8l^4(8L-3)+16l^3(-2+3L+8L^2)+\\ 2l^2(-6-6L+69L^2+32L^3)+2lL(6-L+33L^2)\end{array}\right]W^2 \\
& +\left[\begin{array}{c}16l^4(2L-1)+3L(2+5L-4L^2)+32l^3(-1+L+2L^2)\\ +4l^2(-3-12L+29L^2+8L^3)+4l(-6+15L-14L^2+17L^3)\end{array}\right]L^3W \\
& -3L^4(L-1)^2(3+2L)+12l^2L^4(L-1)^2+12lL^4(L-1)^3
\end{aligned}$$

Notice that for $W,L>2$ and $l>0$ the terms multiplying W^5, W^4, and W^3 are positive. The terms multiplying W^2 and L^3W and the constant are polynomials in l. For $l>0$, all three are increasing in l, as the coefficients of the positive powers of l are positive. Moreover, all three are positive when evaluated are $l=1$, hence for all $l>1$ as well. . In particular, the coefficient of W^2 evaluated at $l=1$ is equal to $-68+112L+270L^2+127L^3>0$. The coefficient of L^3W evaluated at $l=1$ is equal to $-84+82L+139L^2+88L^3>0$. Finally, the constant evaluated at $l=1$ is equal to $3L^4(2L-3)(L-1)^2>0$.

We have proved that $g(L,W,l)>0$. Since $\frac{L}{(-1+L+W)^2(L+2lL+2lW)^2(L+2lL+2L^2+2lW+2LW)^2}>0$, this concludes the proof.

b) Calculation of $b(L,W,l,-0.5)>0$ **for** $l>1$ **and** $\left[\frac{lW}{L}\right]\geq 1$.

We evaluate the function $b(L,W,l,\omega)$ at $\omega=-0.5$ and find the following

expression:

$$\frac{-L * h\left(L,W,l\right)}{(-1+L+W)^2(-3L+2lL+2lW)^2(-3L+2lL+2L^2+2lW+2LW)^2}$$

where $h\left(L,W,l\right)$ can be expressed as a polynomial in L :

$$\begin{aligned}
& h\left(L,W,l\right) \\
= \; & (20l-18)\,L^7+\left[-4l^2(8W-5)+l(-76+92W)+45-36W\right]L^6 \\
& +\left[-36+4(23-10l)l+(81-4l(90+l(-75+16l)))W-2(9-78l+64l^2)W^2\right]L^5 \\
& +\left[\begin{array}{c} 9-36l+20l^2+(-54+388l-528l^2+224l^3-32l^4)W \\ +(36-492l+780l^2-256l^3)W^2+(116l-192l^2)W^3 \end{array}\right]L^4 \\
& +\left[\begin{array}{c} -4l(18-43l+24l^2-4l^3)-4l(-74+234l-168l^2+32l^3)W \\ -4l(52-185l+96l^2)W^2-4l(32l-8)W^3 \end{array}\right]WL^3 \\
& -8l^2W^2\left[-19+56W-30W^2+4W^3+l^2(-6+24W)+l(24-84W+32W^2)\right]L^2 \\
& -16l^3W^3\left[6-3l+(8l-14)W+4W^2\right]L-16l^4W^4(2W-1)
\end{aligned}$$

In what follows, we prove that $h\left(L,W,l\right)<0$ for all $l>0$ and $L,W>2$. The constant term is negative. The coefficient of L is negative because it is the product of a negative term and a quadratic expression in W with a positive coefficient on the square which is positive and increasing at $W=2$, hence for any larger W too. Similarly, the coefficient of L^2 is negative because it is the product of a negative term and a quadratic expression in l with a positive coefficient on the square which is positive and increasing at $l=2$, hence for any larger l too.

The coefficient of L^3 is the product of W, which is positive, and a third degree polynomial in W which can be shown to be negative in the relevant range. In particular, the polynomial has a negative coefficient on the third and second power. At $W=2$, this polynomial is equal to $-56l+236l^2-288l^3-240l^4$ which is negative for all $l>1$. Moreover, its derivative at $W=2$ is equal to $-152l+488l^2-864l^3-128l^4$ which is also negative for all $l>1$. Finally, the fact that this derivative is negative $W=2$ implies that it is also negative for all values of $W>2$, because the negative coefficients on the third and second powers of W guarantee that the function is concave in W for positive W.

The coefficient of L^4 is a third degree polynomial in W which can be shown

to be negative in the relevant range ($l > 1$, $W > 2$). The polynomial has a negative coefficient on the third power. Evaluated at $W = 2$, it takes value $45 - 300l + 548l^2 - 576l^3 - 64l^4 < 0$ for all $l > 1$. Moreover, its derivative w.r.t. W evaluated at $W = 2$ is equal to $90 - 188l + 288l^2 - 800l^3 - 32l^4$ which is also negative for all $l > 1$. Finally, its second derivative w.r.t. W is equal to $-8\left(-9 + 123l - 195l^2 + 64l^3 + (144l - 87)lW\right)$ which is negative at $W = 2$ and decreasing in W for all positive values of W.

The coefficient of L^5 is a quadratic function of W with a negative coefficient on the square, which is negative and decreasing at $W = 3$, hence negative for all larger values of W too. The coefficient of L^6 is a quadratic function of l with a negative coefficient on the square, which is positive for $l = 2$ and negative for all larger values of l. The coefficient of L^7 is positive.

Since the coefficient L^7 is positive, and we want to prove that the whole polynomial in L is negative, we prove that the sum of the terms in L^7 and L^5 is negative.

First, notice that the condition $\frac{lW}{L} \geq \frac{1}{2}$ implies that $L \leq 2lW$, which in turn implies:

$$(20l - 18)\, L^7 < 4\,(20l - 18)\, L^5 l^2 W^2$$

which in turn implies that

$$\begin{aligned}
& (20l - 18)\, L^7 + \left[\begin{array}{c} -36 + 4(23 - 10l)l \\ +(81 - 4l(90 + l(-75 + 16l)))W - 2(9 - 78l + 64l^2)W^2 \end{array}\right] L^5 \\
< \;& 4\,(20l - 18)\, L^5 l^2 W^2 + \left[\begin{array}{c} -36 + 4(23 - 10l)l \\ +(81 - 4l(90 + l(-75 + 16l)))W - 2(9 - 78l + 64l^2)W^2 \end{array}\right] L^5 \\
= \;& \left[\begin{array}{c} (80l - 72)\, l^2 W^2 - 36 + 4(23 - 10l)l \\ +(81 - 4l(90 + l(-75 + 16l)))W - 2(9 - 78l + 64l^2)W^2 \end{array}\right] L^5 \\
= \;& \left[\left(92l - 40l^2 - 36\right) + \left(300l^2 - 64l^3 - 360l + 81\right) W + (-128l^2 + 236l - 90)W^2\right] L^5
\end{aligned}$$

The last expression is a quadratic in W which is negative for all $W > 2$. In particular, it has a negative coefficient on the square, hence it is concave. Evaluated at $W = 2$ it is equal to $-128l^3 + 48l^2 + 316l - 234 < 0$ for all $l > 1$. Moreover, its derivative evaluated at $W = 2$ is equal to $-64l^3 - 212l^2 + 584l - 279 < 0$ for all $l > 1$.

To conclude the proof that the whole polynomial in L is negative, we still need to address the fact that the coefficient of L^6 is positive at $l = 2$.

In particular, we do so by proving that the sum of the terms in L^6 and L^4 is negative at $l = 2$. First, notice that the condition $\frac{lW}{L} \geq \frac{1}{2}$ implies that $L \leq 2lW$, which in turn implies:

$$\begin{aligned} & \left[-4l^2(8W-5)+l(-76+92W)+45-36W\right]/_{l=2}L^6 \\ < \;& 4\left[-4l^2(8W-5)+l(-76+92W)+45-36W\right]/_{l=2}L^4l^2W^2 \end{aligned}$$

which in turn implies that

$$\begin{aligned} = \;& \left[-4l^2(8W-5)+l(-76+92W)+45-36W\right]/_{l=2}L^6 \\ & +L^4\left[\begin{array}{c} 9-36l+20l^2+(-54+388l-528l^2+224l^3-32l^4)W \\ +(36-492l+780l^2-256l^3)W^2+(116l-192l^2)W^3 \end{array}\right]/_{l=2} \\ < \;& 4\left[-4l^2(8W-5)+l(-76+92W)+45-36W\right]/_{l=2}L^4l^2W^2 \\ & +L^4\left[\begin{array}{c} 9-36l+20l^2+(-54+388l-528l^2+224l^3-32l^4)W \\ +(36-492l+780l^2-256l^3)W^2+(116l-192l^2)W^3 \end{array}\right]/_{l=2} \\ = \;& \left(-216W^3-308W^2-110W+17\right)L^4 < 0 \text{ for all } W > 2. \end{aligned}$$

This concludes the proof that $b\left(L,W,l,-0.5\right) > 0$ for $l > 1$.

Calculation of $b\left(L,W,l,0.5\right) > 0$ for $l > 1$ and $\left[\frac{lW}{L}\right] = 0$.

We evaluate the function $b\left(L,W,l,\omega\right)$ at $\omega = 0.5$ and find the following expression:

$$\frac{L * \eta\left(L,W,l\right)}{\left(L+W-1\right)^2\left(-L+2LW+2lW+2Ll+2L^2\right)\left(-L+2lW+2Ll\right)}$$

where $\eta\left(L,W,l\right)$ can be expressed as a polynomial in L : in which all the coefficients, as well as the constant, are positive:

$$
\begin{aligned}
& \eta(L,W,l) \\
= \; & (12l-2)L^7 + \left[W(4l-4) + l(32W-28) + 32l^2W + 12l^2 + 3\right] L^6 \\
& + \left[\begin{array}{c} 64l^3W + l^2W(100W-44) + l^2(28W^2-24) \\ +W^2(6l-2) + lW(30W-72) + 20l + 7W \end{array}\right] L^5 \\
& + \left[\begin{array}{c} 6l^4W + l^3W(156W-96) + l^2W^2(192W-204) + 16l^2W(l^2-1) + 12lW^3 \\ +lW^2(100l^2-60) + (44lW-1) + l(12l-4) + 2W(2W-1) \end{array}\right] L^4 \\
& + \left[\begin{array}{c} l^4W(128W-16) + l^3W^2(300W-288) + 32l^3W + l^2W^3(128W-228) \\ +40l^2W^2 + 20l^2W + lW^3(84l^2-16) + lW(24W-8) \end{array}\right] L^3 \\
& + \left[\begin{array}{c} l^4W^2(192W-48) + l^2W^4(156l-80) + l^3W^3(100W-288) + 64l^3W^2 \\ +32l^2W^5 + 32l^2W^3 + 8l^2W^2 \end{array}\right] L^2 \\
& + \left[l^4W^3(128W-48) + l^3W^4(64W-96) + 32l^3W^3\right] L + 16l^4W^4(2W-1)
\end{aligned}
$$

c) Calculation of $\frac{\partial b(L,W,l,\omega)}{\partial \omega} < 0$ for all $\omega \geq -\frac{W}{L}$ for the case $l = 1$

For $l = 1$, the $b(L,W,l,\omega)$ function and its derivative with respect to ω are

$$
\begin{aligned}
b(L,W,1,\omega) = & \frac{LW(L+W)}{(L+W-1)^2} + \frac{L+W+L\omega}{W+L\omega} \\
& - \frac{(1+L)(W+LW+L\omega)(L+L^2+W+LW+L\omega)}{(L^2+W+LW+L\omega)^2}
\end{aligned}
$$

$$
\frac{\partial b(L,W,1,\omega)}{\partial \omega} = \frac{-L^3\phi(L,W,\omega)}{(W+L\omega)^2(L^2+W+LW+L\omega)^3}
$$

where $\phi(L,W,\omega)$ is the following cubic expression in ω in which all the coefficients, including the constant, are positive.

$$
\begin{aligned}
& \phi(L,W,\omega) \\
= \; & L^5 + 3L^3W + 3L^4W + 4LW^2 + 8L^2W^2 + 4L^3W^2 + 2W^3 + 4LW^3 + 2L^2W^3 \\
& + \omega\left(3L^4 + 8L^2W + 10L^3W + 2L^4W + 4LW^2 + 6L^2W^2 + 2L^3W^2\right) \\
& + \omega^2\left(4L^3 + 2L^4 + L^5 + 2L^2W + 3L^3W + L^4W\right) + \omega^3L^4
\end{aligned}
$$

The sign of the coefficients guarantees that the expression is positive, for all $\omega \geq 0$. To examine the sign of $\phi(L,W,\omega)$ for $w \in [-\frac{W}{L}, 0)$, notice that:

a) $\phi\left(L, W, -\frac{W}{L}\right) = L^2(L+W)^3 > 0$

b) $\phi(L, W, 0) = L^5 + 3L^3W + 3L^4W + 4LW^2 + 8L^2W^2 + 4L^3W^2 + 2W^3 + 4LW^3 + 2L^2W^3 > 0$

c)

$$\begin{aligned}
\frac{\partial \phi(L, W, \omega)}{\partial \omega} &= \left(3L^4 + 8L^2W + 10L^3W + 2L^4W + 4LW^2 + 6L^2W^2 + 2L^3W^2\right) \\
&\quad +2\omega\left(4L^3 + 2L^4 + L^5 + 2L^2W + 3L^3W + L^4W\right) + 3L^4\omega^2 \\
&\geq \left(3L^4 + 8L^2W + 10L^3W + 2L^4W + 4LW^2 + 6L^2W^2 + 2L^3W^2\right) \\
&\quad +2\omega\left(4L^3 + 2L^4 + L^5 + 2L^2W + 3L^3W + L^4W\right) \\
&> \left(3L^4 + 8L^2W + 10L^3W + 2L^4W + 4LW^2 + 6L^2W^2 + 2L^3W^2\right) \\
&\quad -2\frac{W}{L}\left(4L^3 + 2L^4 + L^5 + 2L^2W + 3L^3W + L^4W\right) \\
&= 3L^3(L + 2W) > 0
\end{aligned}$$

where the first inequality follows from the fact that $3L^4\omega^2 \geq 0$ and the second from the fact that $\omega > -\frac{W}{L}$.

Hence we can conclude that $\phi(L, W, \omega)$ is positive and increasing in the whole interval $\left(-\frac{W}{L}, 0\right)$, hence the function $b(L, W, 1, \omega)$ is decreasing for all $\omega > -\frac{W}{L}$.