I. Metainduction - Basic Account: New Solution to the Problem of Induction?

(Mo 10.8. 10.00-11.00) Gerhard Schurz (DCLPS, HHU Düsseldorf)

1. Introduction: The Problem of Induction

Hume's problem: How can we rationally justify the inductive transfer of patters or regularities from past observations to the unobserved future?

Hume's insight: we cannot demonstrate the success (reliability) of induction (I), because all conceivable strategies of justification seem to fail:

– I cannot be justified by logic, because it is logically possible that future $\neq$ past.

– I cannot be justified by observation, because I's conclusions are about the unobserved.

– the only remaining possibility would be to justify I by induction from its past success, but this would either amount to an *infinite regress* (higher-order inductions) or to a *circle*.

● Contrary to claims of several epistemologists (Black 1974, van Cleve 1984, Papineau 1993, ch. 5; Goldman 1999, 85; Lipton 1991, 167ff.; Harman 1986, 33; Psillos 1999, 82):

(Rule-) Circular justifications are epistemically worthless, because with their help one may 'justify' opposite conclusions (Salmon 1957):

Inductive Just. of I:	Anti-Inductive Just. of Anti-I :
Past inductions were successful	Past anti-inductions were not successful
[Therefore by the rule of induction:]	[Therefore by the rule of anti-induction:]
Future inductions will be successful	Future anti-inductions will be successful

Similar refutation strategy are possible in other cases:

Rule-circular 'justification' of inference to the best explanation (IBE): The assumption that IBEs are reliable is the best (available) explanation of the fact that so far, most hypotheses introduced by IBEs have been successful. Therefore, by the IBE rule: IBEs are reliable.

(Douven 2011): rule-circular justification of 'inference to the worst explanation'.

Rule-circular 'justification' of the inference to authority, IA ("If the authority A tells that p, infer that p is true"): A tells that the rule IA is reliable. Therefore rule IA is reliable.

Refutation by inference to the opposite authority.

● If we attempt to justify scientific theories, or real experts, by their explanatory and predictive success, we basically need a justification of induction …

Is a (non-circular) justification of induction impossible (as many epistemologists think)?

The practical significance of this question: if we cannot justify induction, what reason do we have to prefer science over religion …?

2. Hume's Problem Within Bayesianism

In Bayesianism Hume‘s problem is not immediately apparent. But it is there:

● If one assumes a *state-uniform distribution* – a uniform prior distribution over possible worlds (say, binary event sequences) –, then induction becomes impossible:

$P(Fa_{n+1} \mid freq_n(F) = k/n) = 1/2$ for all $k \leq n \in \mathbb{N}$ (Carnap 1956; c†).

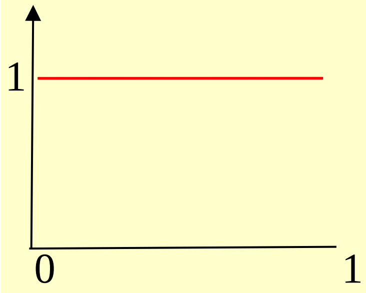
Wolpert's *no-free-lunch theorem* (1996) is a generalization of this result (Schurz 2017).

● On the other hand: if one assumes a *frequency-uniform distribution* – a uniform prior distribution over possible frequencies of binary events – then one obtains Laplacean induction rule: $P(Fa_{n+1} \mid freq_n(F) = k/n) = (k+1)/(n+2)$ for all $k \leq n \in \mathbb{N}$.

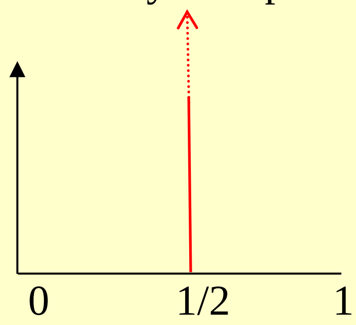
Which prior is the 'right' one? **Moral: all priors are biased in some respect**.

Transformation of prior distributions:

Uniform P-density over
possible sequences (binary coding)

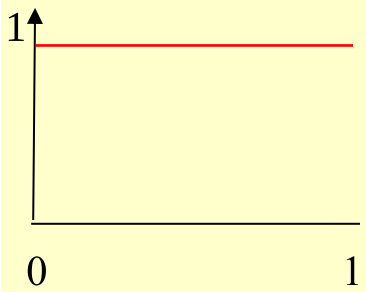


Corresponding "maximally dogmatic"
P-density over possible frequencies



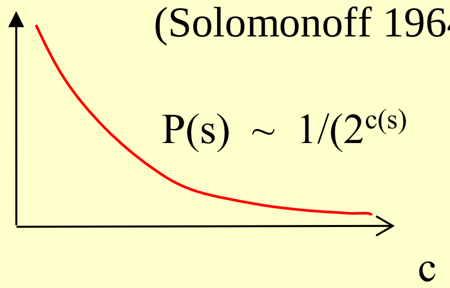
Outwashing of this prior
is impossible!

Uniform P-density over
possible frequencies



Corresponding "inductive" P-density
over algorithmic complexity of sequences

(Solomonoff 1964)



A justification of induction is needed that is independent from an assumed prior.

Is this possible?

3. Optimality Justifications – an Escape?

Schurz (2008, …, 2019): New approach to Hume's problem based on meta-induction.

Distinction: Object-induction (level of events) vs. *meta-induction* (level of methods).

The approach is compatible with Hume's diagnosis that one cannot demonstrate the *reliability* of induction.

It attempts to show something weaker: the *optimality* of induction

– in all possible worlds (including paranormal worlds hosting clairvoyants, anti-inductivistic demons; since otherwise account would be circular)

– among all methods that are *accessible* to the epistemic agent ('access-optimal').

Two crucial features:

● Shift to **optimality**: in induction-hostile worlds, induction may be "best of a bad lot".

● Shift to **meta**-induction (MI) and optimality among **accessible** methods.

General characterization of "meta-induction":

A meta-inductive method favors prediction methods according to their observed success and attempts to predict an optimal combination of their predictions.

Imitate the best, ITB: the simplest meta-inductive method.

Weighted MI methods: weigh predictions of methods according to observed success.

Optimality account is related to *Hans Reichenbach*'s "best alternative" account (1949).

● Problem of Reichenbach's account: focused on object-induction. Result in formal learning theory show: *impossible* to establish optimality w.r.t. all object-level methods.

Given method M → construct M-demonic world w → constr. w-perfect method M* → M* better than M in w (Putnam 1965, Kelly 1996; Skyrms 1975 against Reichenbach).

● But optimality may be possible for MI methods w.r.t. all accessible methods.

Here the last →step is no longer valid, because MI would imitate M*.

Is the restriction to *accessible* methods a drawback?

No, since inaccessible methods are epistemically irrelevant.

On the relation between meta- and object-induction:

● If the universal access-optimality of a particular MI-method could be demonstrated, this would provide an *a priori* justification only of meta-induction (not of object-induction).

● However: the a priori justification of meta-induction implies the following *a posteriori* justification of object-induction:

So far object-inductive methods were (much) more successful* than non-inductive methods of prediction; therefore it is meta-inductively justified to favor object-induction in the future.

This argument is not circular, because of the independent justification of meta-induction.

*Precisely: Until now, ind. methods were often significantly more successful than non-ind. methods, but not vice versa (compatible with fact that sometimes no method is successful).

4. Prediction Games (for the following see Schurz 2019)

A (real-valued) *prediction game* consists of:

(1) An infinite sequence (e) = (e_1, e_2,…) of real-valued events $e_i \in VAL \subseteq [0,1]$ (normalized)

(2) A (finite) set of 'players' Π whose task is to predict next (future) events.

$pred_n(P) \in [0,1]$: prediction of P *for* time (round) n, delivered *at* time n–1.

Important: Players may predict mixtures of events. – Even if events are binary (VAL = {0,1}), predictions may be real-valued. *Application: Probabilistic* predictions.

Players in Π include (2.1): one or several meta-inductivists 'xMI' (x = type of MI),

(2.2) a (finite) set of other players $P_1,\ldots,P_m$ (the non-MI-players): either object-inductivists, or alternative players (e.g., clairvoyants who may have perfect success).

We identify players with prediction methods.

Success evaluation: Normalized loss function $loss(pred_n,e_n) \in [0,1]$.

Natural loss $|e_n - pred_n|$. Our theorems admit many other functions, e.g. *convex* ones.

score $s(pred_n,e_n) := 1 - loss(pred_n,e_n)$

absolute success: $Suc_n(P)$:= P's sum of scores until time n

relative success (success rate) $suc_n(P) := Suc_n(P) / n$.

absolute attractivity of P for xMI (*regret* of xMI wr.t. P): $At_n(P) := Suc_n(P) - Suc_n(xMI)$

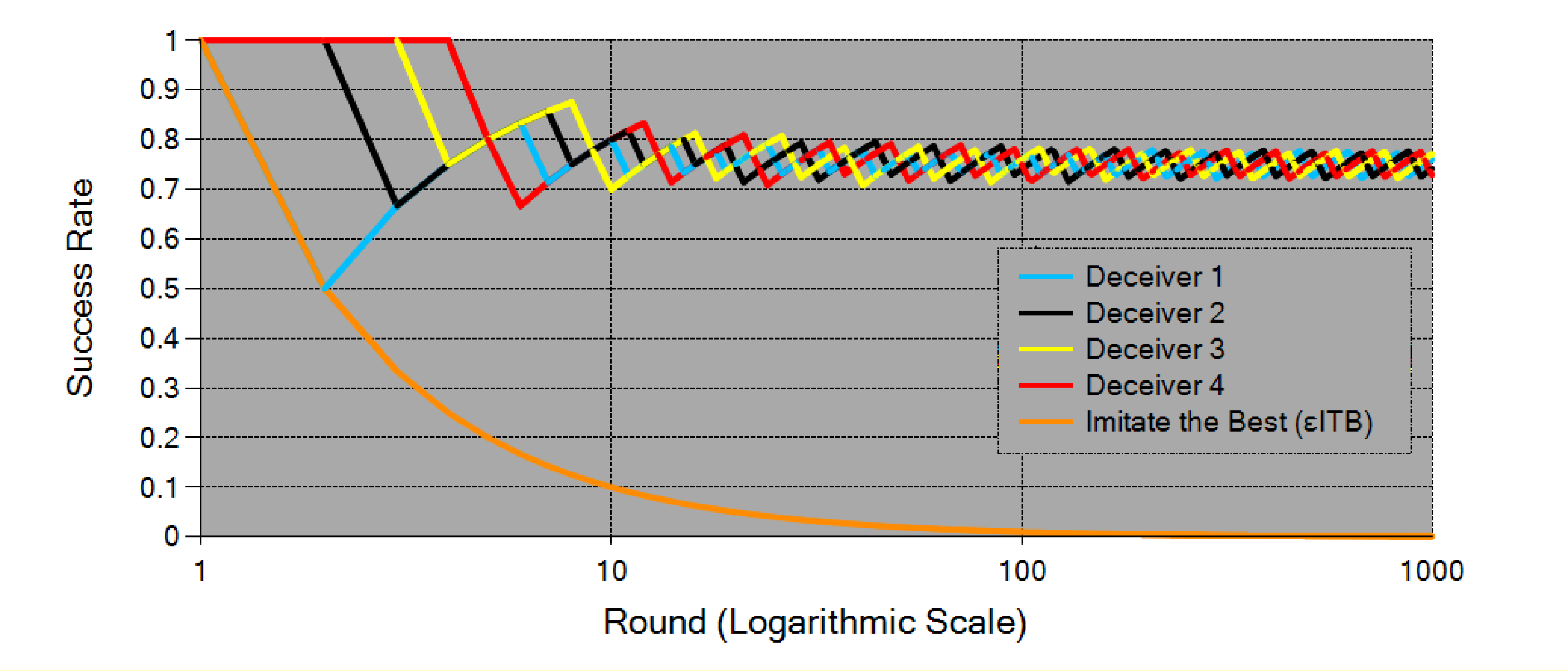
relative attractivity (attr. rate): $at_n(P) := At_n(P) / n$

Theorem 1 – major result about ITB: ITB is only access-optimal in environments with success rates converging to a stable ordering; they must not oscillate forever.

ITB may be deceived by players whose success goes down as soon as they are favored by ITB → this leads to success-oscillations of players modulo the switching threshold ε of ITB.

Example: stock market in a bubble economy. – *Programming* (by Paul Thorn):

if ITB favors a deceiving player P, P predicts incorrectly, else correctly.



- The **delay problem**: observation of change of leader costs time (one score unit).

Theorem 2: No *one-favorite* MI method can be universally access-optimal.

Conclusion: Optimality can only be found in the class of success-weighted MIs.

But not all success-dependent weightings will do.

5. Attractivity-Weighted Meta-Induction

Predictions of weighted meta-induction wMI:

For all times n>1 with $\sum_{1\le i\le m} w_n(P_i) > 0$: $pred_{n+1}(wMI) = \dfrac{\sum_{1\le i\le m} w_n(P_i)\cdot pred_{n+1}(P_i)}{\sum_{1\le i\le m} w_n(P_i)}$.

(If n=0 or $\sum_{1\le i\le m} w_n(P_i) = 0$, wMI predicts by its 'fallback-method'.)

Attractivity-weighting: Simple a.w. meta-inductivist AW: $w_n(P) = \max(at_n(P),0)$.

Exponential a.w. meta-inductivist EAW: $w_n(P) := e^{\eta\cdot n\cdot at_n(P)}$ where $\eta = \sqrt{8\cdot \ln(m)/(n+1)}$.

Crucial: a.w. MI *forgets* players whose regret is negative.

Note: AW forgets immediately; EAW forgets gradually.

There are further variants of AW: e.g. polynomial AW (…).

Universal Optimality Results (long-run; based on Cesa-Bianchi and Lugosi 2006, Schurz 2008, 2019; cf. Shalev-Shwartz and Ben-David 2014, "online learning under expert advice"):

Theorem 3: *Universal long-run access-optimality of (E)AW with tight upper-bounds for short-run losses*

For every prediction game ((e), $\{P_1,\ldots,P_m, xAW\}$) whose loss-function is *convex* in the argument $pred_n$, the following holds for all $n \geq 1$:

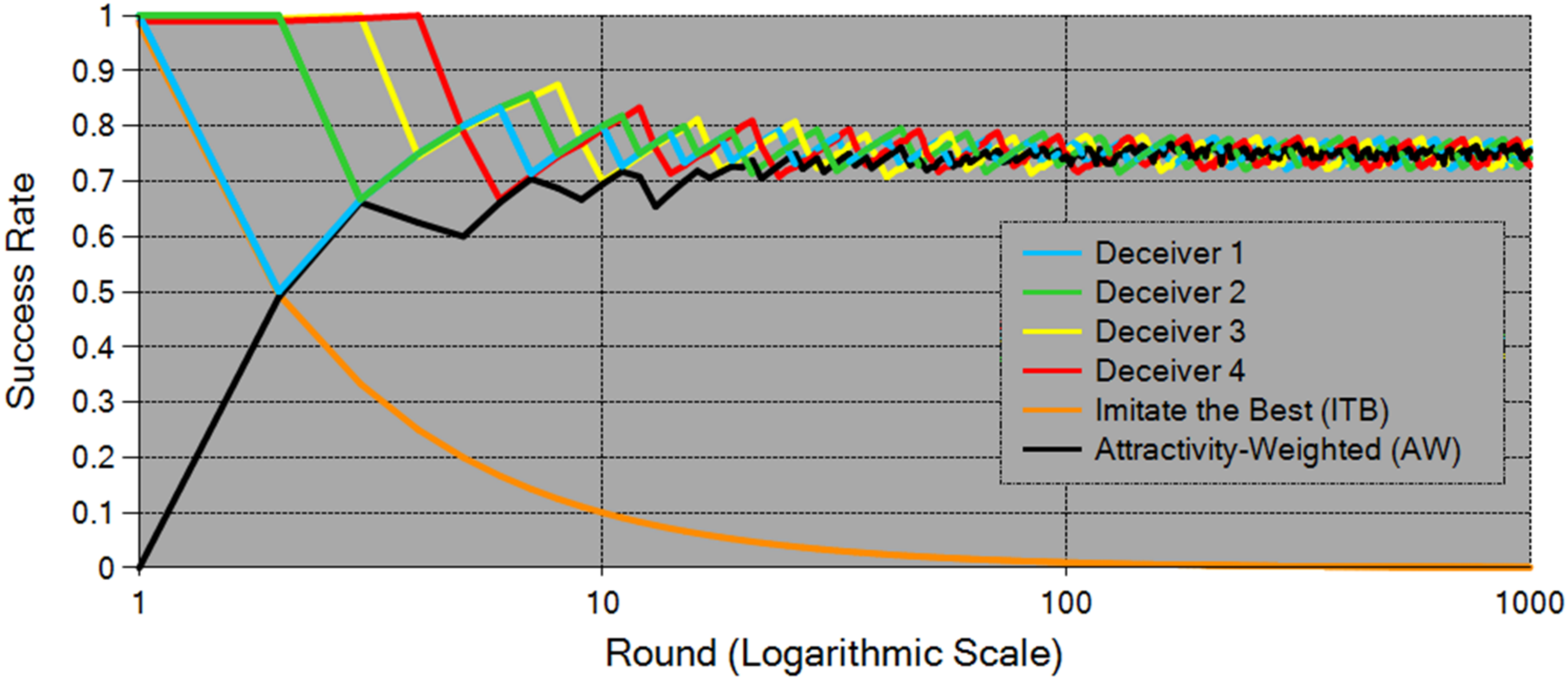
(1) For AW – short-run: $maxsuc_n - suc_n(AW) \leq \sqrt{\frac{m}{n}}$.

(2) For EAW – short-run: $maxsuc_n - suc_n(EAW) \leq 1.78 \cdot \sqrt{2 \cdot \ln(m)/n}$.

(3) Thus for AW and EAW – long-run: $\limsup_{n\to\infty} (maxsuc_n - suc_n(EAW)) \leq 0$.

Two crucial features: (1.) (E)AW cannot be deceived by adversarial players, because if they oscillate in their success-rates, (E)AW predicts the average of their predictions.

Programming:



(2.) Difference between attractivity-weighting and **success-weighting** ('Franklin's rule', cf. Gigerenzer et al. 1999, part III; Jekel et al. 2012, etc.)

Success-weighted MI (SW) does not forget players that are less successful than the MI. Thus, its success cannot converge to the maximal success. SW cannot be access-optimal.

On the relation between (E)AW and ITB:

In scenarios in which ITB is optimal (stable success ordering), (E)AW coverge to ITB in their behavior, with a small delay.

On the relation between AW and EAW (recent simulations with Paul Thorn):

Over all possible sequences: EAW is better in avoiding large regrets than AW, while AW forgets faster and is better in avoiding regrets for regular sequences in which object-induction achieves high success.

II. Metainduction - Extensions of the account (Tue 11.8. 10:00 - 11:00)

6. Discrete Prediction Games

Mixtures of predictions are impossible or not allowed. *Theorem 3 fails.*

$pred_n \in$ discrete event value space VAL = $\{v_1,\ldots, v_q\}$ **Binary games:** VAL = $\{0,1\}$

Theorem 4: No individual (MI) method can be universally optimal in discrete games.

Proof: Take a binary game, an arbitary (MI) method M, an M-demonic event sequence (e), and the two methods 'Always-1' and 'Always-0'.

Then at any time n, M's success rate is 0, while at least one of Always-1 and Always-0 has a succes rate $\geq$ 0.5.

Two methods of transferring theorem 3 to discrete games:

(1.) Randomized a.w.MI – R(E)AW (Cesa-Bianchi and Lugosi 2006):

Each time RAW predicts an event value $v_i \in VAL$ with a probability equal to the normalized weight-sum of all non-MI players predicting v_i (with weights assigned as by AW).

Theorem 5: For arbitrary loss functions: If RAW's choice of prediction is probabilistically independent from predicted event, then:

$maxsuc_n - \overline{suc_n}(RAW) \leq$ the regret bound of AW, where $\overline{suc_n}$ is the (cumulative) *expected success* rate. (Similarly for REAW.)

Definition: $\overline{suc_n}$ (RAW) $=_{def} (1/n)\cdot \Sigma_{1\leq i\leq n}Exp(score_i(RAW))$, where

$Exp(score_i(RAW)) =_{def} \Sigma_{1\leq r\leq q} P(pred_i(RAW)=v_r)\cdot score(v_r,e_i))$. (Likewise for REAW)

● Advantage: the result holds for **arbitrary loss functions** (because the *expected* loss of probabilistic predictions is always linear).

● Strong disadvantage: the optimality of randomized MI excludes deceptive scenarious.

(2.) Collective a.w.MI – *CAW* (Schurz 2008): $AW_1,\ldots, AW_k$.

Each time, a fraction k_i/k of the k meta-inductivists predict the event value $v_i \in Val$ that approximates as close as possible RAW's probability of v_i.

Theorem 6: For arbitrary loss functions:

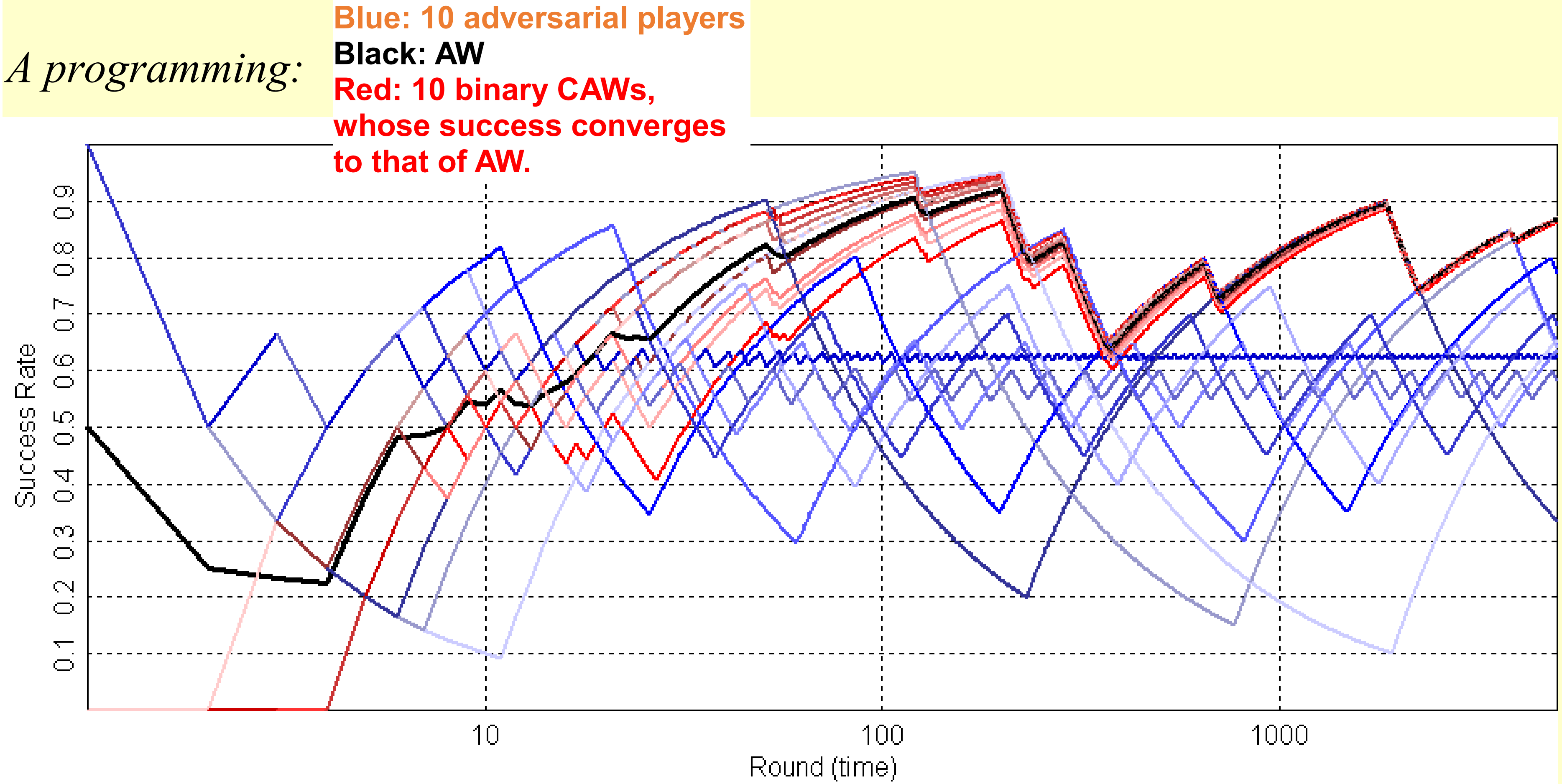
$maxsuc_n - \overline{suc_n}(X) \le$ (E)AW's regret bound $+\frac{q-1}{2k}$, where $\overline{suc_n}$ is the *average success*.

● Disadvantage: The additional loss term of $\frac{q-1}{2k}$. (Can be made small by large k).

● Strong advantage: The approximative optimality of collective MI is **universal**.

● Assuming the CAW's share their success, collective optimality guarantees optimality for every indidvual. Here a *practical* condition becomes directly epistemologically relevant: by epistemic cooperation, the negative result of theorem 4 can be defeated.

A programming:



7. Unboundedly Growing Sets of Methods

Challenge of Arnold (2010) and Sterkenburg (2018, 2019): Theorems are restricted to fixed finite sets of accessible methods.

Defense: Humans' cognitive resources are finitely bounded.

Successor problem *(Sterkenburg):* The set of 'candidate methods' cannot be fixed. We need meta-induction over unboundedly growing sets of methods:

$\Pi(n) = \{P_1,\ldots,P_{m(n)}\}$, where m(n) is monotonically growing.

● The meta-inductivist attributes to all new players a hypothetical *default success* for past times of the game when they were absent.

Otherwise a *fair* comparison is impossible: it may be that before the *entrance time* of a player P it was *much harder* to attain predictive success than *after* t.

Which 'default success' should be attributed?

Solution: EAW attributes to a new player P the so-far success of him-/herself (Chernov and Vovk 2009).

Epistemic advantage: fair.

Technical advantage: makes transfer of theorems 3, 5,6 possible.

Theorem 7: *Access-optimality of EAW_{gr} for growing player sets*:

Then for every prediction game $((e), \{P_1,\ldots,P_{m(n)}, EAW_{gr}\})$:

(1) $maxsuc_n - suc_n(EAW_{gr}) \leq 1.78 \cdot \sqrt{2 \cdot \ln(m(n))/n}$ (the regret bound of EAW).

(2) If m(n) grows *slower than exponential* with n ($\lim_{n\to\infty} m(n)/e^n = 0$):

$\lim_{n\to\infty} (maxsuc_n - suc_n(EAW_{gr})) \leq 0$

(Similarly for REAW, CEAW.)

8. A Result for Goodman-type methods

Assumption: a given language with qualitative primitive predicates (Goodman 1955).

(Goodman's problem of language-relativity has to be solved independently)

A Goodman-method with k switch points is an arbitrary piecemeal combination of k+1 qualitatively defined *basic methods*:

$P_1,\ldots,P_1,\quad P_2,\ldots,P_2\ \ldots\ P_{k+1},\ldots,P_{k+1}.$

kth switch point

1st switch point 2nd switch point

Problem: We shouldn't include in the candidate set too many 'crazy' Goodman-methods.

Theorem 8: There is variant of EAW (the 'fixed share' EAW) that tracks the success rates of the basic methods $P_1,\ldots,P_m$, but is nevertheless access-optimal in regard to all Goodman-type combinations of basic methods whose *switch number* k(n) *grows sublinearly with n*.

9. Further Generalizations and Applications

9.1 Generalization to action games ("multi-armed bandits")

9.2 Results about Dominance (long-run)

There are several equally optimal MI methods (with different short-run properties).

(1) (E)AW dominates every independent method and every meta-method that is not access-optimal.

(2) Not access-optimal meta-methods are: all one-favorite methods, success-weighted MI, linear regression with linear loss function, simply non-inductive meta-methods, …

● **Reconciliation with the no free lunch theorem**: state-uniform probability of infinite event sequences in which MI dominantes these methods is zero (Schurz 2017).

9.3 Application to Bayesian epistemology: probabilistic prediction games - tomorrow.

9.4 Outlook: Applications to Social Epistemology and Cultural Evolution

- **Meta-induction =** success-based **social learning.**

Schurz (2012): *Local Meta-Induction in epistemic neighborhood structures:*

Here, success-information and meta-inductive learning is restricted to local neighborhood structues.

Provided the neighborhoods are overlapping, expert knowledge spreads.

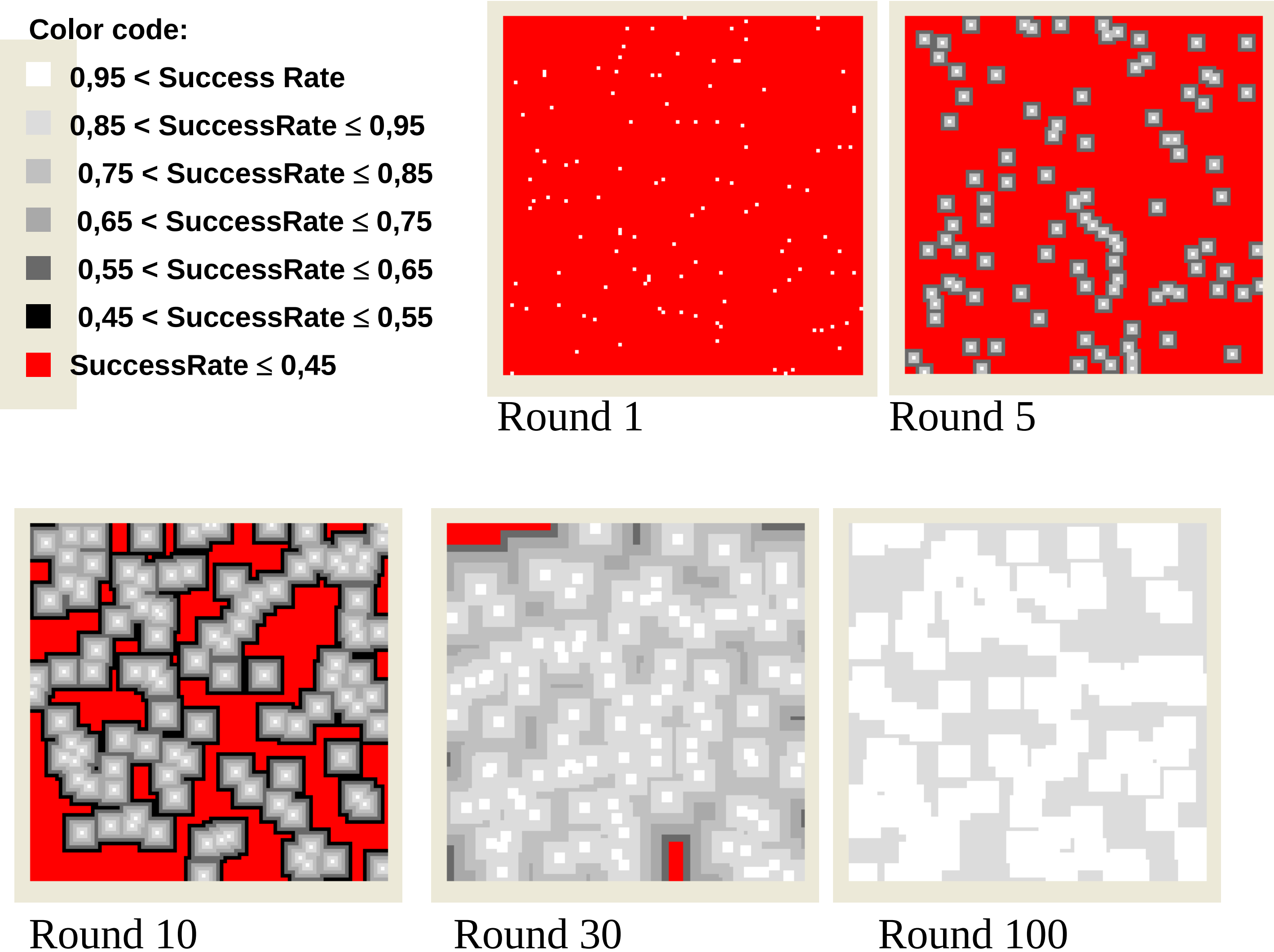


Figure 5: Local Meta-induction spreads reliable knowledge of 1% experts (white spots in round 1) among 99% unreliable nonexperts (red area in round 1) within 100 rounds with 12 cycles per round. In round 200 everything has become white.

Rendell et al. (2010) – computer tournament: Social learners were much more successful than individual learners in the *all-against-all* tournament. But when social learners played against themselves, their success-rate went down (Roger's Paradox).

Conclusion:

(1.) Members of a successful research community should not *only* apply MI, but at the same time attempt to improve their *independent* methods (theories).

(2.) Populations can only survive if they do not only consist of imitators/social learners; a possibly small fraction of independent learners is needed; otherwise extinction.

Douven (forthcoming in BJPS): Optimality account has to be complemented by an *explanation* why induction is not only optimal, but *highly* successful.

He offers an explanation based on *evolutionary programming* of prediction games. Meta-induction is indirectly implemented by evolutionary selection of successful predictors.

III. Bayesian Prediction Games and Meta-inductive Probability Aggregation

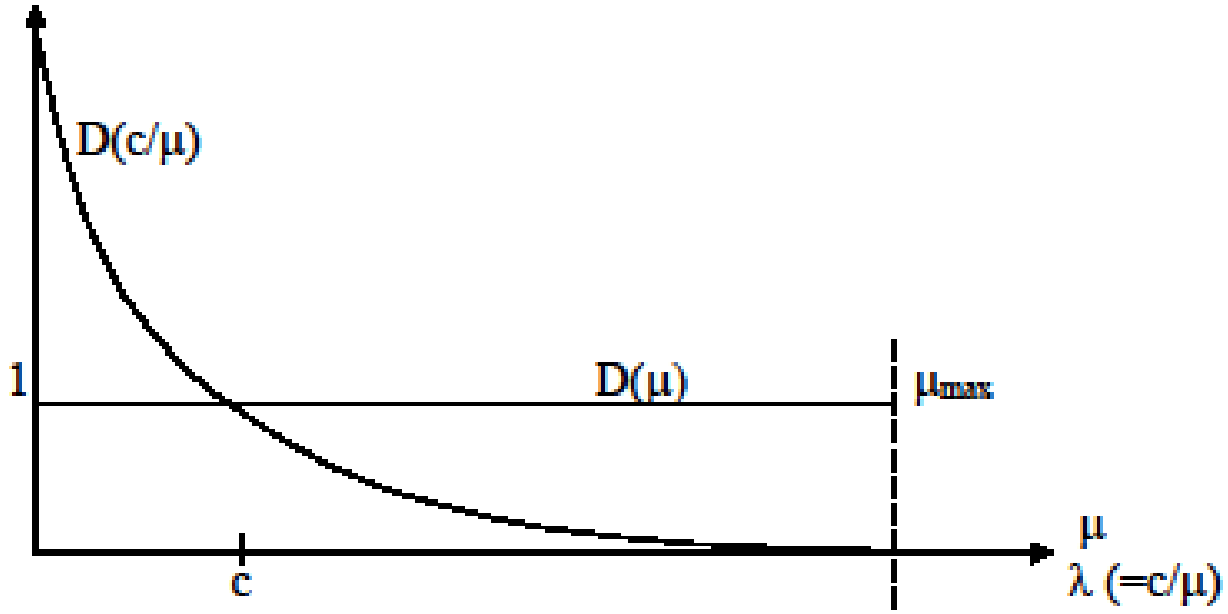
(Wed 12.8.10:00-11:00)

The problem of choosing a prior distribution:

For objective Bayesianism: Equiprobability is language-dependent.

- **Recall section 2:** If one assumes a *state-uniform distribution* – a uniform prior distribution over possible worlds (say, binary event sequences) –, then induction becomes impossible.
- If one assumes a *frequency-uniform distribution* – a uniform prior distribution over possible frequencies of binary events – then one obtains the Laplacean induction rule.

Moreover: uniform distributions are not preserved under *fineness-preserving* language transformations (cf. Gillies 2000, 37-48). Example:



Uniform density for μ (frequency) turns into a non-uniform density for λ (wave-length).

For subjective Bayesianism: Bayesian convergence theorems hold only for prior distributions that are non-dogmatic and (in the infinite case) continuous.

Thus: *not all prior distributions can be outwashed by conditionalizing on increasing amounts of evidence*. For example, the state-uniform distribution cannot be outwashed.

Moral: An a priori justification of particular prior distributions is impossible (Hume's insight). – All a priori choices contain a subjective element.

Proposal: Use meta-induction to choose the optimal distribution function a posteriori.
This determines the optimal 'prior' distribution post-facto.

Recapitulation (from Monday):

General characterization of "meta-induction":

A meta-inductive method favors prediction methods according to their observed success rates and attempts to predict an optimal combination of their predictions.

Imitate the best, ITB: is the simplest meta-inductive method, but not universlly optimal.
Weighted MI methods: weigh predictions of methods according to observed success.

A (real-valued) ***prediction game*** consists of:

(1) An infinite sequence (e) := (e_1, e_2,…) of real-valued events $e_n \in VAL \subseteq [0,1]$.

(2) A (finite) set of accessible methods ('players') Π whose task is to predict next (future) events. $pred_n(P) \in [0,1]$: prediction of P *for* time (round) n, delivered *at* time n–1.

Important: Players may predict mixtures of events. – Even if events are binary (VAL = {0,1}), predictions may be real-valued. *Application: Probabilistic* predictions.

Players in Π include (2.1): one or several meta-inductivists 'xMI' (x = type of MI),

(2.2) a (finite) set of other players $P_1,\ldots,P_m$ (the non-MI-players): either object-inductivists, or alternative players (e.g., clairvoyants who may have perfect success).

Success evaluation: Normalized loss function $loss(pred_n,e_n) \in [0,1]$.

Natural loss $|e_n - pred_n|$. Theorems admit many other loss functions, e.g. *convex* ones.

score $s(pred_n,e_n) := 1 - loss(pred_n,e_n)$

absolute success: $Suc_n(P)$:= P's sum of scores until time n

relative success (success rate) $suc_n(P) := Suc_n(P) / n$.

absolute attractivity of P for xMI (*regret* of xMI wr.t. P): $At_n(P) := Suc_n(P) - Suc_n(xMI)$

relative attractivity (attr. rate): $at_n(P) := At_n(P) / n$

Predictions of weighted meta-induction wMI:

For all times n>1 with $\sum_{1\le i\le m} w_n(P_i) > 0$: $pred_{n+1}(wMI) = \dfrac{\sum_{1\le i\le m} w_n(P_i)\cdot pred_{n+1}(P_i)}{\sum_{1\le i\le m} w_n(P_i)}$.

Attractivity-weighting: Simple a.w. meta-inductivist AW: $w_n(P) = \max(at_n(P),0)$.

Exponential a.w. meta-inductivist EAW: $w_n(P) := e^{\eta\cdot n\cdot at_n(P)}$ where $\eta = \sqrt{8\cdot\ln(m)/(n+1)}$.

Bayesian prediction games:

Prediction games with binary or discrete event values $Val = \{v_1,\ldots,v_q\}$

Predictions are *probability distributions over Val* ('Bayesian predictors).

● ***Question:* When is it reasonable to predict the probability of event values for the purpose of maximizing predictive success?** → Depends on the chosen scoring function.

If the deviation of predicted probability of the actual event from its truth-value (1) is scored by the absolute (linear) distance, then it is **not** optimal to predict probabilities, but to predict truth values: '1' for event value with maximal probability and '0' otherwise ('maximum rule).

Proof (binary case, p = IID event probability; pred = prediction):

$p \cdot pred + (1-p) \cdot (1-pred)$ is maximal if pred = 1 if $p \geq 0.5$ and pred=0 otherwise.

Note: From this one should not infer that linear scoring rules are less adequate (cf. Maher 1990; Fallis 2007). – In my view, the result shows that under linear scoring *the thesis that subjective probabilities are rational estimations of truth values* is false.

(Rather, they are rational estimations of objective probabilities.)

Moral: The pro's and con's of certain scoring functions are context-dependent.

● **A *proper* scoring rule:** A scoring function that maximizes the P-expected score if one predicts P.

Brier (1950): the quadratic loss function, loss(e,pred) = $(e-pred)^2$, constitutes a proper scoring rule. (cf. Selten 1998).

Proof: By differentiating expected quadratic loss w.r.t. pred and setting it zero:

$d[p\cdot(1-pred)^2 + (1-p)\cdot pred^2]/dpred = d[p-2p\cdot pred + pred^2]/dpred = -2p+2pred =! 0$;

hence pred = p.

There are other proper scoring functions, e.g. logarithmic ones (Fallis 2007).

Objective interpretation: Under proper scoring, a rational forecaster will attempt to predict degrees of belief that *match the objective probabilities*, because only then expected success coincides with average success.

A **Bayesian prediction game** is a real-valued prediction game $((e),\{P_1,\ldots,P_m, xMI\}$ with discrete event values $Val = \{v_1,\ldots,v_q\}$ and for all P_i $(1\leq i\leq m)$ and $n \in N$:

(i) P_i's prediction equals P_i's probability distribution over VAL conditional on past evidence:

$pred_{n+1}(P_i) = (r_1,\ldots,r_q)$, where: $r_j = P_{i,n}(e_{n+1}=v_j \,|\, e_1,\ldots,e_n)$,

"$e_1,\ldots,e_n$": the sequence of the past event values, "$e_{n+1}=v_j$": the prediction that the next event value will be v_j, and $P_{i,n}$ = the probability function of player P_i *at* time n.

(ii) If $e_{n+1} = v_k$, then $score(pred_{n+1}(P_i),e_{n+1}) = 1-loss(r_k,1)$, where the loss function is *proper:*

For all P: $Val\rightarrow[0,1]$ and predictions $(s_1,\ldots,s_q) \in [0,1]^q$ (with $\Sigma_{1\leq i\leq q}\, s_i = 1$)

$Exp_p(loss(s_1,\ldots,s_q)) =_{def} \Sigma_{1\leq i\leq q}P(e=v_i)\cdot loss(s_i,1)$ is *minimal* iff $s_i = r_i$ for all $i\in\{1,\ldots,q\}$.

Note: This scoring method is adopted in Cesa-Bianchi and Lugosi (2006, ch. 9), but confined to logarithmic loss function. Brier's (1950) uses a more refined scoring method that adds the loss between the predicted probability and truth value for all event values.

Universal Optimality Result for (E)AW (based on Cesa-Bianchi and Lugosi 2006, Schurz 2008, 2019, Shalev-Shwart and Ben-David 2014) – applied to Bayesian prediction games:

Theorem 9: *Optimality of AW-based probability aggregation:*

For every Bayesian prediction game $((e), \{P_1,\ldots,P_m, xAW\})$:

(1) For AW – short-run: $\text{maxsuc}_n - \text{suc}_n(AW) \leq \sqrt{\frac{m}{n}}$.

(2) For EAW – short-run: $\text{maxsuc}_n - \text{suc}_n(EAW) \leq 1.78\cdot\sqrt{2\cdot\ln(m)/n}$.

(3) For AW and EAW – long-run: $\text{limsup}_{n\to\infty} (\text{maxsuc}_n - \text{suc}_n(AW)) \leq 0$.

$P_{AW,n}$ is an *aggregated* conditional probability function, whose weights are meta-inductively determined based on objective sucess rates (Feldbacher-Escamilla and Schurz 2020) (→ this may solve a problem of probability aggregations; cf. Mongin 2001).

From the aggregated conditional distribution P_{AW} the *optimal prior distribution* over the events can be calculated from the predictive probabilities **post-facto** as follows, where $(v_{i_1},\ldots,v_{i_n})$ is a sequence of n event values at times 1,…,n:

$P_{AW}(v_{i_1},\ldots,v_{i_n}) = \quad \Pi_{1\le t\le n}\, P_{AW}(v_{i_t} | v_{i_1},\ldots,v_{i_t}) \quad (= P_{AW}(v_{i_1}) \cdot P_{AW}(v_{i_2} \mid v_{i_1}) \cdot \ldots)$.

Note: this prior is 'post facto' because the weights of the aggregated P-function depends on the success of the probabilistic predictors and thus on the actual events to be predicted.

Final remark: With the logarithmic loss function Bayesian predictors attain an especially simple mathematical format:

Logarithmic loss function: $\text{loss}(P_{i,n},e_{n+1}) = -\ln(P_{i,n}(e_{n+1}))$. *In words:* the loss of P_i's prediction of e_{n+1} is the negative logarithm of P_i's probability of the actual value of e_{n+1}.

- Disadvantage of logarithmic loss: for $P(e)\rightarrow 0$, $loss(P,e)\rightarrow\infty$, which is rather unnatural.
- Advantage of logarithmic loss: improved regret bound of EAW: ln(m)/n.

With the logarithmic loss function, EAW's weight rule can be transformed into a rule that bears a similarity with a Bayesian updating. One obtains:

$$w_n^{EAW}(P) = e^{-Loss_n(P)} = e^{-\Sigma_{1\le t\le n} -\ln P(e_t|e^{t-1})} = \Pi_{1\le t\le n} P(e_t|e^{t-1}) = P(e^n)$$

(cf. Cesa-Bianchi and Lugosi 2006, 249; Sterkenburg 2018).

Also in this case, the determination of weights and priors is *post facto*, since this equation holds only for the *actual* course of events e^n, which determines the weights (not for all possible courses of events).

Application to Data: Empirical Prediction Games

(Schurz and Thorn 2016, Thorn and Schurz 2019)

Monash University Footy Tipping Competition:

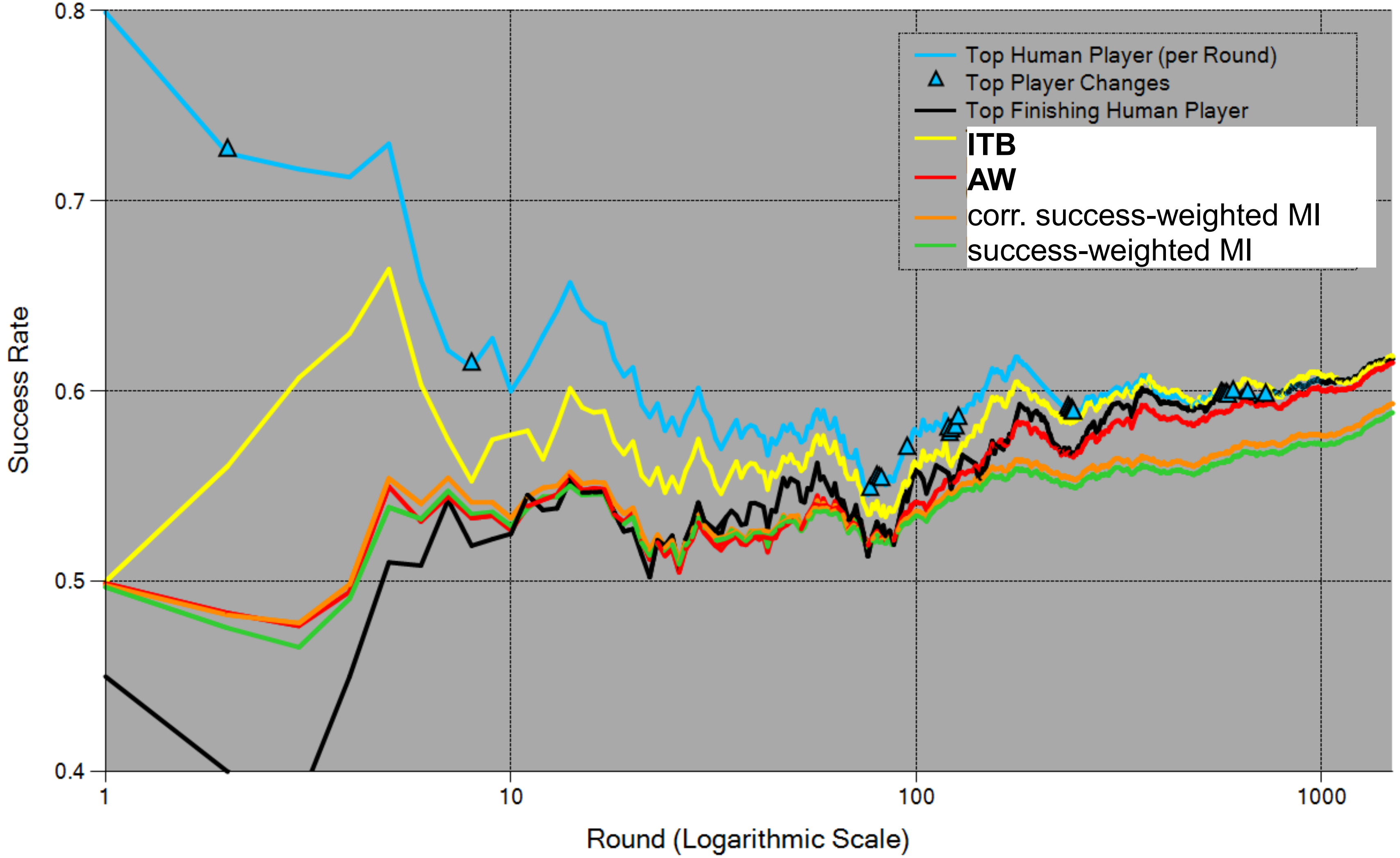
Event-sequence: 1514 matches of the Australian Football League 2005-2012.

1071 human predictors (a "short run" experiment) predicting the winning probability.

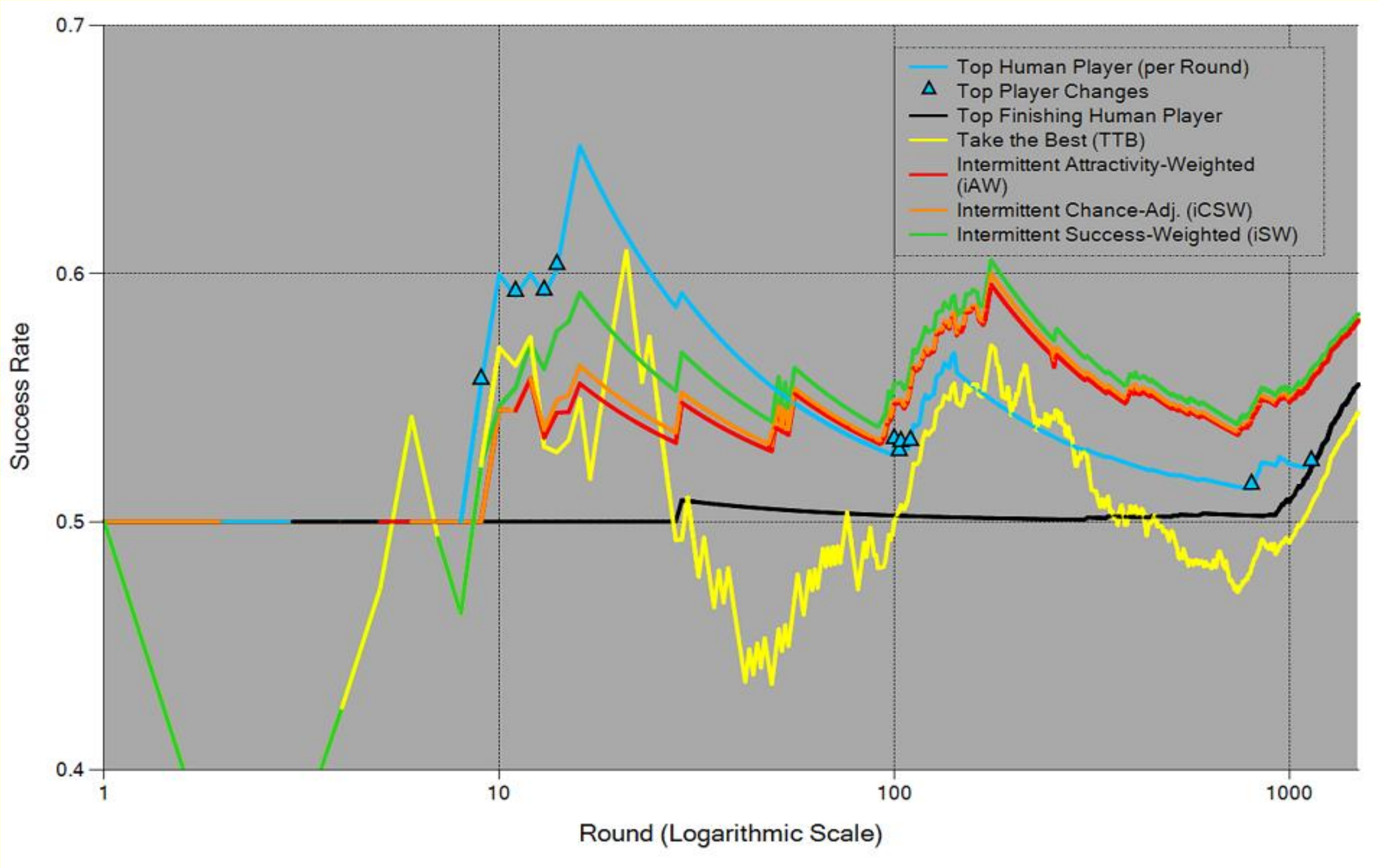
Results: In 6 out of the 8 seasons, there was a different best player; but EAW and AW were always at the top (almost no difference between AW and EAW).

Round	*Worst case regret of EAW*	*Obtained regret of (E)AW*
20	0.29	0.025
100	0.13	0.026
500	0.06	0.006
1500	0.034	0.005

Results for 69 players predicting 50% of time (permanent evaluation)



Results for 50 players with best 'ecological validity', intermittent evaluation



References

Arnold, E. (2010): "Can the Best-Alternative-Justification Solve Hume's Problem? On the Limits of a Promising Approach", *Philosophy of Science* 77, 584-593.

Black, Max (1974). Self-supporting Inductive Arguments. In: Swinburne. R. (ed.), *The Justification of Induction* (pp. 127-134). Oxford: Oxford University Press.

Brier, G. (1950): "Verification of Forecasts Expressed in Terms of Probability", *Monthly Weather Review* 78, 1–3.

Carnap, R. (1950): Logical Foundations of Probability, Univ. of Chicago Press, Chicago. Cesa-Bianchi, Nicolo, and Gabor Lugosi (2006), *Prediction, Learning, and Games*. Cambridge: Cambridge University Press.

Chernov, A., and Vovk, V. (2009): "Prediction with Expert Evaluator's Advice", *Proceedings of the 20th International Conference on Algorithmic Learning Theory*, ALT'09, Springer, Berlin, 8-22.

Douven, Igor (2011). Abduction. *Stanford Encyclopedia of Philosophy* (March 09, 2011), http://plato.stanford.edu/entries/abduction/)

Douven, I. (forthcoming): "Explaining the Success of Induction", *British Journal for the Philosophy of Science*.

Fallis, D. (2007): "Attitudes towards Epistemic Risk and the Value of Experiments", *Studia Logica* 86, 215-246.

Feldbacher-Escamilla, C. and Schurz, G. (2020): "Optimal Probability Aggregation Based on Generalized Brier Scoring", *Annals of Mathematics and Artificial Intelligence* 2020, doi.org/10.1007/s10472-019-09648-4.

Goldman, Alvin (1999). *Knowledge in a Social World*. Oxford: Oxford University Press.

Gigerenzer, G., Todd, P.M., & the ABC Research Group (1999, eds.): *Simple Heuristics That Make Us Smart*, Oxford Univ. Press, Oxford.

Goodman, N. (1955): Fact, Fiction, and Forecast, Athlone Press, Atlantic Highlands.Kelly

Harman, Gilbert (1986). *Change in View*. Cambridge/MA: MIT Press.

Jekel, M., Glöckner, A., Fielder, S., & Bröder, A. (2012). The rationality of different kinds of intuitive decision processes. *Synthese, 198*, 147–160

Kelly, K.T. (1996): The Logic of Reliable Inquiry, Oxford Univ. Press, New York.Putnam

Lipton, Peter (1991). *Inference to the Best Explanation*. London: Routledge.

Maher, P. (1990): "Why Scientists gather Evidence", British Journal for the Philoso-phy of Science 41, 103-119.

Papineau, David (1993). *Philosophical Naturalism*. Oxford: B. Blackwell.

Psillos, Stathis (1999). *Scientific Realism. How Science Tracks Truth*. London and New York: Routledge.

Putnam, H. (1965): "Trial and Error Predicates and a Solution to a Problem of Mostowski", Journal of Symbolic Logic 30, 49-57.

Reichenbach, H. (1949): The Theory of Probability, University of California Press, Berkeley.

Rendell, Luke et al. (2010): "Why Copy Others? Insights from the Social Learning Strategies Tournament", *Science* 328, 208-213.

Rendell, L., Fogarty, L., and Laland, K. (2009): "Roger's Paradox Recast and Resolved: Population Structure and the Evolution of Social Learning Strategies", *Evolution* 64-2, 534-548.

Salmon, W. C. (1957). Should We Attempt to Justify Induction? *Philosophical Studies* 8, 45-47.

Salmon, W. (1974). "The Pragmatic Justification of Induction", in: R. Swinburne, The Justification of Induction, Oxford University Press, Oxford, 85 – 97.

Selten, R. (1998): "Axiomatic Characterization of the Quadratic Scoring Rule", *Experimental Economics* 1, 43-61.

Skyrms, B. (1975): Choice and Chance, Dickenson, Encinco (4th ed. Wadsworth 2000).Schurz, G.(2008): "The Meta-Inductivist's Winning Strategy in the Prediction Game ", *Philosophy of Science* 75, 2008, 278-305.

Schurz, G. (2012): "Meta-Induction in Epistemic Networks and Social Spread of Knowledge", *Episteme* 9, 2012, 151-170.

Schurz, G. (2017): "No Free Lunch Theorem, Inductive Skepticism, and the Optimality of Meta-Induction", *Philosophy of Science* 84, 825-839.

Schurz, G., and Thorn, P. (2016): "The Revenge of Ecological Rationality: Strategy-Selection by Meta-Induction Within Changing Environments", *Minds and Machines* 26(1), 31-59.

Schurz, G,. (2018): "Optimality Justifications: New Foundations for Foundation-Oriented Epistemology", *Synthese* 195, 3877-3897.

Schurz, G. (2019): Hume's Problem Solved: The Optimality of Meta-Induction, MIT Press, Cambridge/MA.

Shalev-Shwartz, S. and S. Ben-David. (2014): *Understanding Machine Learning. From Theory to Algorithms*. New York: Cambridge University Press.

Solomonoff, R.J. (1964): "A Formal Theory of Inductive Inference", Information and Control, 7, 1-22 (part I), 224-254 (part II).

Sterkenburg, T. (2018): "The Meta-Inductive Justification of Induction", Episteme 15, 1-23. Meta-Inductive Justification of Induction. The Pool of Strategies." Philosophy of Science 86: 981-992.

Thorn, P. and Schurz, G. (2019): "Meta-Inductive Prediction based on Attractivity Weighting: An Empirical Performance Evaluation", *Journal of Mathematical Psychology* 89, 13–30.

Van Cleve, James (2003). Is Knowledge Essay – or Impossible? Externalism as the Only Alternative to Skepticism. In: S. Luper (ed.), *The Skeptics: Contemporary Essays*, (pp. 45-59). Aldershot: Ashgate.

Wolpert, David H. 1996. "The Lack of A Priori Distinctions between Learning Algo-rithms." Neural Computation 8/7: 1341-1390.