Why humble farmers may in fact have bigger potatoes A call for street-smart decision-making in sport

For correspondence:
anne.hecksteden@uibk.ac.at

Anne Hecksteden[1,2] , Niklas Keller[3,4,5], Guangze Zhang[6], Tim Meyer[6], and Thomas Hauser[7,8]

[1]Universität Innsbruck, Institute of Sport Science, Innsbruck, Austria, [2] Medical University Innsbruck, Institute of Physiology, Innsbruck, Austria, [3]Simply Rational – The Decision Institute, Berlin, Germany, [4]Technical University Berlin, Institute of Psychology and Ergonomics, [5]Harding Centre for Risk Literacy, Faculty of Health Science, University of Potsdam, [6]Saarland University, Institute of Sports and Preventive Medicine, Saarbrücken, Germany, [7]German Football Association, Medicine and Science, Frankfurt, Germany, [8]Business and Law School, Faculty of Applied Sport Sciences & Personality, Berlin, Germany

Please cite as: Anne Hecksteden, Niklas Keller, Guangze Zhang, Tim Meyer & Thomas Hauser (2023). Why dumb farmers may in fact have bigger potatoes - A call for street-smart decision-making in sport. *SportRχiv*.

All authors have read and approved this version of the manuscript. This article was last modified on March 12th, 2023.

Authors Anne Hecksteden @*sportaerztin, Guangze Zhang Zhang64182011, and Tim Meyer @ ProfTim_Meyer* can be reached on Twitter.

ABSTRACT

The main task of applied sport science is to inform decision-making in sports practice, that is, enabling practitioners to compare the expectable outcomes of different options (e.g. training programs). The "evidence" provided may range from group averages to multivariable prediction models. By contrast, many decisions are still largely based on the subjective, experience-based judgement of athletes and coaches. While for the research scientists this may seem "unscientific" and even "irrational", it is important to realize the different perspectives: Science values novelty, universal validity, methodological rigor, and contributions towards long-term advancement. Practitioners are judged by the performance outcomes of contemporary, specific athletes. This makes out-of-sample predictive accuracy and robustness decisive requirements for useful decision support. At this point, researchers must concede that under the framework conditions of sport (small samples, multifactorial outcomes etc.) near certainty is unattainable, even with cutting-edge methods that might theoretically enable near-perfect accuracy. Rather, the sport ecosystem favors simpler rules, learning by experience, human judgement, and integration across different sources of knowledge. In other words, the focus of practitioners on experience and human judgement, complemented - but not superseded - by scientific evidence is probably street-smart after all. A major downside of this human-driven approach is the lack of science-grade evaluation and transparency. However, methods are available to merge the assets of data- and human-driven strategies and mitigate biases. This work presents the challenges of learning, forecasting and decision-making in sport as well as specific opportunities for turning the prevailing "evidence vs. eminence" contrast into a synergy.

HOW TO MAKE GOOD DECISIONS IN SPORTS

Applied sport science aims to provide the basis for informed decisions in sports practice – and thereby for the effectiveness and safety of exercise training on all levels. While this statement may seem self-evident, framing applied sport science in the context of decision-making (providing guidance for a concrete, new case) rather than inference (using empirical data to gain new insights into the workings of nature) has important consequences. Perhaps most fundamentally, quality criteria for the output shift from the habitual set around novelty, universal validity, and contribution towards the long-term advancement of science, to a more instrumental set around helpfulness towards achieving current goals and validity under the framework conditions of the specific use case.

Figure 1 illustrates how rules learned from past observations are subsequently brought to bear in forecasting and decision-making. Reliably deciding for the best or at least for a better-than-random of several options (e.g. different training intensities, recovery strategies, or time points of return to sport), calls for a - at least implicit - prediction of what would happen with each of the options. This outcome-oriented perspective of the practitioner makes out-of-sample predictive accuracy a critical requirement for useful scientific evidence. Importantly, agreement with the past (e.g. previously collected data; scientists' habitual "hindsight" perspective) is a poor indicator of generalizability to new cases (practitioners' "foresight" perspective) [1]. Negligible out-of-sample predictive performance can occur despite a near-perfect agreement with the observations from which the rule has been learned [1, 2].
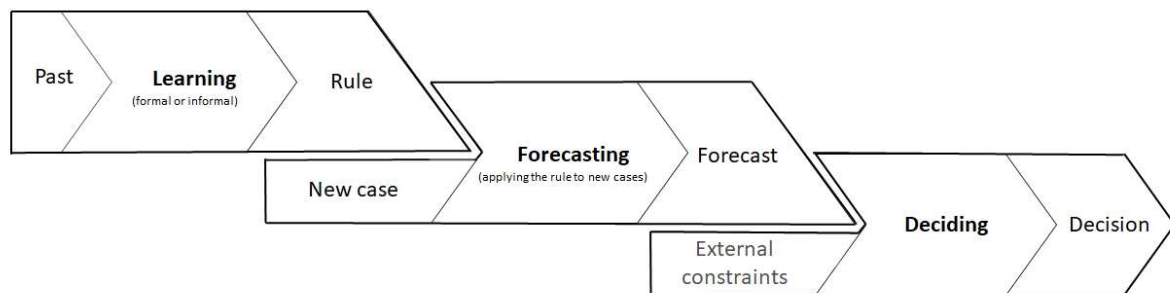

Figure 1.   Learning, forecasting and decision making
             Applied sports science contributes important groundwork for informed decision-
             Making in sports practice – not less and not more.

Prediction models and uncertainty in sports

As sport-related outcomes are usually multifactorially influenced, for a contemporary life scientist, the natural approach to the challenge of "foresight" is arguably a data-driven prediction model based on (multiple-) regression or a more fancy machine learning algorithm. In fact, a comprehensive model could theoretically transform rules previously inferred from data into accurate forecasts - thereby directly linking research and practice. An illustrative example is the field of astronomy, in which physical laws are used to calculate the future positions of celestial bodies with great precision. In the field of sport, however, such a near-clairvoyant model is not only unavailable to date but also unrealistic (given the bewildering complexity of exercise training effects and limited sample sizes [3, 4]), as well as impractical (considering the testing burden associated with assessing the multitude of input variables).

While uncertainty and ambiguity are generally unfavourable in decision making especially if stakes are high, actively embracing the limits of one's knowledge and forecasting tools is an important factor in making rational decisions [5]. An obvious advantage of such "intellectual" [5] or "epistemic" [6] humility is risk and contingency management: Being aware of perhaps being wrong helps mitigate the effects of actual errors (by being attentive and prepared). A less well-recognized benefit is increased freedom in strategy selection and combination: Being aware that "all models are wrong" (even the sophisticated ones currently considered as reference-standard) opens the competition for a greater variety of epistemic approaches [6].

In fact, under real-world constraints (rules inferred from small samples, unknown moderators, measurement error in assessing the specific case etc.) standardized practice based on expected values for a larger reference class can offer superior performance compared to "individualized" decisions based on a multivariable prediction model [7]. Admittedly, deliberately ignoring established influencing factors is counterintuitive. It may be even more surprising that improving predictive performance by limiting model complexity is a matter of course in bioinformatics and machine learning – uniquely "data-driven" fields that are generally thought to seek salvation in complexity and dimensionality [8]. This less-is-more effect can be traced back to the tradeoff between tightly fitting the model to the training data on the one hand and the generalizability of the trained model to other samples and new cases on the other: A highly complex, flexible model will fit limited training data almost perfectly (Remember from highschool that two points can always be fit perfectly with a straight line, three points with a second degree polynomial ...). However, such a close fit to the training data (low bias) is achieved by fitting not only the regularities but also spurious variations leading to marked differences between models trained with different datasets (high variance). However, using a model (or rule) for making

practically-useful forecasts requires its applicability ("generalizability") beyond the cases from which it has been learned (fig. 1). The spuriously high, "useless" performance on the training data is known as "overfitting". Simple models generally provide more robust results (low variance) but may fail to capture as much of the regularities as possible (high bias, underfitting). Optimizing out-of-sample predictive performance requires finding the sweet spot of this bias-variance tradeoff – the location of which is heavily dependent on the amount of available training data. An intuitive illustration can be found in [8]. Note that overfitting is not confined to machine learning but may be seen as a special case of the "narrative fallacy": In hindsight, we will generally find a compelling story of why and how things developed as they did. Projecting this narrative into the future is a completely different story.

Note: The use of the terms "bias" and "variance" differs between machine learning, statistics and psychology. This paragraph reflects the machine learning perspective.

Ecological rationality – generalizing the bias-variance lesson

Selecting the option with the most favourable expected outcome after carefully weighing all available criteria ("maximizing expected utility") has long been considered the essence of rational decision-making. In practice though, the limited amounts of knowledge, information, time, and computational power that human beings possess do not allow for such perfection. The more benign version of the argument is the "effort-accuracy tradeoff": Searching additional information or using additional computation power improves performance but with diminishing returns – and humans stop optimizing when the additional benefit is simply not worth it. The bias-variance tradeoff, however, puts this in a very different perspective: In a world marked by changes and uncertainties, additional effort (that is: more information and complex computation) can actually reduce forecasting accuracy! Put more generally, the real-world performance of a forecasting or decision-making strategy (and therefore the rational choice between strategies) depends on its fit with the framework conditions of the specific use case rather than on the strategies theoretical performance potential under ideal conditions. Table 1 summarizes the key ideas of ecological rationality.

Table 1: Ecological rationality – Key ideas
Don't let the perfect be the enemy of the good!
Renouncing an approach that would be able to provide perfect results under ideal conditions for a simpler alternative can improve performance under real-world circumstances.
There is no golden bullet!
The optimal strategy cannot be identified in general but depends on preconditions of the specific usecase.
The diversity of epistemic approaches is an asset.
The above insights can be generalized beyond tailoring machine learning algorithms to a broader range of forecasting and decision making strategies. Examples for such alternatives range from expert intuition ("gut feeling" [9]) and deliberately simple decision rules (fast-and-frugal heuristics [10]), to debiasing subjective judgements ("crowd intelligence" [11, 12]) and the strategic integration across several sources of information and knowledge

THE SPORT ECOSYSTEM

Key aspects of ecology

The prediction ecology of a specific use case primarily concerns the framework conditions for learning rules that can subsequently be used to make useful predictions for new cases. In a wider sense, the set of options between which to decide, as well as practical constraints and evaluation criteria regarding the decision-making process and the decision maker are also part of the decision ecology. A clear and concise overview is complicated by varying terminologies and focal points associated with different approaches and the fields within which they have been developed. Here we aim to provide a unified overview by identifying four main aspects: Predictability, explanatory (cues), criterion variables, and the number of precedents to learn from.

Predictability

Predictability of a specific outcome can be limited by random chance playing a role in it (aleatoric uncertainty) and/or by practical epistemic constraints (epistemic uncertainty). An example of high aleatoric uncertainty is contact injuries in team sports which are due to the

player being "in the wrong place and situation at the wrong time" with few (if any) risk factors that could – even in principle - be identified in advance [2]. By contrast, overload injuries are an example of mainly epistemic uncertainty: The loading capacities of e.g. a runner's metatarsal bones and the exercise-induced (local) load are existing quantities and knowable in principle, but we cannot assess or estimate them precisely. Importantly, predictability is sensitive to the timeframe. In many situations, short-term "anticipation" is easier and more accurate than long-term forecasting. A theory-based idea of predictability for the outcome in question is crucial for topic and strategy selection, risk management, and spotting exaggerated claims of model performance and overconfidence in subjective judgements.

In the field of sports predictability varies widely between negligible (contact injuries in team sports) and good (change in running performance with change in body weight [13, 14]). Many outcomes fall in between: They are predictable to some (potentially meaningful) extent and gradual improvements rely on identifying dominant features and smart strategy selection/combination.

Note that there are several classifications of "worlds" tightly linked to predictability [15-19]. While the fields of origin and the dimensions differ, a unifying rationale is tentatively illustrated by the small tree in figure 2. There are only two nodes: (1) Are the governing rules known or do they have to be learned from limited amounts of previous observations? And (2) Can we realistically hope to learn rules with near-perfect out-of-sample predictive accuracy under real-world conditions? The world of sport is located in the lower right, reflecting limited predictability and considerable uncertainty.
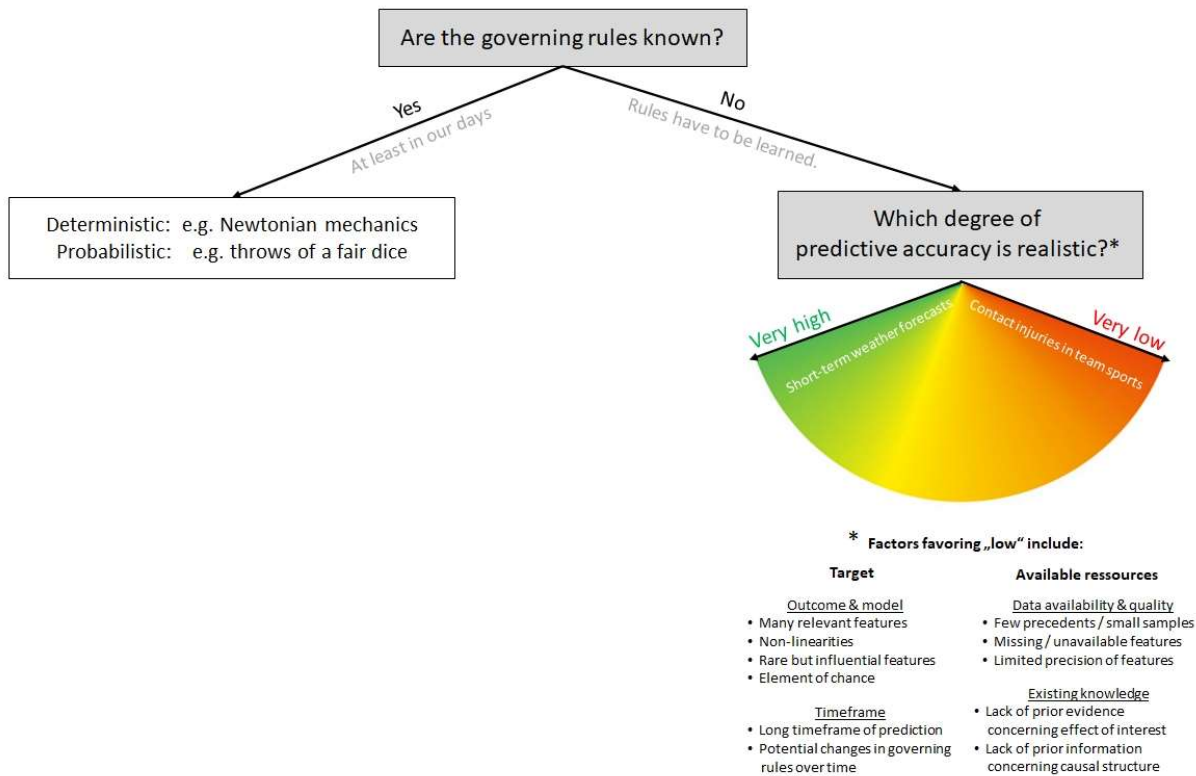
Are the governing rules known?

Yes
At least in our days

No
Rules have to be learned.

Deterministic: e.g. Newtonian mechanics
Probabilistic: e.g. throws of a fair dice

Which degree of
predictive accuracy is realistic?*

Very high
Short-term weather forecasts

Contact injuries in team sports
Very low

* Factors favoring „low" include:

Target

Outcome & model
• Many relevant features
• Non-linearities
• Rare but influential features
• Element of chance

Timeframe
• Long timeframe of prediction
• Potential changes in governing rules over time

Available ressources

Data availability & quality
• Few precedents / small samples
• Missing / unavailable features
• Limited precision of features

Existing knowledge
• Lack of prior evidence concerning effect of interest
• Lack of prior information concerning causal structure

Figure 2. Predictability in different "worlds" -
Can we realistically expect to procure generalizable rules?

Note: The classification of a specific case may change over time at both nodes. Example for node 1: The laws of Newtonian mechanics once had to be discovered, but have since aquired the status of „law of nature". Example for node 2: Conceptual and technological advances can gradually improve predictability e.g. in weather forecasting.

Cues

Using currently available cues (explanatory variables) to predict the state of a complex system at a future time point is the essence of forecasting - and thereby of evaluating different options between which to decide (fig. 1). The **number** of cues, their relative importance and potential interactions determine the complexity of the rule to be learned, and thereby the number of precedents needed to do so. The **type** of relevant cues (manifest/objective quantity, latent/complex construct, subjective perception, social) is another important aspect for

matching a forecasting strategy to the environmental structure. Further aspects of the set of cues are their **redundancy** (overlap in meaning), **accessibility** (e.g. testing burden), **availability** (at the time point of making the forecast), and finally the **uncertainty of cue values** (e.g. due to measurement error).

The general, cue-related specifics of the sport ecosystem mainly concern number and type. Most outcomes of major interest in sports are multifactorially influenced with at least a considerable subset of influencing factors being of noteworthy and consistent importance. Although it is important not to confound "cue" with "cause", a multitude of relevant explanatory factors is highly plausible. Moreover, many conceptually important explanatory factors are not manifest, directly observable quantities (e.g. the speed or the number of sprints) but complex constructs (e.g. recovery needs or movement quality) that have to be inferred from indicators of their various dimensions. Taking recovery needs as an example, indicators may include blood born markers, heart rate and heart rate variability measures and tests of neuromuscular performance [20]. While this approach is objective and scalable, it further increases the number of parameters to be fit and thereby the sample size required for learning. As a complementary asset of the sport ecosystem, the corporeality of exercise regularly enables direct access to the target construct as a subjective perception. The human brain evolved to integrate exercise-related cues into a feeling (athlete) or impression (coach) of fatigue, movement quality etc. Again taking recovery needs as an example, questionnaire results almost uniformly outperform objective indicators – at least in the context of scientific studies with no conflicts of interest on the side of athletes or coaches [20].

Criterion

To develop and evaluate forecasting accuracy, it is crucial to compare predicted and actual outcomes. Therefore, the timely availability of an unambiguous criterion is an important aspect of a learning environment. In machine learning this consideration is embodied in the concept of "supervised learning", but it is equally important for (human) learning from experience. The above considerations regarding manifest, directly observable quantities and complex, latent constructs also apply to the criterion variable.

Precedents

The number of similar cases from which to learn is fundamental for obtaining generalizable rules. This consideration is embodied by the statistical proverb "Repetition is the

key to separate trait from chance." While in the life science context, the critical quantity is the number of cases or events, the rule also applies to learning by experience. Importantly, increasing the number of explanatory factors can't compensate for a limited number of precedents but aggravates the risk of overfitting.

A characteristic feature of high-level sport is the small number of athletes [3] and even outside high-performance environments sample sizes in sport science are usually limited. Therefore, "greedy" approaches (high-dimensional biomarkers / "omics", "deep learning") usually fail to provide useful out-of-sample predictive accuracy [3]. Note that experienced coaches may have had more opportunities for learning than any scientific trial.

The sport ecosystem – further relevant aspects

Beyond the peculiar expression of the 4 main characteristics above, the sport ecosystem includes further relevant specifics, 3 of which will be discussed below.

The ecosystem in high-performance sport is populated by outliers

High-performance athletes are, more-or-less by definition, exceptions from the rule. This impedes the generalizability of learning outcomes and complicates the identification of implausible results.

Specific constraints on decision-making

For example, in team sports a specific number of players has to be lined up for a competitive match – even if all players have a high predicted probability of getting injured. Moreover, the trade-off between expected consequences for individual health and team success, respectively, varies between players. These aspects differ from decision-making in other areas of health care where decisions are based exclusively on the expected (health) consequences for the concerned individual [2].

Importance of avoiding the big mistakes

Athletic training and performance development are long-term and therefore involve a large number of decisions ranging from strategic to mundane. In this context, it is important to keep in mind that a single big mistake may outweigh the positive effects of a large number of gradual optimizations. Therefore, it is crucial to identify forecasts that are "way off" or cases that

are "off limits" and for which an otherwise successful model may not be applicable (e.g. due to rare but influential characteristics [21]). Moreover, as for predictability in general, it is important to have a theory-based expectation regarding the distribution of forecasting errors. It makes a big difference for risk management whether errors are more or less normally distributed or if long streaks of accurate forecasts are punctuated by complete failures [21]. Unfortunately, infrequent but massive errors are not well represented by common measures of predictive accuracy. Therefore, plausibility checks are essential for robust decision-making with imperfect knowledge. Importantly, this requires cross-comparison between different sources of knowledge - especially when "plausible" may not be approximated by "within a group-based reference range". To date, such non-algorithmic critical thinking and common sense are still a privilege of humans.

The sport ecosystem in a nutshell

Taken together, the key challenge of the sport ecosystem is complexity complicated by a small, sometimes tiny number of precedents to learn from. On the assets side, there is the corporeality of physical exercise which provides direct subjective access to complex exercise-related features, the feedback provided by daily practice and competition performance, and the longstanding, immersive experience of professional coaches, support staff and athletes. From the perspective of task analysis, the priority of avoiding big mistakes (robustness) and sport-specific external constraints on decision-making have to be taken into account.

LEARNING, FORECASTING & DECISION-MAKING STRATEGIES

In the context of sports, only two contrasting strategies are generally considered: (1) The life science approach including group-based scientific evidence and data-driven prediction models and (2) Experiential knowledge and expert intuition. However, this "evidence vs. eminence" dichotomy ignores the diversity of available options. In particular, there are two potential amendments with promising fit to the sports ecosystem: Deliberately frugal decision rules (heuristics) and debiasing and aggregating subjective human judgements ("crowd intelligence"). Finally, the individual strategies are not mutually exclusive but may be synergistically combined. The following sections discuss assets, drawbacks and potential synergies. A conceptual overview is provided in figure 3.
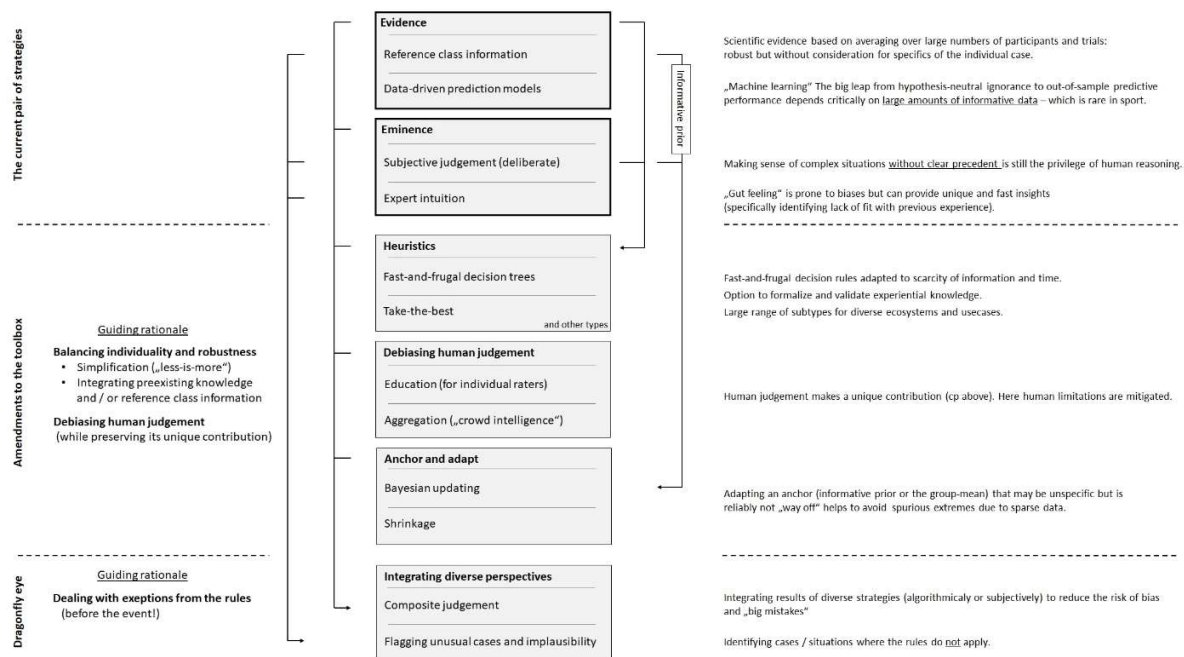
Figure 3 content (strategy diagram):

The current pair of strategies

Evidence	Scientific evidence based on averaging over large numbers of participants and trials: robust but without consideration for specifics of the individual case.
Reference class information	
Data-driven prediction models	„Machine learning" The big leap from hypothesis-neutral ignorance to out-of-sample predictive performance depends critically on large amounts of informative data – which is rare in sport.
Eminence	Making sense of complex situations without clear precedent is still the privilege of human reasoning.
Subjective judgement (deliberate)	
Expert intuition	„Gut feeling" is prone to biases but can provide unique and fast insights (specifically identifying lack of fit with previous experience).

Informative prior

Amendments to the toolbox

Guiding rationale

Balancing individuality and robustness
- Simplification („less-is-more")
- Integrating preexisting knowledge and / or reference class information

Debiasing human judgement (while preserving its unique contribution)

Heuristics	Fast-and-frugal decision rules adapted to scarcity of information and time. Option to formalize and validate experiential knowledge. Large range of subtypes for diverse ecosystems and usecases.
Fast-and-frugal decision trees	
Take-the-best and other types	
Debiasing human judgement	Human judgement makes a unique contribution (cp above). Here human limitations are mitigated.
Education (for individual raters)	
Aggregation („crowd intelligence")	
Anchor and adapt	Adapting an anchor (informative prior or the group-mean) that may be unspecific but is reliably not „way off" helps to avoid spurious extremes due to sparse data.
Bayesian updating	
Shrinkage	

Dragonfly eye

Guiding rationale

Dealing with exceptions from the rules (before the event!)

Integrating diverse perspectives	Integrating results of diverse strategies (algorithmicaly or subjectively) to reduce the risk of bias and „big mistakes"
Composite judgement	
Flagging unusual cases and implausibility	Identifying cases / situations where the rules do not apply.

Figure 3.   Strategy selection and combination in the sport ecosystem – A first serve

The life-science approach ("Evidence")

In many ecosystems, standardized expectations based on a large reference group (e.g. results from large randomized controlled trials or meta-analyses) are superior to expert judgements [22] and a hard-to-beat benchmark for data-driven, individualized predictions [7]. However, it has to be kept in mind that the predictive value of trials with sample sizes typical for sport science is low [3, 23]. This means that many (perhaps most [24]) novel findings are false. While these busts will eventually be sorted out during the research process [25], they make early adoption a hazardous business for practitioners.

Data-driven prediction models aim on reducing uncertainty by considering a comprehensive set of explanatory variables. They combine the theoretical potential for (near-) perfect predictive performance with objectivity and scalability (e.g. by implementing the trained model in a digital decision support system). Moreover, computation capacity and large amounts of data (from wearables, player tracking, power meters, smartphone apps etc.) are readily available today. However, critical requirements for unleashing the potential of data-driven prediction models are a large number of precedents and an informative (!) panel of explanatory variables. As already pointed out, these requirements are generally not met by the sport

ecosystem. Importantly, this does not rule out that in some sport-related applications data-driven prediction models may be helpful – particularly when a limited number of dominant cues or patterns can be identified [2] and/or as part of composite strategies [3, 26]. A recent illustrative example for the latter is the combination of data-driven prediction with coaches subjective judgment in talent identification [9].

Beyond predictive accuracy, interpretability is an important asset in the context of decision support. While in theory "black box" predictions of a relevant target (e.g. injury risk) made with a well-validated model can be useful, ideally, predictive accuracy coincides with causal interpretability. In other words: The explanatory variables relate to determinants in causal concepts and it is known which cue values drive a specific prediction. An illustrative example is the monitoring of injury risk: While an accurate estimate of injury risk may in itself be worthwhile (e.g. to avoid exposure when the risk estimate is high), knowing the factors that lead to an elevated risk would enable a more targeted response.

Finally, it has to be kept in mind that although current software packages make "machine learning" doable for the subject matter scientist, this ease is deceiving and a lack of expertise (or rigor) in the finer details of model fitting and validation can easily lead to spuriously high (!) estimates of model performance [2]. A salient but regularly overlocked pitfall is information leakage [27], the risk of which is particularly high when working with longitudinal (e.g. monitoring) data [2].

Learning by experience and expert judgement ("Eminence")

Arguably, experience-based subjective judgments are still the prevailing basis of decision-making in sports. While from the life-scientists perspective this may sometimes seem "unscientific", "irrational" or even stubborn, in fact, characteristic features of the sport ecosystem favour this approach. To begin with, experienced professional coaches and other support staff typically had access to more precedents than can be included in any scientific trials. Together with the direct subjective perceptibility of complex exercise-related cues, regular feedback provided during daily practice and competition, and guidance from formal training, this favours the build-up of robust experiential knowledge. Moreover, making viable forecasts in complex situations with very few clearly identifiable precedents is a characteristic feature of human reasoning. This human faculty exploits higher-order mental capacities such as thinking in analogies to make sense of diverse and incomplete information and remains hard to emulate for artificial intelligence.

Note that subjective assessments do not necessarily arise from an unconscious "black box". While this is a characteristic feature of intuitions and "gut feelings", subjective judgements and forecasts can be the result of targeted information search and conscious reasoning with an explicit line of argument as well as an estimate of uncertainty [28, 29]. While the assets of the latter are well supported [28, 29] (particularly when complexity is combined with sparse data as it is the case in sport) the potential contribution of expert intuition and "gut feelings" is less clear. Arguably, expert intuitions should be particularly considered for spotting abnormalities e.g. cases that do not belong to the reference class despite fulfilling formal inclusion criteria [30] (figure 3). However, this remains to be empirically verified in the context of sport.

Deliberately simple decision rules (Heuristics)

Heuristics are simple "rules of thumb" that enable fast decisions without the effort and resources needed for considering all available information. Generally, heuristics are viewed in light of an effort-accuracy trade-off: "Quick-and-dirty" solutions, necessary to get along with the deluge of everyday decisions that do not merit the effort of optimization. However, as already pointed out, simple rules can also be more accurate than extensive strategies that consider more cues and use more extensive computation methods [10]. Empirical results supporting a "less-is-more" effect have been reported in a wide range of fields [7, 31-38].

A structured introduction to the science of heuristics is beyond the scope of this work and has been provided by experts in the field [10, 39]. However, it is important to identify two main perspectives: In the seminal work of Tverski and Kahneman [40] heuristics are simplifications of judgemental operations that are unconsciously used by humans and rely on subjective cues such as representativeness (how much the specific case evokes a certain class) or availability (how easily similar cases come to mind). While the authors explicitly state that "in general, these heuristics are quite useful", the focus is on the biases associated with such intuitive short-cuts e.g. insensitivity to base rates, sample size, and predictability. The deficiencies of heuristics are demonstrated using the "rational" judgement or choice as a comparator. However, while this is straight forward for situations in which the optimal solution is known, in most practically relevant situations the optimal solution is unknown or even unknowable. Therefore, Gigerenzer and colleagues modified this view by defining heuristics as efficient judgemental operations that deliberately use only part of the potentially available information and simple computation [10, 39]. Emphasize is put on exploiting "less-is-more" effects and on formalizing and evaluating heuristics (e.g. for use in decision support tools [41, 42]). This integrates heuristics at par into the larger toolset of forecasting and decision support.

Regarding fit with the sport ecology, simple models (heuristics) are generally favoured by sparse data. Moreover, the corporeality of exercise offers the option to leverage the innate capacities of the human mind (e.g. perceiving physical exertion or recognizing movement patterns that indicate it in others) and thereby favors the validity of subjective cues. Finally, formalizing experiential knowledge as heuristics [41, 42] offering a potential hub between . Taken together, the general fit between heuristics and the sports ecosystem seems to be almost exemplary. Readers interested in the rationales and rules for selecting specific heuristics are referred to [43].

Taming and harnessing subjective judgements (Crowd intelligence)

Despite the positive perspective on experiential knowledge and expert judgment presented above, there are also major downsides of this "human-driven" approach. These include limitations in attention, time and memory as well as the numerous biases introduced e.g. by limited or irrelevant information and wishful thinking [40]. Moreover, the informal process and verbal or even implicit judgements (as opposed to quantitative probability estimates) complicate the evaluation of performance. Taken together, subjective judgements are a serious option when trying to make forecasts in complex situations with incomplete information and very few specific precedents e.g. when trying to mitigate injury risk in an elite athlete. However, the risk of bias and a lack of (objective) verification of performance are downsides of this uniquely human contribution.

In fields that regularly deal with this conundrum at high stakes (e.g. intelligence analysis [44, 45]) techniques for mitigating these limitations have been developed. First and foremost, objective evaluation of subjective judgements is enabled by unambiguous targets (including criterion and timeline) and quantitative estimates [29, 44]. Of course, this requires a commitment to accountability, feedback and continuous improvement on the part of the raters. As a next step, the accuracy of individual raters may be increased by feedback and advice on good judgement practice (e.g. incremental updating of reference class information [46]). Beyond these basic measures, further improvements are mainly achieved by having not one but many raters [11, 44] – in exact analogy to averaging several measurements e.g. of VO2max [47] to reduce the impact of measurement error. The concept of "crowd intelligence" or "wisdom of the crowd" dates back to the beginning of the democratic era [12] and posits that in many situations aggregating subjective judgements from a large number of independent raters (on average) leads to a more accurate estimate than the judgement of a single rater [48] – even in the case of superior expertise, experience [12] and access to exclusive information [44]. While initially

simple averaging was used [12], today more sophisticated methods for aggregation are available [11, 29]. Implementing the collection, aggregation and presentation of subjective judgments in a web application or smartphone app can be the final step of taming subjective judgements and integrating them into the harnessed team of forecasting and decision support tools [49]. Taken together, aggregated subjective judgments are a promising option for improving decision support in high-level sports – specifically in "unique" situations in which statistical learning is doomed to fail, clearcut heuristics are not available and the popularity of a sport provides a large and motivated "crowd" (e.g. football).

Integrating diverse sources of information and knowledge

In the sports ecosystem, each of the approaches to learning, forecasting and decision-making comes with considerable limitations. Therefore, it seems promising to search for synergies and complementary combinations – in particular between data-driven and human-driven strategies but also between existing knowledge and new data. Mundane (yet essential) examples are the support of experiential learning by formal training [43] (e.g. for obtaining a coaching licence) and common-sense-based plausibility checks. While there are countless ways to formally combine sources of information and knowledge, it may be helpful to identify two main categories: (1) Building upon preexisting knowledge and (2) Single-stage integration of results from diverse strategies.

Leveraging prior knowledge – Bayesian updating, shrinkage, and causal inference

When the small number of precedents and / or the acceptability of extensive study requirements are limiting factors (as is typically the case in sport) prior knowledge - which may concern the magnitude and/or the causal structure of the effect in question - may be used to augment current data.

Anchor and adapt: Fortunately, at least in high-performance sport, we usually have a defendable prior expectation about the direction and magnitude of the effect(-s) in question. Options to formally implement a "hub" with new data range from using base rates as anchors[50], over Bayesian updating of informative priors [51, 52] and shrinking individual forecasts towards the group average [53], to pre-trained models in machine learning [2, 3]. Ultimately, these methods balance individuality and robustness by including an "anchor" based

on a larger reference class. The following references provide some worked examples for the integration of preexisting knowledge and current data in sport [2, 51, 52, 54].

Leveraging causal knowledge: Insights into the causal structure of the effect under investigation can be used to gain more information from available data, specifically for improving out-of-sample predictive accuracy, robustness and interpretability [55]. While fully implementing causal inference arguably requires expert collaboration, explicitly specifying subject matter knowledge (or assumptions) in a causal diagram is already an important step enabling transparent scrutiny and identification of pitfalls [55]. Note that existing causal knowledge or respective assumptions are involved in any trial even if the statistical analysis is purely data-driven and model-free (e.g. for selecting proper stratification criteria and standardization measures).

Integrating diverse perspectives – Dragonfly eye strategies and triangulation

Supplementary knowledge and different perspectives may also arise in parallel, e.g. by applying several of the above strategies to the same use case. As already pointed, integrating judgements from diverse and independent raters can reliably improve accuracy and avoid extreme outliers. This principle of "crowd intelligence" also applies to non-human sources. A salient example is ensemble methods in machine learning such as random forests (which rely on purposefully increased diversity and independence followed by aggregation). The combination of forecasts across different machine learning methods is referred to as "consensus forecasting" or "committee machine". Of course, different, potentially complementary access routes may also be combined in a non-algorithmic way to gain a human-driven, composite assessment (dragonfly eye strategy) or to identify spurious extremes that may arise as artefacts of a particular method (plausibility control). In the context of scientific research, increasing the robustness of insights by combining diverse lines of evidence is known as "triangulation" [56].

CONCLUSION

Taken together, the above considerations call for (deliberate and reflective) street-smart decision-making in sport: Waiving extensive proceedings or authorities that are axiomatically considered optimal (even if they would theoretically be under ideal conditions) for strictly outcome-oriented approaches that are less elegant and ambitious but adapted to and proven in the environment in question. Specifically, in the sport ecosystem, key amendments to data-driven evidence and individual eminence are:

- Using the head start provided by <u>preexisting knowledge</u>
- Targeting the sweet spot of model complexity by <u>deliberate simplification</u>
- Harnessing uniquely <u>human contributions</u>
- Replacing gold-standard dogmas by using <u>synergies between diverse approaches</u>

Contributions

Contributed to development of concept: AH, NK, TH

Drafted and/or revised the article: AH, GC, NK, TM, TH

Approved the submitted version for publication: AH, GC, NK, TM, TH

Data and Supplementary Material Accessibility

Not applicable

REFERENCES

1. Hastie T, Tibshirani R, Friedman J. The Elements of Statistical Learning - Data mining, Inference, and Prediction. 2nd ed. Springer; 2017.
2. Hecksteden A, Schmartz G, Egyptien Y, Keller A, Meyer T. Forecasting soccer injuries by combining screening, monitoring and machine learning. Science and Medicine in Football. 2022. epub ahead of print. https://doi.org/10.1080/24733938.2022.2095006
3. Hecksteden A, Kellner R, Donath L. Dealing with small samples in football research. Science and Medicine in Football. Science and Medicine in Football. 2022;6:389-397. https://doi.org/10.1080/24733938.2021.1978106.

4. Greenland S., Mansournia M.A., Altman DG. Sparse data bias: a problem hiding in plain sight. BMJ. 2016; https://doi.org/10.1136/bmj,i1981

5. Porter T, Elnakouri A, Meyers EA, Shibayama T, Jayawickreme E, Grossmann I. Predictors and consequences of intellectual humility. Nat Rev Psychol. 2022;1:524-536. https://doi.org/10.1038/s44159-022-00081-9

6. Mellers B, Tetlock P, Arkes HR. Forecasting tournaments, epistemic humility and attitude depolarization. Cognition. 2019;188:19-26. https://doi.org/10.1016/j.cognition.2018.10.021

7. Senn S. Mastering variation: variance components and personalised medicine. Stat Med. 2016;35:966-977. https://doi.org/10.1002/sim.6739

8. Gigerenzer G, Brighton H. Homo Heuristicus: Why biased minds make better inferences. Topics in Cognitive Science. 2009;1:107-43. https://doi.org/10.1111/j.1756-8765.2008.01006.x.

9. Sieghartsleitner R, Zuber C, Zibung M, Conzelmann A. Science or Coaches' Eye? - Both! Beneficial Collaboration of Multidimensional Measurements and Coach Assessments for Efficient Talent Selection in Elite Youth Football. J Sports Sci Med. 2019;18:32-43.

10. Gigerenzer G, Gaissmaier W. Heuristic decision making. Annu Rev Psychol. 2011;62:451-82. https://doi.org/10.1146/annurev-psych-120709-145346

11. Ungar L, Mellers B, Satopää V, Baron J, Tetlock P, Ramos J, et al. The Good Judgment Project: A large scale test of different methods of combining expert predictions. Association for the Advancement of Artificial Intelligence. 2012; retrieved from: https://www.cis.upenn.edu/~ungar/papers/forecast_AAAI_MAGG.pdf.

12. Galton F. Vox populi. Nature. 1907;75:450-1.

13. Cureton KJ, Sparling PB, Evans BW, Johnson SM, Kong UD, Purvis JW. Effect of experimental alterations in excess weight on aerobic capacity and distance running performance. Med Sci Sports. 1978;10:194-9.

14. Zacharogiannis E, Paradisis G, Magos S, Plavoukos I, Dagli F, Pilianidis T, et al. The Effect Of Acute Body Mass Reduction On Metabolism And Endurance Running Performance. Med Sci Sports Exerc. 2017;49:194. https://doi.org/10.1249/01.mss.0000517367.65054.59

15. Savage LJ. The Foundations of Statistics,2nd ed. Dover Publications Inc.; 1972.

16. Reiter R. On closed world data bases. In: H. G, editor. Logic and Data Bases: Plenum Press, New York; 1978.

17. Hanheide M, Göbelbecker M, Horn GS, Pronobis A, Sjöö K, Aydemir A, et al. Robot task planning and exploration in open and uncertain worlds. Artificial Intelligence. 2017;247:119-50.

18. Lendrem DW, Lendrem BC, Woods D, Rowland-Jones R, Burke M, Chatfield M, et al. Lost in space: design of experiments and scientific exploration in a Hogarth Universe. Drug Discov Today. 2015;20:1365-1371. https://doi.org/10.1016/j.drudis.2015.09.015

19. Hogarth RM. Intuition: a challenge for psychological research and decision making. Psychol Inquiry. 2010;21:338-53.

20. Kellmann M, Bertollo M, Bosquet L, Brink M, Coutts AJ, Duffield R, et al. Recovery and Performance in Sport: Consensus Statement. Int J Sports Physiol Perform. 2018;19:1-6. https://doi.org/10.1123/ijspp.2017-0759

21. Taleb NN. Black Swans and the Domains of Statistics. The American Statistician. 2007;61:1-3. https://doi.org/10.1198/000313007X219996

22. Grove WM, Zald DH, Lebow BS, Snitz BE, Nelson C. Clinical versus mechanical prediction: a meta-analysis. Psychol Assess. 2000;12:19-30.

23. Mesquida C, Murphy J, Lakens D, Warne J. Replication concerns in sports and exercise science: a narrative review of selected methodological issues in the field. R Soc Open Sci. 2022;9:220946. https://doi.org/10.1098/rsos.220946

24. Ioannidis JP. Why most published research findings are false. PLoS Med. 2005;2:e124. https://doi.org/10.1371/journal.pmed.0020124

25. Ioannidis JP. Why replication has more scientific value than original discovery. Behav Brain Sci. 2018;41(E137). https://doi.org/10.1017/S0140525X18000729.

26. Dellerman D, Ebel P, Söllner M, Leimeister J. Hybrid Intelligence. Bus Inf Syst Eng. 2019;61:637-43.

27. Gibney E. Could machine learning fuel a reproducibility crisis in science? Nature. 2022;epub ahead of print. https://doi.org/10.1038/d41586-022-02035-w

28. Mellers B, Ungar L, Baron J, Ramos J, Gurcay B, Fincher K, et al. Psychological strategies for winning a geopolitical forecasting tournament. Psychol Sci. 2014 1;25:1106-1115. https://doi.org/10.1177/0956797614524255

29. Mellers BA, Tetlock PE. From discipline-centered rivalries to solution-centered science: Producing better probability estimates for policy makers. Am Psychol. 2019;74:290-300. https://doi.org/10.1037/amp0000429

30. Klein G, Calderwood R, Clinton-Cirocco A. Rapid decision making on the fire ground. Prodeedings of the Human Factors and Ergonomics Society Annual Meeting. 1986;30. https://doi.org/10.1177/154193128603000616

31. Wübben M, Wagenheim F. Instant customer base analysis: managerial heuristics often "get it right". J Mark. 2008;72:82-93.

32. Artinger F, Kozodoi N, Wangenheim F, Gigerenzer G. Recency: prediction with smart data. Am Mark Assoc Winter Conf Proc 2018;29:L2.

33. Barber B, Odean T. Trading Is Hazardous to Your Wealth: The Common Stock Investment Performance of Individual Investors. The Journal of Finance. 2000;55:773-806.

34. Kadlec D. Why US funds are not up to par. Time. 1997; January 27:32-33.

35. Snook B, Zito M, Bennell C, Taylor P. On the complexity and accuracy of geographic profiling strategies. J Quant Criminol. 2005;21:1-26.

36. Green LA, Mehr D. What alters physicians' decisions to admit to the coronary care unit? Journal of Family Practice. 1997;45:219-26.

37. Lichtman A. Predicting the next president: The keys to the White House. Rowman and Littlefield. 2016.

38. Serwe S, Frings C. Who will win Wimbledon? The recognition heuristic in predicting sports events. J Behav Decis Mak. 2006;19:321-32.

39. Gigerenzer G, Todd PM. Simple heuristics that make us smart: Oxford University Press; 1999.

40. Tversky A, Kahneman D. Judgment under Uncertainty: Heuristics and Biases. Science. 1974;185:1124-1131. https://doi.org/10.1126/science.185.4157.1124

41. Keller N, Katikopoulos K. On the role of psychological heuristics in operational research; and a demonstration in military stability operations. European Journal of Operational Research. 2015. https://doi.org/10.1016/j.ejor.2015.07.023

42. Keller N, Czienskowski U, Feufel M. Tying up loose ends: a method for constructing and evaluating decision aids that meet blunt and sharp-end goals. Ergonomics. 2014. https://doi.org/10.1080/00140139.2014.917204

43. Gigerenzer G, Reb J, Luan S. Smart Heuristics for Individuals, Teams, and Organizations. Annual Review of Organizational Psychology and Organizational Behavior. 2022;9:171-198. https://doi.org/10.1146/annurev-orgpsych-012420-090506

44. Tetlock PE, Mellers BA, Scoblic JP. Bringing probability judgments into policy debates via forecasting tournaments. Science. 2017;355:481-483. https://doi.org/10.1126/science.aal3147

45. Ulfelder J. Using the "Wisdom of (Expert) Crowds" To Forecast Mass Atrocities (Report). 2014. https://doi.org/10.2139%2Fssrn.2418980

46. Atanasov P, Witkowski J, Ungar L, Mellers B, Tetlock P. Small steps to accuracy: Incremental belief updaters are better forecasters. Organizational Behavior and Human Decision Processes. 2020;160:19-35. https://doi.org/10.1016/j.obhdp.2020.02.001

47. Bouchard C, Rankinen T. Individual differences in response to regular physical activity. Med Sci Sports Exerc. 2001;33(6 Suppl):S446-51; discussion S52-3.

48. Surowiecki J. The wisom of crowd: Why the many are smarter than the few and how collektive wisdom shapes business, economies, societies and nations. New York: Anchor; 2005.

49. https://goodjudgment.com/, Accessed 24.06.2022.

50. Kent DM, Steyerberg E, van Klaveren D. Personalized evidence based medicine: predictive approaches to heterogeneous treatment effects. BMJ. 2018;10:k4245. https://doi.org/10.1136/bmj.k4245

51. Hecksteden A, Pitsch W, Julian R, Pfeiffer M, Kellmann M, Ferrauti A, et al. A New Method to Individualize Monitoring of Muscle Recovery in Athletes. Int J Sports Physiol Perform. 2017;12:1137-42. https://doi.org/10.1123/ijspp.2016-0120

52. Hecksteden A, Skorski S, Egger F, Buder F, Kellner R, Meyer T. Dwarfs on the shoulders of giants: Bayesian analysis with informative priors in elite sports research and decision making. Frontiers in Sports and Active Living. 2022. https://doi.org/10.3389/fspor.2022.793603..

53. Senn S. Transposed conditionals, shrinkage, and direct and indirect unbiasedness. Epidemiology. 2008;19:652-654; discussion 7-8. https://doi.org/10.1097/EDE.0b013e318181b3e3

54. Sottas PE, Baume N, Saudan C, Schweizer C, Kamber M, Saugy M. Bayesian detection of abnormal values in longitudinal biomarkers with an application to T/E ratio. Biostatistics. 2007;8:285-96. https://doi.org/10.1093/biostatistics/kxl009

55. Pearl J. Causal inference in statistics: An overview. Statistics Surveys. 2009;3:96-146. https://doi.org/10.1214/09-SS057

56. Munafo MR, Davey Smith G. Robust research needs many lines of evidence. Nature. 2018;553:399-401. https://doi.org/10.1038/d41586-018-01023-3

REFERENCES

[1] A. J. Figueredo and P. S. A. Wolf. "Assortative pairing and life history strategy - a cross-cultural study". In:Human Nature 20 (2009), pp. 317–330. URL: www.google.com.