

# What makes a good sport?

SIR — Professional sports, to survive, must attract fans, and so are under adaptive pressure to satisfy fans' desires by providing some combination of grace, strength, brilliance, violence and excitement. This is done, we suggest, with a finely calibrated mix of skill and chance in the outcome. Contests with too much chance are pointless as measures of relative ability. Those with too little chance in the mix provide no sus-

while those whose games are largely chance have long ones. Thus these sports, differing enormously in their particulars, converge towards the same reliability in a season. In contrast to what one might expect, season length is not an arbitrary product of historical, meteorological or other such constraints. Instead, almost all the variance in season length can be explained as an adaptation to fans' desires for the proper mix of

RELIABILITY AND SEASON LENGTH OF PROFESSIONAL LEAGUE SPORTS

Sport	Single-game reliability	Games per season	Season reliability
Baseball	0.008	162	0.566
Hockey	0.043	84	0.789
Soccer	0.065	42	0.703
Basketball	0.089	82	0.890
Football	0.118	16	0.681
Rugby Union	0.136	22	0.776
Rugby League	0.139	30	0.829

The split-half reliability coefficient, representing the extent to which a team's success in half the games predicts its success in the other half, can be adjusted (using the Spearman-Brown formula) to estimate the reliability of the entire season (twice as many games as the split half) or the reliability of a one-game season. This procedure was followed for 7 sports in 3 recent seasons. Season length is strongly negatively related to single-game reliability,  $r(s) = -87, P < 0.01$ . (For further details of methods, please contact the author.)

pense. The superior team should probably, but not certainly, win. Across all sorts of sports, those with the proper mix can thrive, while those that are too reliable or insufficiently reliable should never catch on. We find evidence that nine professional sports, from rugby to chess, control their reliability to create outcomes of roughly equal parts skill and chance. It is not that fans actually know how reliable a sport is, but rather their satisfaction depends on a sport having the right blend of skill and chance. Sports reflect this preference, shaped by the aggregated satisfaction of the public.

This notion of an optimal reliability for sports was tested by examining the extent to which sports regulate their reliability with the length of their seasons. Given a particular level of reliability within each game (which is influenced by spreading talent between teams, allowing small differences to have large consequences, rolling dice and the like), the reliability of the season is a function of the number of games played. A sport whose games are highly unreliable can increase the proportion of skill in the outcome by playing many games in a season. Similarly, a sport that, for structural reasons, has reliable games can increase the role of chance by having a short season. We calculated for seven league sports the reliability of each game, and compared this to the number of games played in a season (see table).

This demonstrated almost perfect compensation. The sports whose single games reliably assess talent have short seasons,

do fans of these sports have a level of reliability that they find most satisfying, but, in the aggregate, they must be exquisitely sensitive to the flow of information in sports to shape them so precisely.

If there is an optimal level of reliability, then it ought to appear not just in league sports, but in different sports with different structures. The predictability of two sports, tennis and chess, that depend on tournaments to assess skill is examined by measuring the extent to which players' rankings predict their success. While the nature of these sports and the structure of the contests are dramatically different from the league sports, the balance in the outcome between talent and chance is the same (details available from the author on request). This provides more support for the view that there is a level of predictability that is most satisfying, and that sports, regardless of the details, aim for this balance. As with natural selection, it need not be the case that sports deliberately strike this balance, but rather that those that do, succeed.

**Nicholas Christenfeld**

Department of Psychology,  
University of California, San Diego,  
La Jolla, California 92093-0109, USA  
e-mail: nicko@ucsd.edu

## Flaws in CTBT

SIR — Carrigan *et al.*<sup>1</sup> have recently shown how measurements of trace gas emissions on geological faults might be used as indicators of underground nuclear testing. Their experiment suggests that, of the  $6.7 \times 10^{15}$  becquerel (Bq) of xenon-133 produced by a one-kilotonne underground nuclear explosion, assuming that seepage through geological faults is similar to the conditions of the experiment, as much as 41 Bq m<sup>-3</sup> could be present in the atmosphere, near the test site, after a few days.

This is within the sensitivity of existing techniques. A detector with a sensitivity of 20 Bq m<sup>-3</sup> will be field tested next year<sup>2</sup>. Such a detector would be appropriate in the mode proposed in the paper, namely in a well-timed challenge inspection on site. More sensitive detectors (in the mBq m<sup>-3</sup> range) potentially exist<sup>3</sup>.

The text of the Comprehensive Test Ban Treaty (CTBT)<sup>4</sup> speaks of equipping with xenon detectors some of the 80 or so fixed sites that will constitute the system of monitoring detectors designed to ensure compliance with the treaty. In this mode, detectors with a sensitivity in Bq m<sup>-3</sup>, or even in mBq m<sup>-3</sup>, would have a very high probability of yielding a false negative result — that is, of failing to detect tests of low yield (a few kilotonnes) if they do not take place in the direct vicinity of the detector.

Xenon detectors are interesting precisely for low-yield tests, for which other signatures are weak. But to be efficient, as ref. 1 clearly demonstrates, they should not be part of a fixed monitoring system, but should instead be used in on-site challenge inspections.

The experiment reported in ref. 1 thus shows that, in its present form, the CTBT does not have a verification architecture optimized for the best use of xenon detectors and ultimately to maximize the chance of successful verification.

**Benoit Morel**

Departments of Engineering  
and Public Policy,  
Carnegie-Mellon University,  
Pittsburgh, Pennsylvania 15213, USA

1. Carrigan, C *et al.* *Nature* **383**, 528 (1996).
2. Perkins, R. & Casey, L. *Radio-xenons: Their Role in Monitoring a Comprehensive Test Ban Treaty* US DOE preprint: DOE/RL-96-51, June 1996.
3. Schell, W. *et al.* *Nucl. Instrument and Methods in Phys. Res. Sect. A* (in the press).
4. *Protocol of the Comprehensive Test Ban Treaty, Part 1: The International Monitoring System and Data Centre Functions*, Ch. C: *Radio-nuclide monitoring*, Para. 10 (UN Document A/50/1027, September 1996.).

## Einstein for kg?

SIR — The kilogram is the only SI basic unit that is a multiple of another unit, the gram. I propose the adoption of a new name for the basic unit, which should stand on its own. I suggest einstein (symbol E).

Nobody will dispute the importance of Einstein's contributions to science in general, and to the discovery of the equivalence of mass and energy in particular. I believe he should be honoured by having an SI basic unit named after him. The gram has a long tradition and deep roots in science, so it should remain as 1 g = 1 mE.

**Doo Jung Jin**

Environmental Restoration Management  
Technology Department,  
Columbia Basin College,  
Hanford, Washington 99301-3397, USA  
e-mail: djjin@ctc.edu