You cannot be serious! Public understanding of technology with special reference to "Hawk-Eye"

Harry Collins and Robert Evans

Public understanding of science, though it approaches the specialist knowledge of experts only in rare circumstances, can be enhanced more broadly in respect of the *processes* of science and technology. The public understanding of measurement errors and confidence intervals could be enhanced if "sports decision aids," such as the Hawk-Eye system, were to present their results in a different way. There is a danger that Hawk-Eye as used could inadvertently cause naïve viewers to overestimate the ability of technological devices to resolve disagreement among humans because measurement errors are not made salient. For example, virtual reconstructions can easily be taken to show "exactly what really happened." Suggestions are made for how confidence levels might be measured and represented and "health warnings" attached to reconstructions. A general principle for the use of sports decision aids is put forward. A set of open questions about Hawk-Eye is presented which, if answered, could help inform discussions of its use and accuracy.

Keywords: sports decision aids, Hawk-Eye, public understanding of technology, simulations, cricket, tennis.

1. Introduction

At the heart of the debate about public understanding of science is the relationship between esoteric knowledge and ubiquitous knowledge. The debate turns on the extent to which the ubiquitous knowledge alone is enough to make judgments that touch on esoteric knowledge. The discredited "deficit model" held that public discomfort with science and technology was caused by a lack of specialist knowledge: if only the public could share in the specialist knowledge of science then their values—in respect of such things as the desirability of new technologies—would align with those of the scientific community. This was obviously incorrect since, in addition to the difficulty of making the esoteric knowledge of specialists available to non-specialists, scientists themselves—who, by definition, understand more science than anyone—disagree about both truth and values. Social studies of science showed that science was itself shot through with ordinary thinking and decision-making: science is not an automatic procedure for generating either truth or uniform opinions—ordinary human judgment lies at its heart. For the kind of science that is likely to cause debates in the public domain, this means that ordinary people's judgment is not dissimilar to scientists' judgment in some respects.

This finding has given rise to a widespread tendency in contemporary science and technology studies (STS) to treat the public as more or less continuous with experts when it comes to technological decision-making in the public domain.[1] More recently this approach has been subject to a new challenge, concentrating on differentials, not in access to the truth, but in expertise and experience (Collins and Evans, 2002, 2007).

This paper takes the position on public understanding that has been put forward in books such as the *Golem* series (Collins and Pinch, 1993, 1998, 2005). The position is that, first, there is a deficit in public understanding of the technicalities of science and, second, there is a deficit in understanding the processes of science; while not much can be done about the former, the latter can be, to some extent, remedied. The *Golem* series was intended to improve public understanding of scientific processes.

Understanding the degree of certainty that attends any finding or claim is part of understanding the process of science. Thus, *The Golem at Large* (Collins and Pinch, 1998) pointed to the uncertainty surrounding technologies that had been presented to the public as fully understood—such as the Patriot anti-missile missile and the space shuttle.[2] Here we analyse a case of technology in the public domain which, like Patriot and the space shuttle, can give the impression that its outcomes are not subject to the errors that are always associated with measurement and prediction. Where the Hawk-Eye system differs from these cases, however, is in its impact and relative accessibility. Because Hawk-Eye and similar technologies are becoming an increasingly common part of televized sports coverage then making their capacities and technological limits more clearly visible in their day-to-day use has the potential to promote, across a wide section of society, a more nuanced and widespread understanding of the statistics of uncertainty.

To teach a wide public about the nuances of statistical decision-making under uncertainty may seem over-ambitious in the light of studies that show that, on the whole, the public have a poor grasp of statistical inference.[3] Nevertheless, there are grounds for optimism. We know that, when groups of the public have a burning interest in issues involving mathematical or statistical reasoning, high levels of understanding can be attained. Thus, Lave (1988) showed that the public are good at arithmetical calculations when it comes to supermarket transactions or weight-watching diets; gamblers with no significant education can develop a remarkable grasp of the statistics related to their passion; and some sports fans have a thorough knowledge of the statistics of their favored team's performance and how these relate to the performance of other teams.[4] The "circle is squared" because participation is the key to gaining knowledge and the public become participants, or quasi-participants, in the statistical and mathematical domains in question rather than remaining mere onlookers. We propose that this potential for a still more widespread understanding of the domain of statistical uncertainty should be actualized through the presentation of measurement errors and confidence intervals in the outputs of automated sports decision aids. As the most casual observations make evident, committed sports fans watching television act as quasi-participants in the decisions made by umpires and referees!

2. Method

The substance around which this paper turns is an analysis of technical aids designed to supplement or replace the decision-making of cricket and tennis umpires, football referees and the like. The device known as "Hawk-Eye" is used as the principal illustrative example as it is the most well-known of the commercial systems and, as we discuss later on, it is currently being used to make decisions in major tennis competitions. In addition, because it has been

used for several years it has a well developed website and has been the subject of a range of media coverage and one or two published articles. In what follows, nearly all the material we use in the analysis is drawn from such sources and so can be readily checked. Aside from some initial inquiries we were unable to obtain significant information directly from Hawk-Eye Innovations. As explained below, we discovered toward the end of the analysis that a number of our questions and proposals had already been put by contributors to newspaper websites. We cannot find any detailed response to the newspaper queries either, suggesting that our experience is not unrepresentative.

To gain as much information as we could, we used the major search engines and data-bases systematically to search popular websites and the academic literature. Specifically, three people spent a total of about 15 hours searching Google and Google Scholar (we looked at the first 20 pages that were returned) plus Web of Science (there were only three, irrele-vant, hits) for articles relating to Hawk-Eye (spelt in various different ways).[5] Of the infor-mation we uncovered, the most useful was on newspaper websites, on a discussion site called "Cricinfo" and on Hawk-Eye Innovations' own website. We also examined the original patent application, which is available online from the European Patent Office (reference number WO 01/41884). We discovered an article in *Scientific American* (Fischetti, 2007) and a recently published analysis of line calls in tennis (Mather, 2008).

We found only one article about Hawk-Eye in an engineering journal but it was very hard to access in both senses of the term:[6] it has been referred to in the main body of the paper and we attempt to explain its contents more fully in the paper's final note. We do not know how this paper, which reports development of a device under contract with Hawk-Eye Innovations, bears on the technology which is currently used but, in any case, no substantive difference would be made to the analysis if it included more reference to this paper unless it is true, as the paper implies, that Hawk-Eye takes its television feeds from existing television network cameras; this would indicate lower camera frame rates than are discussed in this paper.[7]

Reliance on public domain data means that our information about how Hawk-Eye works is not complete. We do our best to warn about the gaps in our knowledge by inserting through-out the paper the symbol, "{caveat *n*}." The "*n*" part of the symbol is a number indicating the particular pieces of missing information. These are recast as a series of questions in note 8, each cross-referenced to the paper by the number.[8] The questions in note 8 are a subset of those we would have explored had our research project involved the kind of interactions with scien-tists and technologists that have supported our many case studies of science and technology over the years. Note 8 lists, then, a series of open inquiries that others might like to pursue.

3. Hawk-Eye and measurement error

To repeat, as it is currently used we believe that Hawk-Eye could inadvertently mislead the public about the degree of certainty that can be brought to a scientific measurement. We are not the first to notice this. For example, contributors to both the *Timesonline* and *Daily Telegraph* websites in July 2007 questioned whether Hawk-Eye can measure the position of a fast moving tennis ball to the nearest millimeter as is implied by some of its reports.[9] We, however, suggest that the potential for misunderstanding Hawk-Eye's capabilities could be obviated by incorporating information about measurement error into the presentation and we develop the concerns of the contributors to the websites in a more systematic way by linking them to the concept of measurement error.

Thus, humans and machines make errors of two kinds. "Systematic errors" are repeated errors that have a similar effect each time; the causes of such errors can sometimes be understood

and their impact can be predicted and compensated for. "Random errors" cannot be predicted except that their typical size, and the shape of the random distribution, can be estimated. They cannot be compensated for but they can be taken into account in assigning a degree of confidence to a measurement. Decision aid devices are often presented as though it must be a good thing if they make decisions that are more accurate than those of humans but we show that "more accurate" can mean different things. To know what is meant by "more accurate" both random errors and systematic errors in humans must be taken into account. Such a consideration gives rise to what we call the "Automated Decision Principle," which is a proposal for how sports decision aids should be used and turns on consideration of both kinds of error.

4. What are sports decision aids?

In sport, decision aids are meant to help referees and umpires make decisions. Sometimes the action is very fast, or the official's view is obscured, and this makes it hard to make a good judgment; in these cases decision aids can help. A well-known and relatively uncontroversial example is the television replays used by referees in rugby union to determine whether a ball has been fairly touched down so as to merit the award of a "try." It could be said that a much simpler and much older aid is the string net which hangs beneath a basketball hoop; the purpose of this net is to slow and mark the descent of the ball so it is easy to see that it has passed through the hoop. Other such devices used in cricket to detect whether the ball has touched the bat include "Snicko," which detects slight sounds, and "Hot Spot" which uses heat-sensitive cameras to indicate contact; in ice hockey the puck and the line interact electronically to indicate whether a goal has been scored. Here we examine more closely an innovatory design of decision aid utilizing artificial intelligence techniques and known as "Hawk-Eye." It is regularly used in cricket and tennis though in the former case it is television viewers alone who benefit from its contribution.[10]

5. The technology of Hawk-Eye

Hawk-Eye, as we understand it is a video processing system combining a number of cameras and a computer to store and process the data. We believe that the cameras—the patent application specifies six but also acknowledges that not all cameras will produce usable data—track the flight of the ball and that these camera feeds are then used by the computer to reconstruct the trajectory of the ball by analysing the pixels in each frame of each relevant camera feed.[11] The field of play is also modeled within the system, as are some of the rules relating to the game. By combining the trajectory of the ball with the model of the pitch and the database of rules, the path of the ball can be reconstructed against the background of the main features of the playing area as a virtual reality and a decision given (e.g. should the batsman be given out in the case of cricket or, in the case of tennis, did the ball land inside or outside the line). The reconstruction can be shown to television viewers.

It is this representation rather than the internal technicalities of Hawk-Eye that are of main interest when considering the public understanding of science. The cameras will have a limited frame rate and the position of the ball between frames has to be interpolated by the software. In the patent application, this process is not described in detail but it is clear that the analysis is statistical, with predicted and actual paths being compared in order to generate the final trajectory.

The explanatory framework of this paper adopts certain simplifying conventions that, we believe, do not affect the principles of the argument. In this we essentially follow Hawk-Eye's own practice in describing its operation: its website presents the device as calculating a trajectory from a series of images of the ball captured in television frames.[12] We are, then, unconcerned with the exact statistical algorithm used in the interpolation of the path of the ball. We take it that whatever the algorithm the accuracy of the final result has an upper limit set by the accuracy of the data. We take it that the crucial data concerning the path of the ball consist of a series of "data points." In the first instance information is generated from camera frames containing certain pixels which are taken to signify the ball and other pixels which are taken to signify the line, wicket, or whatever is needed to locate the ball-pixels in the space of the playing area. When we refer to a "data point" we mean a three-dimensional reconstruction of the position of the ball which will involve information from two or more cameras. The idea of a data point serves the analysis adequately whether or not such data points are ever actually constructed as the Hawk-Eye processor runs. When we refer to "frame rate" we are talking about individual cameras. It might be that the "effective frame rate" of Hawk-Eye is higher than this because it combines frames from more than one camera to establish the trajectory of the ball in any one dimension. We simply do not know. Therefore we assume "effective frame rate" is whatever frame rate we find mentioned in our sources. Any rough calculations we make are based on these seemingly plausible assumptions but they are open to correction {caveat 1}.

In our language, then, each effective frame provides a single "data point" and we would expect the reconstructions to be more accurate with higher frame rates as this would minimize the distance between data points. In addition, having more data points would allow more complex curves to be fitted. Thus, two data points could provide information that can be used to infer the straight-line direction and velocity of the ball (subject to errors). For trajectories that depart from a straight line, more data points would be needed {caveat 2}.[13]

In some circumstances, Hawk-Eye projects a hypothetical trajectory beyond its last data point. One such case is the "leg-before-wicket" (lbw—see below) decision in cricket in which the continued trajectory of the ball beyond the last data point is projected forward and then represented on a virtual reality display that shows the television viewer what "would" have happened had the ball not hit the batsman. We start with the lbw situation because the technicalities (if not the game) are relatively easy to understand and it was the lbw decision that gave rise to our initial questions. The second part of the technical analysis concerns the tennis line-calling system.

6. The lbw rule and the Hawk-Eye system

Cricket is a complicated game. We must explain some of the rules here on the assumption that not every reader of this journal will know them. We will, however, assume that readers who are not familiar with cricket will know baseball.

Cricket is like baseball in that it involves the equivalent of a pitcher and a batter, known, respectively, as the "bowler" and the "batsman."[14] The bowler "bowls" the cricket ball to the batsman and, as in baseball, the batsman tries to hit it. Unlike baseball, there is no limit to the number of balls the batsman may receive—on a good day a batsman may face hundreds of balls before being out. In international matches, one game may continue for up to five days.

As in baseball, there are a number of ways of being out, such as when one of the "fielders" catches the ball before it hits the ground. In cricket the batsman stands in front of a "wicket" (otherwise know as "the stumps") that he has to defend with his bat. If the ball hits the wicket, the batsman is out—there is no equivalent in baseball. The wicket is a set of three

vertical sticks or "stumps." The wicket is 28 inches high and 9 inches wide overall. The top of each stump has a shallow groove cut at right angles to the direction from which the ball is coming; two smaller sticks, known as "bails," are carefully balanced in these grooves, the ends of the two bails touching each other where they meet in the middle of the groove cut in the central stump (see Figure 2, below). The working, and universally accepted, definition of "the ball hitting the wicket" is that one or both of the delicately balanced bails fall to the ground—the wicket must be "broken." On very rare occasions a ball grazes the stumps, or rolls very gently against them, but no bail falls; in such a case the batsman is not out because the wicket has not been broken.

In cricket, the bowler nearly always directs the ball in such a way that it hits the ground before it reaches the batsman and it usually then bounces toward the batsman's legs. The batsman wears a "pad" to protect each leg. Each pad is an armored sheath running from ankle to just above the knee. The ball is very hard, about as hard as wood at the beginning of the game, though it begins to soften slightly as the hours pass (the same ball is used for many hours before it is changed). The ball can sometimes be bowled at more than 90 mph (miles per hour). Allowing the ball to hit the pads is an integral part of the game. Clearly, the batsman would never be out if he simply stood in front of the wicket, kept his bat out of the way, and allowed the ball to hit him or his pads.[15] To make that impossible the notoriously complicated "lbw rule" says that a batsman is out in certain restricted circumstances if the pads alone stop a ball that would otherwise hit the wicket—this counts as out in virtue of "leg before wicket." In the normal way, the umpire, who stands at the point from which the bowler bowls the ball, is the sole judge of whether the ball a) falls within the restrictions and b) would have gone on to hit the wicket. The question that concerns us here is the decision about whether the ball would have gone on to break the wicket if it had not hit the batsman's pads.[16]

One of the earliest uses of Hawk-Eye was to project the path of the ball after it hit the batsman's pad in an attempt to judge whether it would have gone on to hit the wicket. Figure 1 is a two-dimensional schematic version of the situation from side-on. The ball, traveling from left to right, bounces and then hits the batsman's pad-protected leg. The dotted portion of the trajectory is what has to be judged or estimated. Television viewers see a three-dimensional virtual reality representation of the projected path of the ball against a virtual cricket field and they can see it either hitting or missing the wicket. For a number of years after the introduction of Hawk-Eye, cricket commentators would simply remark on what Hawk-Eye showed on the screen, giving the impression, perhaps inadvertently, that the virtual reality represented exactly what would actually have happened had the pad not been struck. This is where our analysis of Hawk-Eye begins.

A cricket ball is not uniformly spherical. Around its "equator" it has a raised seam and the two "hemispheres" become more asymmetrical as the game goes on. The trajectory of the ball after it hits the ground can vary enormously. The bounce depends on the speed, the hardness and texture of the ball—which changes during the game, the state of the ground at the exact point of the bounce, the spin on the ball and the position of the seam. The "swing"—which is the aerodynamically induced curve in the flight of the ball, which can be in any plane—depends on the ball's speed, its spin, its state, its orientation, the orientation of the seam and the state of the atmosphere. As a result, what happens to the ball after it bounces is not going to be fully predictable from its pre-bounce trajectory so that, as far as we can see, Hawk-Eye has to estimate the post-bounce trajectory largely or entirely from the post-bounce behavior of the ball {caveat 3} for which it can gather data between the bounce and impact on the pad. This certainly seems to be the implication of the claim made by Paul Hawkins, the Director of Hawk-Eye Innovations, in response to a criticism of Dennis Lillee, the Australian fast bowler:
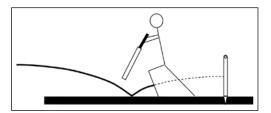
Figure 1. A two-dimensional schematic of a potential lbw situation.

"… Hawk-Eye simply observes and then calculates the actual trajectory of the ball. Whether the cause of this trajectory was due to atmospheric conditions, the wicket, or the ball hitting the seam is irrelevant from a Hawk-Eye perspective. Hawk-Eye just tracks what happened—it does not try to predict nor to answer why it happened."

So, if the ball rears up unexpectedly after hitting the seam or a crack on the pitch, Hawk-Eye will track the trajectory off the pitch to predict the future course of the ball. Similarly, the tracking system will come into play if the ball shoots along the ground after hitting a dry spot on the pitch. (Hawkins quoted by S. Rajesh, "Give Hawk-Eye a Chance," on Cricinfo website, 18 December 2003)[17]

Our concern in analysing what Hawk-Eye can do is to understand more fully what it means to "track" and "predict" the path of the ball. Predictions are extrapolations and the accuracy of these extrapolations is limited by, among other things, the quality of the data. No measurement is ever exact. Heisenberg established this as a deep principle of physics with the "uncertainty principle," but here we are talking of macroscopic measuring processes such as are discussed by, say Thomas Kuhn in his 1961 paper on measurement and, of course, by physicists and most other scientists as a matter of ordinary fact in their day-to-day work. As a result, it is normal in science to associate a measurement with an estimate of its potential error.

A decision is not a measurement. A decision is binary like the "guilty/not guilty" decision of an English jury; in cricket the batsman is either "OUT" or "NOT OUT." The process of what we will call "digitization" is used to turn inexact measurements into discrete decisions. In most sports, the referees or umpires are the people who do the digitization and what we are discussing here could be described as technical aids to digitization. The bails in cricket are one such aid to digitization. As discussed above, it can sometimes be difficult to tell whether the ball has touched the wicket or not and the falling of a bail converts this uncertainty into one of two discrete possibilities which have merely to be "read off" by the umpire.

In the case of lbw decisions, the bails can't help because the flight of the ball has been stopped before it gets to the wicket and so the wicket is never broken. Instead, the umpire has to judge whether the bails would have been dislodged had the ball continued on its trajectory and passed "through" the batsman instead of being stopped by his legs. Here Hawk-Eye's reconstruction and extrapolation appears to have the potential to take on the role of the bails by showing what would have happened.

In practice, however, Hawk-Eye's digitization cannot correspond exactly with that of the bails in all cases because there are some circumstances in which Hawk-Eye will not be able to predict with certainty whether the ball really would break the wicket if it had not hit the batsman's pad. To take the extreme circumstance where the ball just touches the wicket, whether the bails would fall would depend on how firmly they are sitting in their grooves, how rigidly the stumps are held in the ground (the tapered part at the base of each stump is pushed into the turf to hold them upright), or whether the bails and stumps are wet or dry.

Again, the ball is not exactly spherical so whether it would cause a bail to fail can depend on its orientation as it passes the wicket. We assume that predicting all these things is beyond the capacity of both Hawk-Eye {caveat 4} and human umpires, but all we are doing in this paragraph is establishing *the principle* of the imperfection all such measurements.

The usual way of coping with measurement error in experimental science is to report a confidence interval of the errors. The width of the confidence interval is a function of two things: the confidence level which is chosen by the experimenter and the dispersion of the distribution of errors. If the dispersion of errors is known, then each prediction can be associated with a confidence interval defined by a chosen level of confidence such as 95 percent or 99 percent: the first would mean that it is estimated that there is only a 5 percent chance that the error is greater than the outer limits indicated by the 95 percent confidence interval; the second would mean that there is only a 1 percent chance that the error is greater than the (wider) limits indicated by the 99 percent confidence interval.[18]

To anticipate, we have not found any detailed indications of the dispersions of Hawk-Eye's errors in the public domain. The nearest we could find are the following quotations from Paul Hawkins when interviewed by S. Rajesh on the "Cricinfo" website:

> … Hawk-Eye has shown that balls pitched on roughly the same area on the wicket have passed the stumps at widely varying heights. And in tests conducted, thousands of deliveries were bowled from a bowling machine and filmed by Hawk-Eye. The camera feeds were cut about two metres from the stumps, approximately the point where the batsman would normally intercept the ball. When the ball hit the wicket, Hawk-Eye was able to determine, to within about 5 mm, the point of impact.

> … "Hawk-Eye requires between 1 to 2 feet of travel after the ball has pitched to be able to accurately track the ball out of the bounce (this is significantly less than an umpire requires). In instances when this does not happen, a Hawk-Eye replay is not offered to TV." (18 December 2003)[19]

> … in most cases Hawk-Eye's output is accurate to within five millimetres in predicting the path of the ball. The accuracy levels are highest when the ball has traveled a fair distance after pitching, but even when the point of contact is very close to the pitch of the ball, the accuracy levels are still within 20mm. (13 June 2006)[20]

We don't know if the developers of Hawk-Eye have attempted to estimate the dispersion of their errors in a way that is more exact than is indicated in the above quotations but we have not found any such report {caveat 5}.

As the quotations intimate, the size of the error will be affected by the length of the bounce-to-pad trajectory but it will also depend on other factors. The bounce-to-pad trajectory is not of fixed length because the position of the bounce is variable and the batsman can move forward or back. The length of the projected pad-to-wicket trajectory also depends on where the batsman is standing when struck. The accuracy of Hawk-Eye's estimates are likely to have a direct dependence on the distance between the bounce and the moment of impact on the pad (the longer the better) and an inverse dependence on the distance between the impact on the pad and the wicket (the shorter the better) {caveat 6}.

In the case of the human umpire making an lbw decision it is acknowledged that the accuracy of the judgment is affected by how close the batsman is to the wicket when the pads are struck by the ball. If the batsman whose pads are struck is well forward in his stance then he is rarely given out. In this way, *human judges deliberately introduce a systematic error*

into their judgments that favors the batsman—the so-called "benefit of the doubt" rule. The importance of this rule will become clear later.

None of the information that we have found in the public domain gives any indication of overall dispersion of Hawk-Eye's errors and the only indication about how the error increases with short bounce-to-pad trajectory that we can find is given in the quotation above. We can find no systematic information about how the size of the error relates to position at which the ball bounced, the speed of the ball, the length of pad-to-wicket trajectory, the length of bounce-to-pad trajectory, the degree of spin, the degree of swing, the nature of the pitch surface, or the nature of the atmosphere. Although the patent application does acknowledge that the distance of the camera to the ball is important (e.g. frame sizes are set to make the ball appear as large as possible) and that the position of the sun matters (e.g. it is important to distinguish the ball from its shadow) it does not tell us how variations in these or other parameters affect the accuracy with which the position of the ball can be tracked.

Unfortunately, the importance of accounting for these errors is greatest when the decision is most difficult. Whenever the projected point of impact is well away from the edge of the wicket there is unlikely to be any real doubt about the "correct" decision and so Hawk-Eye's errors are probably "hardly worth reporting."[21] Where the judgment is much more difficult—for both human and machine—then the potential error becomes crucial. To understand the issues it is necessary, first, to acknowledge that there is a distribution of measurement error and, second, to describe the characteristics of this distribution (its mean and dispersion). At worst these should be described for the general case but it would be much better to provide separate analyses for each of the main conditions that can affect the accuracy of the prediction. For example, it would be nice to know the dispersion of errors associated with fast balls and slow balls, with different lengths and ratios of the various trajectories and, perhaps, with different conditions of the pitch, the atmosphere, and the ball— e.g. is Hawk-Eye likely to make bigger errors in conditions when the ball tends to swerve as it travels through the air after hitting the wicket?

The frame rate of the cameras will affect the accuracy of the prediction. Though the technical article referred to above suggests that Hawk-Eye takes its feed from standard broadcast cameras (which would have a frame rate of around 30 per second) we will assume that the frame rate is 120 frames per second, as reported in 2004 on the Cricinfo website {caveat 2}.[22] In this case, if the ball is traveling at 80 mph, it would travel one foot between frames. This would make sense of Hawk-Eye's claim to need between one and two feet to make a prediction. In the worst case scenario, a ball traveling at 120 feet per second with a frame rate of 120 per second, would need a minimum of two feet to provide three data points which should be enough to calculate some kinds of curved trajectory (subject to errors) {caveat 7}. Fortunately this lack of sure knowledge of frame rate and method of calculation does not affect the general principle of the argument {see caveats}.

How error could be measured in the case of lbw

We do not know if data from the tests involving "thousands" of balls (see above) have been preserved {caveat 8}. If they have they might give an initial indication of the distribution of errors. The beginnings of a more complete analysis could be made if, in a test like this, the cut-off point for the camera feed was systematically varied and the bounce point of the ball was systematically varied so that a more complete range of potential errors was analysed. In a still better test these parameters, and other parameters that affect the post-bounce behavior of the ball, would be recorded and measured so that the likely size of error could be reported
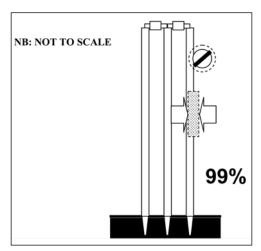
Figure 2. Some possible ways of indicating Hawk-Eye's possible measurement errors.[24]

for different circumstances. Some of this may be beyond the ability of Hawk-Eye to measure but if it is then it could be clearly stated {caveat 8}.

Our own preference would be as follows. The error should be estimated from empirical tests either in the way suggested above or in some other way. Reports on the method of testing and its degree of completeness should be made readily available. Subsequently, using whatever knowledge of dispersion of measurement error was available, confidence levels would be associated with Hawk-Eye reconstructions in real time {caveat 9}. The graphic shown to television viewers could be adjusted either to show something like an "error bar," or "error circle," around the projected position of the ball, or to indicate it in some other way such as numerical confidence level for the prediction that the wicket would have been struck/not struck. Figure 2 (the putative errors are not drawn to scale) indicates some of the possibilities though no doubt it could be improved upon.[23]

If such changes were implemented, commentators might remark, "Hawk-Eye was 99.9% sure the ball was going to hit the wicket so the umpire was right," or "Hawk-Eye was only 90% sure the ball was going to hit the wicket—the umpire should not have given it out," or some such. This way, not only would Hawk-Eye's abilities be presented in a clearer and less easily misunderstood way, but the technology itself could fulfil a valuable role in educating the public about the way uncertainties are turned into decisions. What we are suggesting is not much more than is mentioned in the Hawk-Eye patent that we have found. There it is claimed that the HIT–MISS decision made by the apparatus is based on: "whether the probability of the ball going on to hit the stumps is high e.g. above a given probability threshold" (p. 8). We are asking for a more complete explanation of what the threshold is, what the probability is on any occasion, and for this information to be offered to the public.

7. Tennis line calls

In tennis, unlike cricket where it is used only to enhance television coverage, Hawk-Eye is now being used to take decisions. In high level tennis tournaments players can make "challenges" to the umpire's decisions and, if the player's challenge is supported by Hawk-Eye, the

original decision is overturned. In the summer of 2007 Hawk-Eye figured in at least three disputed line calls in which the ball was called OUT but, after a challenge from one of the players, was subsequently called IN by Hawk-Eye. Two of these are worth reporting in detail as they bear on the argument that follows:

Disputed line call A—Dubai

As reported in the Gulfnews website (3 March 2007), in a match in Dubai between Nadal and Youzhny, a challenge made by Youzhny was supported by Hawk-Eye:

> World No 2 Rafael Nadal has questioned the efficiency of the new Hawk Eye line calling technology.
>
> Thursday's first set between Nadal and Youzhny ended in a controversy with the tie-break score at 6–5 in favour of the Russian. Nadal thought a ball from the Russian had landed wide.
>
> So the Hawk Eye was pressed into service and it showed the ball had skimmed the line.
>
> But Nadal, chair umpire Roland Herfel of Germany and even Youzhny believed that the ball had landed wide after watching the Hawk Eye. But officials are bound to accept the Hawk Eye ruling.
>
> "The mark of the ball was still on court and it was outside. But in the challenge it was in, so that's unbelievable. The Hawk Eye system is not perfect," fumed Nadal.
>
> "I told the chair umpire: 'Look, the ball is out' and he said: 'I know'.
>
> …
>
> Even Youzhny agreed the ball appeared to have gone out. "I saw the mark, but I just challenged because it was a very important point," the Russian said.[25]

Disputed line call B—Wimbledon

The technology was also central to a disputed line call in the Wimbledon men's final between Federer and Nadal. Nadal hit a ball which appeared to television viewers, to the umpire, and to Federer as impacting well behind the baseline, but Hawk-Eye called it IN. Federer appealed to the umpire but the umpire accepted the Hawk-Eye judgment. The following is the account from the *Daily Telegraph* newspaper website. The story is dated 10 July 2007.

> Federer, a tennis conservative, has always been against the introduction of Hawk-Eye, and he was as angry as he had ever been on Centre Court when an "out" call on one of Nadal's shots was successfully challenged by the Spaniard in the fourth set. The Hawk-Eye replay suggested that the ball had hit the baseline; Federer thought otherwise. It was then that Federer asked umpire Carlos Ramos whether the machine could be turned off. Ramos declined but also seemed to suggest that he had thought the ball had landed long.
>
> The Hawk-Eye review gave Nadal a break point, which he converted for a 3–0 lead, and Federer continued to complain during the change of ends. "How in the world was that ball in? S***. Look at the score now. It's killing me, Hawk-Eye is killing me," the Swiss said. So, a system which was introduced to prevent McEnroe-style rants at officialdom actually left one of the sport's gentlest champions fuming.[26]

Technology and tennis

In some ways, the technical aspect of the tennis case is easier than the lbw case because Hawk-Eye has more data points to go on: it can follow the ball's trajectory right up to the point of impact and, sometimes, beyond. (Though we do not know if it uses post-bounce data points in its calculations {caveat 10}.) Given that no combination of cameras provides infinite frame rate the computers still have to project forward (and back?) to generate the virtual trajectory. Again, we don't know frame rates so we don't know how much projection has to be done. Tennis balls can be served at up to 150 mph (*c*.220 feet per second) so in this respect the problem is worse for tennis than for cricket (we don't know how fast balls travel in the course of a rally).

In the case of tennis, there exists no traditional physical method for digitizing line calls which is intrinsic to the game as with the bails in cricket. Decisions are traditionally made on the basis of fallible human observation—which is to say that the digitization is normally done by human beings. It seems quite likely that Hawk-Eye could do better than a human umpire in most circumstances but, once more, this is not the same as saying it is always correct.[27] Again, an argument from first principles suggests that it is bound to make occasional mistakes.

Hawk-Eye reports on its own website that the mean error in the position of the tennis ball as measured by its system is 3.6 mm. Again, what it does not report is the distribution and dispersion of errors or the conditions under which errors are greater or smaller. Here, if we take the 3.6 mm as the "mean deviation" of the errors, we can do some simple "as if" calculations {see caveats}. These calculations do not necessarily bear on the actual performance of Hawk-Eye because we have too little information but they indicate the kind of thinking and calculating that might be done. These calculations assume that the distribution of Hawk-Eye's errors is the normal distribution and that there is no systematic error in Hawk-Eye's measurements as systematic error is normally understood (but see Section 8 below).

Dispersions of errors are usefully reported in terms of the "standard deviation." If the distribution of the errors was the well-known and frequently encountered "normal distribution" {caveat 5}, then if the mean deviation is 3.6 mm the standard deviation would be about 3.6 mm × 1.25 = 4.5 mm.[28] Because, in a normal distribution, 95 percent of the points lie within approximately 2 standard deviations of the mean and 99 percent lie within about 2.6 standard deviations, we can estimate some putative confidence intervals. In this case we could say that in 5 percent of Hawk-Eye's predictions (that is 1 in 20), the error could be greater than about 9 mm and in 1 percent it could be greater than 11.7 mm. The physics of the situation means that there could be an absolute upper cut-off point for the errors and this could be smaller than the calculation from an assumed normal distribution would imply, but we have no firm information as to whether this is the case. Even if the numbers we have calculated are correct this would not mean that Hawk-Eye's call would be wrong every time it makes a significant mistake. This is because rightness and wrongness in terms of the binary decision (IN or OUT) depends on the direction of the error. Nevertheless, if the figures were correct it would be likely to be wrong on some of those occasions and the incidents described above could have been such occasions. According to Hawk-Eye Innovations' own website, in the case of the Federer–Nadal call, Hawk-Eye called the ball IN by only 1 mm: the possibility for mistakes is clear even if we look no further than the 3.6 mm mean deviation {see caveats}.

But what does the mean error of 3.6 mm in tennis imply? Is this the mean measured for all shots including, say, lobs and low fast drives or serves {caveat 11}? Just as in the case of varying kinds of ball in cricket, it seems likely that the error will not be equally dispersed in tennis for different kinds of shot. For example, it is likely to be greatest in the direction of
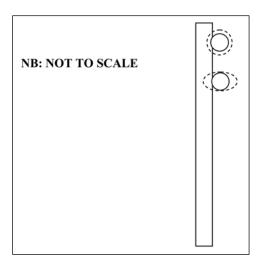
Figure 3. Is the error concentrated in the direction of travel of the ball and to what extent?

travel of a fast moving ball. In this case velocity across the line of travel is zero, or almost zero, but in the direction of travel small errors in measurement will make a big difference to the position of the ball. In Figure 3 the back line of a court is shown with a ball clipping the back edge. Uncertainty is indicated by a dotted circle surrounding the topmost ball. The diagram is only very roughly to scale at best but the circle is meant to show the error associated with between 2 and 3 standard deviations assuming a normal distribution and the other assumptions made above. In other words, on these assumptions, between one time in 20 and one time in 100 the actual position of the ball will be nestled up somewhere against the inside of the dotted circle. The lower ball shows roughly what the error would look like if it was concentrated into the direction of flight of the ball. Again, scale is not accurate and the degree of elongation of the oval might be exaggerated but we cannot tell in the absence of more information. If the 3.6 mm mean is in fact averaged over all kinds of shot it could be that in the case of fast drives or serves the elongation should be even more exaggerated. Here we are not in a position to make a positive claim about these things, merely indicate possibilities that have not been discussed in the public domain as far as we can see.

In an initial e-mail exchange with Hawk-Eye Innovations' Tennis Operations Manager, we were referred to the International Tennis Federation if we wished to understand the methods of testing Hawk-Eye's errors. The International Tennis Federation (ITF) provides details of its testing procedures for automated line-callers on a website. We understand the ITF has the true position of the ball measured with very high-speed cameras. The crucial passages read as follows:

A4.5 Accuracy and Reliability

The decision-making success rate (i.e. "in" or "out" decisions) for all balls bouncing between 100 mm inside the line and 50 mm outside the line should be 100% with a tolerance of ± 5 mm.

The average absolute discrepancy for all impacts on a single line on court should be no more than 5 mm.

The maximum discrepancy between the system's measure of the distance from the line and the true distance should be 10 mm for all impacts.

> The system should be capable of making the correct in/out decision if a ball legally crosses a line from outside to inside.

We found these rules difficult to understand. Initially, we could not understand how IN/OUT decision-making can be 100 percent accurate if there is a tolerance of 5 mm. On the face of it, these statements seem incompatible—a ball could be 5 mm OUT and still be called IN. We thought that even if we forget about distribution of errors and just accept the 5 mm at face value, if Hawk-Eye was taking its measure of accuracy from the ITF {caveat 12} the Federer–Nadal disputed ball might well have been OUT by nearly a quarter of an inch, even though Hawk-Eye called it IN. If we accept Hawk-Eye Innovations' own figure of 3.6 mm average error and its claim that the ball was 1 mm IN, the possibility for a mistake is still obvious. The ITF appeared to agree and in response to our inquiries (all of which took place on 26 January 2008), their spokesman said:

> … in general, if the ball landed sufficiently close to the edge of the line, there is a chance that Hawk-Eye could make the wrong call.

Hawk-Eye Innovations' own website contains a discussion of this specific line call. The introductory paragraph remarks:

> This document provides more information about the line call that Roger Federer questioned during the Wimbledon Men's Singles Final on Sunday 8th July. Whilst it is unable to prove conclusively that the ball was 1mm IN as shown by Hawk-Eye, it can show that 1mm IN is a likely [*sic*].[29]

The ITF was also able to clear up at least some of our confusion in a speedy way. Here is the gist of the initial response from the ITF.

> All decisions made by a line-calling system ("in" or "out") must be correct, unless the ball lands within 5 mm of the outside edge of the line, when an incorrect decision is allowed, providing that the absolute error in the system's measured impact location is no more than 10 mm.
>
> Example 1.
>
> True impact location: 4 mm "out".
>
> System's measured impact location: 2 mm "in".
>
> Outcome: Acceptable (wrong decision, but absolute discrepancy < 10 mm).
>
> Example 2.
>
> True impact location: 4 mm "out".
>
> System's measured impact location: 8 mm "in".
>
> Outcome: Unacceptable (wrong decision, absolute discrepancy > 10 mm).

In sum, the ITF accepts errors of up to 10 mm for individual impacts, and the system may still pass the accuracy test overall (contingent on meeting the other performance criteria).

Incidentally, we asked the ITF how many impacts were involved in their tests. They explained:

> Over the full evaluation, at least 80, and normally 100–120. Of these, around 10% land within 5 mm of the line.
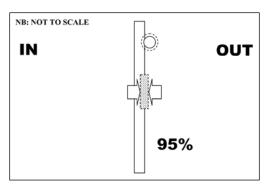
Figure 4. Possible ways of indicating possible error in tennis line calls.

Thus, it could be that the ITF tested Hawk-Eye's performance in the crucial zone around the edge of a line on only around 6 to 15 impacts of ball with court.

On random error in tennis, it seems to us that the contribution of Hawk-Eye would be much better understood if, just as in cricket, it were admitted that on a few occasions it will be wrong and that each prediction were associated with a confidence interval. Each line call provided by Hawk-Eye in real time should be associated with a claim about the confidence. Figure 4 suggests ways in which these possibilities could be indicated to the public (though the error circle might need to be elongated as in Figure 3).

8. Decision-maker or decision aid?

Whilst random error can be dealt with by estimating confidence intervals, the case of systematic error is more complex. Whilst the common-sense idea might be to off-set it in such a way as to return the mean error to zero, applying this principle in the case of tennis raises more subtle questions about the role of human and machine judges.

The Hawk-Eye Innovations' website makes the claim, supported by stills from a high-speed camera, that the human eye, and television replays, can be systematically misleading under certain circumstances.[30] The ball may just touch the line but skid so that it is still in contact with the ground when it bounces upward well beyond the line, giving the impression that it did not, in fact, touch the line at all. The website seems to show, then, that it is possible for the ball to appear OUT on a television replay or to the naked eye but still be just IN. Barring unknown sources of human or machine "malfunction," the disagreement between humans and Hawk-Eye in the case of the Wimbledon dispute has, then, two possible explanations. The first is the one given in the preceding sections, namely that it was a random measurement error in Hawk-Eye {see caveats}. The second, suggested in Hawk-Eye Innovations' own analysis, is that the disagreement results from a systematic error in human judgment.

A different way of looking at the problem comes from the sociology of technology.[31] It might be that Hawk-Eye could become *defined* as the "decision-maker" rather than the "decision aid" such that questions concerning its accuracy would no longer be relevant. This is already beginning to happen in tennis, though at the moment it occurs only when a player makes "a challenge" and Hawk-Eye's reconstruction fulfills the same definitive role as the bails in cricket. Making Hawk-Eye the authority in this way could resolve a number of problems by providing readily acceptable explanations for otherwise borderline decisions. Thus, in cricket, it is not unknown for the ball to brush against the wicket but fail to dislodge the bails.

No one claims the bails are "inaccurate" because everyone accepts that bail displacement is the digital definition of hitting the wicket. Similarly, if Hawk-Eye's decisions were to be made the defining criterion of lbw, IN or OUT in tennis, and so forth, players would come to talk in terms of "bad luck" if a call went against their own judgment rather than "inaccuracy," just as they now talk of bad luck rather than inaccuracy if a bail does not fall when the ball gently touches a stump. The question is: how would this change the game?

In cricket there could be substantial changes. For instance, it has been argued that if Hawk-Eye's "face value" lbw projections were taken as the defining criterion, many cricket games in which it was used would be much shorter, perhaps leading to a financial crisis. As discussed above, in lbw decisions made by umpires, a systematic error is deliberately introduced via the well-established rule that the batsman gets the benefit of the doubt. Since there is a lot of doubt in human lbw decisions, it is often quite hard to get a decision made in the bowler's favor. It has been suggested that Hawk-Eye's projections, if taken literally, would greatly increase the number of lbw decisions unless batsmen started to play differently {caveat 13}.[32]

In tennis similar considerations apply though they are of a more subtle nature. Let us assume that the Federer–Nadal ball was actually 1 mm IN as Hawk-Eye called it. That is, let us suppose there was no significant random error in this case but that a fast traveling ball had distorted and skidded on hitting the surface such that though there was a small fraction of a second when the trailing portion of the skin of the ball was in contact with the line, no human eye would spot it. Let us suppose, as Hawk-Eye Innovations' website argues, that in most such cases humans are likely to call the ball OUT. It follows that in nearly all pre-Hawk-Eye matches such a case would have been called OUT—as the umpire did in fact first call it. Assuming there is such a systematic bias, the introduction of Hawk-Eye as the authority would change the game of tennis meaning, among other things, that umpires and players would be less confident in stopping a rally when the ball appeared OUT to the naked eye. This raises the question: do we (as supporters, viewers or players) want the games to change in these ways?

In passing, it should be noted that the question of whether Hawk-Eye would bring about such a change {see caveats} is now being obscured. Those who watched the Australian Open tennis championship in the early part of 2008 will have noticed that where Hawk-Eye was used to make a decision a television replay was not offered—it was Hawk-Eye or nothing. The viewer had nothing to go on as regards any question of whether Hawk-Eye was right or not. This seems an unnecessary restriction and allowing viewers to see both the normal replay and the Hawk-Eye reconstruction would allow for a more informed debate about which kind of decision-maker is preferable. As it is, this debate has been closed off because only one set of evidence is available.

How should we use sports decision aids?

Understanding random and systematic errors is, then, intimately related to changes in sport. In cricket we have seen the possibility that if Hawk-Eye's decisions were taken at face value and used to replace the umpire in the matter of lbw, it might cause a financial crisis. In tennis the potential changes concern players and umpires as they make decisions about whether to stop a rally if the ball appears OUT to the naked eye. In both cricket and tennis, and presumably other sports not yet considered, it would mean that rules would be applied differently in the top echelons of the game, where the devices were in use, as compared to the lower echelons; top level games would become still more different from lower level games than they are now. Finally, the game as seen by the viewer would become different to the game as seen by the technology.
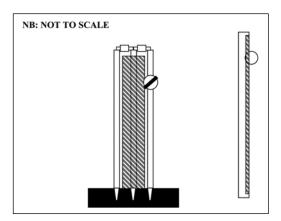
Figure 5. Possible ways to reproduce human systematic error.

The previous paragraph invites a resolution in the form of three rules of thumb:

Don't make the cure worse than the disease.
Minimize the gulf between the technically assisted game and the non-technically assisted game.
Don't disappoint the spectator.

Applying these principles suggests that technological decision aids should be adjusted to make the *same* systematic errors as are made by existing human judges so the game changes as little as possible.

When it comes to random errors things are different. No one likes to see games decided by bizarre decisions made by harassed umpires and referees but the spectacle is becoming more and more common thanks to television replays. There is little doubt that devices such as Hawk-Eye are more *reliable* than human decision-makers where "reliable" means they will make the same decision again and again in the same circumstances and they are not likely to make bizarrely wrong decisions as a result of lapses of attention. Even forgetting about the bizarre, devices like Hawk-Eye are almost certain to make smaller random errors than referees and umpires in circumstances where the ball is not close to a critical edge and as computer processing capacity and speeds increase they should improve further. (A possible counter example would be the ball that bounces on or just in front of the batsman's foot in cricket in which case Hawk-Eye cannot gather any data but if the batsman is well back a human umpire may be able to make a sound "OUT" decision {caveat 14}.)

Combining these concerns leads to an overarching rule covering both random and systematic error which we will call the "Automated Decision Principle":

THE AUTOMATED DECISION PRINCIPLE

Use automated sports decision aids to reproduce human systematic error while minimizing random error; explain what is done and assign confidence levels to automated judgments.

This approach maintains human systematic errors as an integral part of the game. In contrast, as things stand, the role of sports decision devices is implicitly defined as to make more *accurate* decisions than umpires and referees—which implies correcting both random and human systematic errors.

The point is made clear in the case of the Federer–Nadal challenge. Under the Automated Decision Principle, even if we neglect random error and accept that the ball really was one millimeter IN, it would still have been called OUT because all humans would see it as OUT. Figure 5 summarizes the way human systematic error could be reproduced by Hawk-Eye-like devices.

On the left of Figure 5 is shown a wicket. To reproduce the "benefit of the doubt" rule it would only be necessary for Hawk-Eye to make its decisions on a smaller virtual set of stumps indicated by the shaded box. Balls predicted to hit the wicket in the area outside that box would count as "NOT OUT" on the basis of benefit of the doubt. The same result could be produced by giving a high cut-off point—such as 99.9 percent to the confidence demanded before an "OUT" decision was made.[33]

On the right of Figure 5 is shown a tennis line at the rear of the court with a shaded area. If it was predicted that the skin of the ball touched this shaded area, the call would still be "OUT." Unfortunately, in tennis the matter is more complicated because the size of the shaded area should depend on the speed of the ball. For example, in the case of a lob the shaded area would be smaller if our analysis is correct {see caveats}.

The actual decision about how to apply these rules would have to be made by the various sports' governing bodies. How large should the virtual wicket be, or what would be the correct cut-off point for confidence level in the case of lbw, if the human benefit of the doubt rule were to be reproduced? To do it properly some observation and analysis of existing human practices would be a good idea. In tennis, high-speed cameras could be used to measure human judges' propensities to call OUT even when balls moving at various speeds just touch the line. This information would help to establish an appropriate rule for Hawk-Eye-like devices. This is how systematic error could be handled.

In the case of random error the matter is more complicated. In the case where a ball impacted well away from the in–out edge, decision-making devices could, essentially, replace human line-callers. They would improve upon human line-callers because they could obviate cases where the human caller was momentarily unsighted or when a bizarre call is made for some reason.

Where the ball was close to the in–out edge something more sophisticated would be needed even after systematic error had been discounted. Imagine that random measurement error could be reliably estimated and consider a line call where the confidence was less than, say, 99 percent. The tennis authorities could adopt a "benefit of the doubt rule"—always in or always out. Or they could have the call decided at random—as effectively happens now but without acknowledgement or systematic understanding of the bias in the errors. Or they could ask for the point to be replayed. (Most neutral viewers of the Federer–Nadal match would probably have been happier to see that point replayed rather than called IN.) Exactly what the cut-off point should be, and what the rule should be, or whether there should be a series of rules for different cut-off points, ball speeds, and surfaces, is not something that can be properly thought through without knowing the dispersion of errors in the technological assistant. It might even be that the umpire should retain the final say in these cases, using the reported confidence associated with each automated call as an aid.

To proceed in this way, at the very least the information indicated by the questions listed in note 8 would need to be provided or technical explanations given for its irrelevance. Or the doubts should be resolved by estimating error in some other demonstrable and accountable way and reporting the levels of error clearly.

9. Artificial intelligence, micro-worlds and virtual realities

The difference between Hawk-Eye and human judgment can be understood in a more generalized way that pertains to the entire enterprise known as "artificial intelligence." This is

the difference between the real world and what has been called a "micro-world."[34] Hawk-Eye called the Federer–Nadal ball IN by *one millimeter*. Such a call could be made only in a "micro-world"—the world of Hawk-Eye's virtual reality. In real life, the edge of a line painted on grass cannot be defined to an accuracy of one millimeter. First because grass and paint are not like that, and second, because, even given perfect paint and a perfect surface to draw on, the apparatus used to paint the line is unlikely to maintain its accuracy to one millimeter over the width of the court. Furthermore, tennis balls are furry and it is not clear that their edges can be defined to an accuracy of one millimeter. In short, in the real world of tennis we do not quite know what "touching the line" means. In the real world of tennis it is also possible that a ball that touches the perfectly defined virtual line in the supposedly equivalent micro-world {caveat 15} might not touch the fuzzy edged and not-exactly straight real line actually painted on the court. In short, at Wimbledon there is no such thing as "in by one millimeter."

A frequently encountered mistake in artificial intelligence is to take micro-worlds to stand for, or even to be superior to, real worlds and to take possibilities that could pertain in a micro-world (stacking of blocks by an automatic crane, exact measurement, exact machine translation, exact speech-transcription, and so forth) to pertain in the real world. To some extent this may be happening here. The micro-world ethos would certainly encourage the claim that where Hawk-Eye's decisions differ systematically from those of humans, it is Hawk-Eye that should be taken as the authority because it is the "more accurate." In sum, uncritical acceptance of the artificial intelligence approach directs the use of sports decision devices away from the Automated Decision Principle.

10. Public understanding

A device like Hawk-Eye, which is squarely in the public domain, should be properly understood by the public whose lives it affects. Furthermore, devices like Hawk-Eye could have a valuable role to play in public education the benefits of which would spread to all technological decision-making in the public domain. It is vital that people understand uncertainty and come to understand that some decisions that are made for the best are bound to turn out to be wrong because of the levels of uncertainty that attend every decision. This paper has set out some of the ways in which this public understanding could be enhanced—by clearly stating levels of uncertainty with every measurement made by widely watched sports decision aids. A new role for sports commentators would be to interpret and explain these levels for the viewer. Sports commentators could become very useful educators of the public.

One specific change, not mentioned so far, is so urgent and important that it might even be appropriate to institute it in law. The law would make it compulsory to stamp every simulation with a "health warning." Every television picture, even a live one, is presented to the viewer in a form that can be represented as an array of numbers representing the color and brightness of each pixel on the screen. We have reached, or will soon reach, the point where arrays of equal richness can be generated in real time by simulation programs. At this point, there will be nothing in the quality of the picture that enables the viewer to tell whether they are watching the real thing or a simulation. The potential can already be seen in the cinema where the arrays are still larger, though the cinema has the advantage that it does not have to generate the numbers in real time.

It is already the case that even the less rich virtual reality graphics used by Hawk-Eye are too seductive. Even now the viewer needs to be reminded that they are not watching something that happened but a picture which has been produced by a calculation based on imperfect data. The graphic, impressive though it looks, is no more accurate than the data on which the calcula-

tions that generate it are based. As suggested above, these graphics should be accompanied by visual error bars and/or numerical statements of confidence as well as the "health warning."[35]

Another useful visual aid would be to represent the number of data points that contributed to each reconstruction (though, see the final note, we now understand that this may be difficult). On the other hand, taking ball speed, bounce point, trajectory length, frame rate, and so forth into account, it might be possible to calculate a probable number of "virtual data points" associated with any one reconstruction.

Access to information

As indicated at the outset, this analysis reports, *inter alia*, a piece of research on public access to technological information. In earlier work (Collins and Evans, 2002, 2007) we analysed the difficulties of public interaction with technology into the "problem of legitimacy" and the "problem of extension." The problem of extension is to understand the limits of the distribution of expertise among the public. Resolving the problem of legitimacy must involve setting enough information before the public to encourage acceptance of new technologies. In so far as we represent the public in this paper, our research reveals that, if previous experience were to apply to this case, then with Hawk-Eye as we have encountered it, and where those tennis fans who contribute to newspaper websites are concerned, there could be a problem of legitimacy. What we elsewhere call "local discrimination" tends toward the negative when relevant information is not provided.[36] The British government has encountered just this effect in respect of statements about the safety of the Sellafield nuclear processing plant, the safety of British beef, and the safety of genetically modified crops. The unfortunate result was a build-up of distrust in the government with unfortunate consequences when it came to government reassurances about the safety of the MMR (measles, mumps and rubella) vaccine.[37] The goal of any public technology has to be, not only to provide accurate information, but to provide it in a way that emphasizes its trustworthiness. We believe this can be done most effectively if information is given freely, openly and accountably.

11. Conclusion

This number of caveats and open questions make it clear that this paper is far from complete as far as the details of its technical argument are concerned. Luckily, most of the technical points made turn on principles rather than particulars. It would help if the questions found in note 8 were answered though even answering these would still leave us well short of being able to complete a full statistical analysis of the errors involved. In our view, a sports decision aid should be as transparent as possible in the presentation of its potential errors. It should also, in so far as is possible, analyse and explain the systematic differences between the measurements it makes and the tradition of human decision-making in the sports to which it is applied. The "Automated Decision Principle" implies that the decision aid should be adjusted to match the human systematic error which is part of the tradition of the sport to which it is applied.

Sport has changed as a result of television replays whether one likes it or not. For example, at least some television viewers find that soccer is being spoiled for them by the number of blatantly incorrect refereeing decisions visible on television replays. Sports decision aids, including television replays (as in rugby union), have a valuable role to play in undoing some of this damage. But the exact way all these things are used depends on a prior

understanding of the relationship between what these devices can do and the way normal human judgment works. Automated sports decision aids, if their capabilities were presented in a transparent way, could add still more to the enjoyment of sport and, in addition, to a better public understanding of the limits and possibilities of technology. If the Automated Decision Principle were followed, and with the increasing speed of computation potentially making automated decision devices more accurate and more capable of analysing and presenting, in real time, the magnitude of uncertainty associated with any reconstruction, the future for the technology as an *aid to human judgment*, seems bright.[38]

Acknowledgements

Notes

1 The phrase "lay expert" is a symptom of this tendency. For an example of the sentiment see ESRC (1999). Policy documents and other reports that appear to reflect the tendency include Council for Science and Technology (2005), House of Lords (2000), Office of Science and Technology (2002) and Parliamentary Office of Science and Technology (2001). An overview of sociological work relating to expertise is provided in Evans (2008).

2 In some cases, this uncertainty is made explicit. For example, the UK's central bank, the Bank of England, presents its inflation forecasts as a "fan diagram" in which the range of possible outcomes, together with their likelihood, are presented as graph in which the forecast appears not as a line of fixed width but as a line that fans out as it moves further into the future. For further details see: http://www.bankofengland.co.uk/publications/inflationreport/index.htm

3 See the work by Kahneman and Tversky (Kahneman and Tversky, 1996; Kahneman, Slovic and Tversky, 1982), who identified the heuristics and biases that lead people to make less than optimal choices. More anecdotally, the range of social science textbooks that offer to explain statistics to the terrified suggests that, at the very least, there is a perception that even university students approach research methods courses from a very low base (see e.g. Rowntree, 1981, for a classic or Garner, 2005, for a more recent version).

4 For example, cricket and baseball are a statisticians' paradise.

5 Jamie Lewis put in the most time in searching the existing websites.

6 Since, whether we like it or not, this paper is largely based on material in the public domain we thought it worthwhile to treat this source slightly differently (discussing it fully only in the last note to the paper) so that our argument can be used as an example of what an assiduous researcher can do using only readily available sources. The technical article was difficult to access in one sense because it was written in esoteric language such as we would have discussed with the authors or other experts in the normal way. In the other sense of accessible, we could not download the article from home without paying a fee or joining a professional organization nor could we find the authors' homepages to ask them directly for a copy. It was possible to discuss the paper because we found we could download it from a Cardiff University based computer. Google Scholar reports that this article was cited by nine other articles. We checked all those we could download from home but nothing we found bore on the questions discussed in this paper. There were discussions of how to track a ball using a single camera, discussions of better kinds of software, and so on. Those we could not download gave no indication in their titles that they would bear on the concerns of this paper.

7 Which would make the potential uncertainties greater.

8 Each question can be cross-referenced for relevance to the text by number. In the text the numbers appear as "*n*" in the symbol {caveat *n*}. Even if all the answers to these questions were known the situation, though it would have been vastly improved, would not amount to a full statistical analysis of the errors.
   Question 1: What is the frame rate of the cameras used by Hawk-Eye? Does Hawk-Eye combine the output of different cameras to produce an "effective frame rate" which is higher than the actual frame rate of the individual cameras? If "yes," is it possible to put a number on the effective frame rate?

Question 2: Does Hawk-Eye use feeds from regular cameras or does it use dedicated cameras? Does the frame rate vary from game to game and sport to sport? Is the accuracy of Hawk-Eye's measurements of the position of a ball in any one frame known? Is it known whether and how this accuracy changes when the light changes and the distance from camera to ball changes?

Question 3: In the case of lbw is the trajectory of the ball estimated entirely from the post-bounce trajectory or does the pre-bounce trajectory contribute and, if so, how? Are historical data generated by the statistical analysis used in data processing to improve the accuracy of predicted trajectories as in computer learning programs?

Question 4: Can Hawk-Eye measure and make use of the spin of the ball in cricket? Is it able to predict the orientation of the ball when it passes the wicket?

Question 5: Has the distribution of Hawk-Eye's errors been measured? If "yes," are the results available for consultation? If the errors have been measured, are they roughly normally distributed? If not, how are they distributed?

Question 6: How, in addition to the 20 mm figure given for "short" bounce-to-pad trajectory, is the accuracy of Hawk-Eye affected by longer and shorter bounce-to-pad trajectories and pad-to-wicket trajectories?

Question 7: Does Hawk-Eye ever report a result based on only two data points? Does Hawk-Eye directly measure post-bounce swing? Does Hawk-Eye measure the rate of change of post-bounce swing?

Question 8: Have the data from the lbw test been preserved? Could they be used to provide an initial indication of distribution of errors? Is it possible to make a more careful analysis of distribution of errors? What would be the limits of such an exercise?

Question 9: Is Hawk-Eye able to measure and report uncertainties in its predictions in real time?

Question 10: In tennis are post-bounce data points used in calculating the bounce point?

Question 11: Is the error for tennis shots greater in the line of travel of the ball than across the line? The mean error for tennis is given as 3.6 mm—is that an average across all ball velocities and directions? If "yes," does that mean that the mean error in the direction of travel is likely to be larger than 3.6 mm for the class of fast moving balls?

Question 12: In the case of tennis, does Hawk-Eye utilize measures of accuracy in addition to those demanded by the ITF test? If "yes," what are they?

Question 13: Has any systematic analysis of human and Hawk-Eye lbw calls been made? If "yes," does it show that the number of lbw decisions given OUT would go up if Hawk-Eye's decisions were applied automatically? If "yes," by how much?

Question 14: Is it the case that a human umpire could make a good decision that Hawk-Eye would be unable to make in the case where a ball bounced very near to a batsman's pad and the batsman was well back and very near to the wicket?

Question 15: Can Hawk-Eye's virtual reconstruction of the line take into account small departures from straightness in the way the line is painted?

Question 16: Is Hawk-Eye capable of recognizing and representing, in real time, the data points from effective single frames in its virtual reality representation?

Question 17: When used in tennis is Hawk-Eye less accurate when there is no post-bounce or only a short post-bounce trajectory? Has this been tested? Did the ITF tests include cases where the ball was stopped almost immediately after the bounce?

Question 18: Is it the case that the outside edge of the line in the micro-world is reconstructed by adding a nominal half-line width to a virtual central line? If "no," how is it done?

9 These website comments reinforce our optimism about the potential for sport to be a medium of statistics education. The relevant URLs are: http://www.timesonline.co.uk/tol/sport/tennis/article2051307.ece and http://www.telegraph.co.uk/sport/main.jhtml?xml=/sport/2007/07/10/sthawk110.xml (both accessed 7 April 2008)

10 Though we understand it has been tried as an umpiring aid in cricket.

11 Although the patent application specifies six cameras in its diagrammatic representation of the system, it also states that the position of the ball can be calculated using only two cameras. The additional cameras are needed to ensure sufficient data are available (e.g. one camera might be blocked by a player).

12 In, for example, Hawk-Eye's analysis of a Collingwood lbw decision: http://www.hawkeyeinnovations.co.uk/UserFiles/File/Collingwood%20LBW%20Dismissal%20Oval%20Test.pdf (accessed 10 April 2008).

13 The technical paper published in 2003 indicates that the actual estimation of the path of the ball may start with what are called "tracklets," which we take to be straight lines between pairs of data points or between pairs of estimates of ball position from two camera frames. As far as we can see this makes no difference to the principles being discussed here. There is only so much information available however it is used and that information cannot be better than what is found in the initial camera frames.

14 Cricket is also played up to international level by women but the number of women who play is relatively limited and the women's game does not compare in salience or tradition with the men's game—as it does in, say, tennis. Throughout this paper we will refer to the men's game.

15 The game is a draw if the batsmen are not all out so simply blocking with the pads could be an advantage to a losing side.

16 It would take about a page to explain the lbw rule fully. Hawk-Eye is also useful in respect of other aspects of the rule, but the element of decision discussed here is enough to make the crucial point.

17 The URL for this web page is: http://content-ind.cricinfo.com/ci/content/story/125938.html (accessed 15 February 2008).

S. Rajesh's quotations of the words and views of Paul Hawkins on the "Cricinfo" website will be referred to several times in this paper. Rajesh works for Cricinfo and is identified on this page as its "assistant editor." The relationship between Cricinfo and Hawk-Eye Innovations appears to be close. For the period between summer 2006 and summer 2007 they were both part of the same company.

18 We are going to talk of confidence intervals as though they are one dimensional though we should really talk of two-dimensional "confidence regions"; the principle of the argument is unaffected.

19 http://content-ind.cricinfo.com/ci/content/story/125938.html

20 http://content-www.cricinfo.com/ci/content/story/250359.html; 20 mm is nearly an inch.

21 Both in cricket and in tennis Hawk-Eye's predictions are likely to be both more reliable and more accurate than human judgment when the ball is not close to the critical zone.

22 http://content-uk.cricinfo.com/ci/content/story/136222.html (accessed 7 April 2008). Interestingly, the sideways-on reconstruction in the case of the Collingwood lbw decision (http://www.hawkeyeinnovations.co.uk/UserFiles/File/Collingwood%20LBW%20Dismissal%20Oval%20Test.pdf) would suggest a frame rate of about half this (accessed 10 April 2008).

23 This is different from the current HIT–MISS calculation as presented on Hawk-Eye Innovations' website. This shows the position the ball would need to be in if it were to just miss the stumps and them plots the reconstructed trajectory with no indication of uncertainty or measurement error. See: http://www.hawkeyeinnovations.co.uk/UserFiles/File/Collingwood%20LBW%20Dismissal%20Oval%20Test.pdf

24 This figure and subsequent figures are drawn to very approximate scale only. Dotted lines in this and subsequent figures indicate range of error only inexactly. Likewise, drawings and percentage figures do not necessarily correspond.

25 The URL is: http://archive.gulfnews.com/indepth/dubaitennis2007/men/10108379.html (accessed 15 February 2008).

26 The URL is: http://www.telegraph.co.uk/sport/main.jhtml?view=DETAILS&grid=A1YourView&xml=/sport/2007/07/10/sthawk110.xml (accessed 15 February 2008). As with the *Timesonline* page mentioned earlier, many of the comments posted in response to this article raise similar points to the ones made in this paper and suggest that at least some tennis fans understand the concept of measurement error.

27 An indication of the size of errors made by humans is given in Mather (2008). It is clear from his paper that humans can make large errors of judgment and we reiterate that where the decision does not concern a bounce very close to the line, Hawk-Eye is, so far as we can judge from the information we have to hand, much more accurate than a human judge. From Mather (private communication) we learn that roughly 13% of 1,380 challenges that he analysed had a bounce point within 5 mm of the line according to Hawk-Eye and roughly 25% within 10 mm. This might give a sense of the proportion of challenges for which the confidence levels might be "worth reporting." Mather's own analysis of human observational error could be to some extent affected by the fact that he takes Hawk-Eye to be the reference point for accuracy even when the decision is close to the line.

28 This is known as the "Peters Method" (Peters, 1856). Thanks to a referee for pointing out that a more accurate figure would be 1.253.

29 The full explanation is available from the Hawk-Eye website at: http://www.hawkeyeinnovations.co.uk/UserFiles/File/Hawk-Eye%20Line%20call%20explained_Final.pdf (accessed 15 February 2008).

30 See also Mather (2008).

31 See, for example, Bijker, Hughes and Pinch (1987); Bijker (1995).

32 Hawkins is quoted by S. Rajesh, as saying that one in five lbw OUT decisions are incorrect and Hawk-Eye could remedy this (Cricinfo, 25 September 2004). But we do not know if this would balance the benefit of the doubt decisions without a more complete analysis. We do not know if such an analysis has been done {caveat 13}.

33 After we completed this part of the analysis we found that the matter had already been considered. S. Rajesh quotes Hawkins:

At the moment, concerns over using Hawk-Eye as an aid for umpires in lbw decisions centre on the number of marginal cases which would go in favour of the bowlers. While that's a reasonable worry, Hawkins insists that there can be a benefit of doubt built within the system, so that close decisions—when the ball is grazing the outside of the stumps, or striking the pad marginally in line—go in favour of the batsmen, just as the spirit of the law indicates it should. (Cricinfo, 13 June 2006)

34 A difference famously explored in Winograd and Flores's (1987) discussion of Winograd's own program designed to stack colored blocks. Nowadays a micro-world might be referred to as a virtual reality.

35 The authors have encountered university professors with a high level of technical ability in the sciences who had not realized that when they were watching Hawk-Eye "replays" of tennis shots they were watching virtual reality reconstructions subject to error.

36 See Collins and Evans (2002 or 2007) for problems of legitimacy and discrimination and Collins and Evans (2007) for "local discrimination."

37 See, e.g. Wynne, 1982; Welsh, 2000; Lacey, 1994; Horlick-Jones et al., 2007; Boyce, 2007.

38 The professional article that we could not download from home is: Owens, N., Harris, C. and Stennett, C. (2003) "Hawk-Eye Tennis System," in *VIE 2003: International Conference on Visual Information Engineering: Ideas, Applications, Experience*, 7–9 July, University of Surrey, Guildford, pp. 182–5. The paper seems to have been written by employees of Roke Manor Research—a private engineering consultancy—which appears to have developed the tennis line-calling system on behalf of Hawk-Eye Innovations. The paper acknowledges that Hawk-Eye Innovations paid for the work. Crucially, and perhaps surprisingly, the paper does not consider in any detail the questions concerning accuracy analysed above. Rather it explains the broad principles of the system. The paper is extremely densely written and includes some technical terms that only a specialist would understand. Nevertheless the general drift of the paper can be grasped after a couple of readings if one has had some experience with technical literature. That said, what follows is written under the standing caveat that we are not engineers and would, in the normal way of our research, have taken this paper as a basis for discussion with its authors, or similarly qualified persons, rather than as something that we would try to interpret entirely on our own.

We believe we can learn from the paper that the prospect of showing what we call "data points" on the reconstructed image of the flight path of the ball {caveat 7; caveat 16} is slim since each flight path is built up from what are called "tracklets" which are short two-dimensional sectors of the path between television images. The tracklets produced by the three cameras are built up into a three-dimensional image quite late in the reconstruction process so there may be no equivalent of what we call "data points"—only an unfiltered jumble of data point "candidates" pertaining to individual cameras. This, however, does not affect the principle of the arguments that we put forward using the notion of "data points." Indeed, our analysis may be better for being more accessible; in the last resort, the apparatus does have to build its model of the trajectory from data gathered from a series of camera frames so the data point idea works even if these data are assembled in a more complex way than our analysis intimates. What is very, and somewhat surprisingly, unclear in the technical paper is how the bounce point is determined and how its accuracy might be affected by frame rate.

   We believe the paper shows that it is worth asking questions about the accuracy with which the position of the ball is located in any one frame. The paper (written in 2003) suggests that Hawk-Eye uses output from the existing broadcast television cameras rather than utilizing special cameras of its own with higher frame rates {caveat 2}. The paper explains that cameras have to be calibrated and the "observed" ball position at any one time has to take account of pan, tilt, and zoom. The paper seems to say that the position of the ball has to be interpolated from the position of the lines on the tennis court and that these are sometimes obscured or out of frame; it seems to say that these lines are themselves reconstructed from virtual line centers. In other words, there is at least some room for error even in constructing the model of the court from which the location of the ball in any one frame is estimated. It does say that, depending on the shutter speed of the camera and the speed of the ball, the ball may sometimes appear as a circle and sometimes a "sausage-shaped streak." The paper says that in a long shot the ball may be represented on the image by as few as two pixels.

The paper does seem to illustrate the difficulty of reconstructing short trajectories, in this case in reference to the half-volley in tennis. Thus a passage on the fourth page of the paper reads:

> Special code was written to deal with the case of a half volley. Often the upward bouncing portion of the half volley was too short to be extracted, and so an estimated "dummy" track was inserted in the impact point chain. The bounce point was determined using a one-sided Kalman Filter applied to the incoming track. Configurable bounce retardation percentages were used to estimate the initial velocity of the upward bouncing track at the instant just after the bounce. This was used to create a purely ballistic intermediate track, with an acceleration a of $[0, 0, -g]T$. This dummy track was inserted in the chain of tracks, providing an unbroken set of tracks for the visualisation module.

We understand this to mean (under the standing caveat that we are not engineers), that the upward portion of bounce in the case of a half-volley often did not allow the collection of usable data. It was therefore estimated to be a function of the velocity prior to bounce, the loss of speed due to the bounce, and the deceleration due to gravity.

The paper appears to imply that in tennis both pre-bounce and post-bounce trajectories or "tracks" are used in the estimation of bounce point {caveat 17}.

We understand the paper to reveal that the court lines are continually reconstructed in real time. The lines are photographed, a virtual center is constructed from the photograph for each line in short sections and these are then joined to reproduce the straight center of the lines as a mathematical object with zero width. The position of the edge of the line is then reconstructed, we guess, by adding a nominal half-line width to this virtual central line. This would seem to suggest another possible source of difference between the micro-world and the real world {caveat 18}. It is worth noting that all these uncertainties could be resolved in a short conversation or a few e-mails.

The very density of the description of how the position of the ball and the bounce was estimated, and how the track seen by the viewer was reconstructed, seems to reinforce the argument made above that the only sure test of the accuracy of a device such as Hawk-Eye is direct measurement of errors in comparison to some more absolute test of position such as a very high speed camera.

References

Bijker, W.E. (1995) *Of Bicycles, Bakelites, and Bulbs: Toward a Theory of Sociotechnical Change*. Cambridge, MA and London: MIT Press.

Bijker, W.E., Hughes, T.P. and Pinch, T.J., eds (1987) *The Social Construction of Technological Systems: New Directions in the Sociology and History of Technology*. Cambridge, MA and London: MIT Press.

Boyce, T. (2007) *Health, Risk and News: The MMR Vaccine and the Media*. New York: Peter Lang.

Collins, H.M. and Evans, R. (2002) "The Third Wave of Science Studies: Studies of Expertise and Experience," *Social Studies of Science* 32: 235–96.

Collins, H.M. and Evans, R. (2007) *Rethinking Expertise*. Chicago: University of Chicago Press.

Collins, H.M., and Pinch, T. (1993) *The Golem: What Everyone Should Know about Science*. Cambridge and New York: Cambridge University Press.

Collins, H.M. and Pinch, T.J. (1998) *The Golem at Large: What You Should Know about Technology*. Cambridge and New York: Cambridge University Press.

Collins, H.M. and Pinch, T.J. (2005) *Dr Golem: How to Think about Medicine*. Chicago: University of Chicago Press.

Council for Science and Technology (2005) "Policy through Dialog." URL: www2.cst.gov.uk/cst/reports/files/policy-through-dialogue/report.pdf (accessed 15 February 2008).

ESRC (Economic and Social Research Council) (1999) "The Politics of GM Food: Risk, Science and Public Trust," ESRC Special Briefing No. 5, October. Swindon: ESRC.

Evans, R. (2008) "The Sociology of Expertise: The Distribution of Social Fluency," *Sociology Compass* 2(1): 281–98. URL: http://dx.doi.org/10.1111/j.1751-9020.2007.00062.x

Fischetti, M. (2007) "In or Out?," *Scientific American* July: 96–7.

Garner, R. (2005) *The Joy of Stats: A Short Guide to Introductory Statistics in the Social Sciences*. Peterborough, Ontario: Broadview Press.

Horlick-Jones, T., Walls, J., Rowe, G., Pidgeon, N., Poortinga, W., O'Riordan, T. and Murdock, G. (2007) *The GM Debate: Risk, Politics and Public Engagement*. London: Routledge.

House of Lords (2000) *Science and Society: Science and Technology Select Committee*, Third Report. London: HMSO. URL: http://www.parliament.the-stationery-office.co.uk/pa/ld199900/ldselect/ldsctech/38/3801.htm (accessed 23 January 2002).

Kahneman, D. and Tversky, A. (1996) "On the Reality of Cognitive Illusions," *Psychological Review* 103: 582–91.

Kahneman, D., Slovic, P. and Tversky, A. (1982) *Judgement Under Uncertainty: Heuristics and Biases*. Cambridge: Cambridge University Press.

Kuhn, T.S. (1961) "The Function of Measurement in Modern Physical Science," *Isis* 52: 162–76.

Lacey, R. (1994) *Mad Cow Disease: The History of BSE in Britain*. St Helier: Cypsela.

Lave, J. (1988) *Cognition in Practice*. Cambridge: Cambridge University Press.

Mather, G. (2008) "Perceptual Uncertainty and Line-Call Challenges in Professional Tennis," *Proceedings of the Royal Society B* 10.1098/rspb.2008.0211

Office of Science and Technology (OST) (2002) *The Government's Approach to Public Dialogue on Science and Technology*. London: OST. URL: http://www.ost.gov.uk/society/public_dialogue.htm (accessed 27 September 2005).

Parliamentary Office of Science and Technology (2001) *Open Channels: Public Dialogue in Science and Technology*. Report No. 153, March. London: HMSO.

Peters, C.A.F. (1856) "Uber die Bestimmung des wahrscheinlichen Fehlers einer Beobachtung aus den Abweichungen der Beobachtungen von ihrem arithmetischen Mittel," *Astronomische Nachrichen*, 44: 30–1.

Rowntree, D. (1981) *Statistics without Tears: A Primer for Non-Mathematicians*. Harmondsworth: Penguin.

Welsh, I. (2000) *Mobilising Modernity: The Nuclear Moment*. London: Routledge.

Winograd, T. and Flores, F. (1987) *Understanding Computers and Cognition*. Norwood, NJ: Ablex.

Wynne, B. (1982) *Rationality and Ritual: The Windscale Inquiry and Nuclear Decisions in Britain*. Chalfont St Giles: British Society for the History of Science.

Authors

Harry Collins is Distinguished Research Professor and Director of the Centre for the Study of Knowledge Expertise and Science (KES) at Cardiff University and one-time President of the Society for Social Studies of Science and holder of its Bernal Award for Lifetime Contribution to Social Studies of Science. His research is on the sociology of scientific knowledge, with a concentration on gravitational wave physics, artificial intelligence, and, more recently, on the nature of expertise. Correspondence e-mail: CollinsHM@cf.ac.uk

Robert Evans is a Senior Lecturer in Sociology at the Cardiff School of Social Sciences. His main interests are in Science and Technology Studies (STS) and he has published on economic modelling, interdisciplinary working and the nature of expertise. He is currently working on a project evaluating the relationship between deliberative methods and qualitative research more generally and, with colleagues in the Centre for the Study of Knowledge Expertise and Science (KES), is investigating the imitation game as a methodology for researching expertise.
Address: Glamorgan Building, King Edward VII Avenue, Cardiff, CF10 3WT, UK; fax: +44 (0)29 2987 4175; email: EvansRJ1@Cardiff.ac.uk