A Physicist looks at Sport

Professor John D Barrow FRS

25 October 2005

You might think I'm just going to tell you about watching Match of the Day at 10.30 on a Saturday evening. But we are going to look at a variety of different situations in a number of sports and see how some basic physics illuminates what goes on. There is of course vastly more sport that I can't talk about than I can, and if the sport that you do is not included, then I can only apologise.

I want to start off with talking about balance. If you watch gymnastics or if you go to the circus and watch a tightrope walker walking along a high wire – I was hoping that one of the Gresham staff was going to demonstrate, but sadly not! – you will notice that they often tend to carry a long pole as they balance. The first question is, why do they do that? Why do people carry a long pole when they want to balance on a wire?

If you go around and ask people what is going on, you will get a variety of answers, even from engineers and mathematicians. They think may be the pole makes your centre of gravity lower, so that when your centre of gravity is lower, you are more stable than when it is raised. In fact, the pole makes your centre of gravity higher; you're worse off carrying the pole in that respect. If you want to maintain balance, obviously you want to be symmetrical about the centre. Generally, if you are trying to balance on the ground, you want to broaden your base, so with your legs apart, you can balance better than with your legs together. But the key to understanding balance and why these tightrope walkers are carrying a pole is the concept of moment of inertia.

The thing that determines whether objects can move easily or not is called their moment of inertia, and what this measures is the distribution of mass in a body. If the mass is primarily far away from the centre, you have a high inertia; if it is close to the centre and concentrated, you have a low inertia. So if you have two bodies which have the same diameter, the same radius and the same mass, but one is a hollow shell and one is a solid sphere, if you try to move them, if you try to spin them or roll them, you will find the hollow object moves more slowly; it has a greater inertia. If you roll them both down the same slope, the solid object will roll down more quickly to the bottom than the hollow one. What determines your inertia is your mass and your size, but also some measure of the concentration of the mass. So when the mass is far away, you have a large inertia. Now, this is what is going on with the tightrope walker with the pole: by carrying the pole, you distribute mass far away from your centre, you increase your inertia, and that means that when you move, you move more slowly. So when you wobble, you wobble like an oscillator with a very long period, and so you have lots of time to correct yourself.

If I try to balance a pointer on my finger when it I s contracted and rather small, it's much harder to do it than when I completely extend it, when its inertia is larger, and again, it's because when it wobbles, it wobbles more slowly in this case. What's going on here is that the moment of inertia, the distribution of mass, is being increased, and it is harder and slower to move.

There are many sports where this inertia effect is important, and particularly where any type of rotation or spin is involved. If you are a gymnast or a high board diver and you want to undergo several somersaults as part of a manoeuvre before you get to the water or the ground, then clearly you want to rotate as fast as you can in order to have room to do more somersaults. So if you want to rotate fast, you want to have a small inertia, and that means you want to have your mass distribution more concentrated; that's why high board divers will tend to move into this tuck position of a ball: they will spin more rapidly. If you look at an ice skater pirouetting, they will start to spin with their arms out: their inertia is high. As they draw their arms in, their inertia becomes smaller, and they will spin faster.

There are other sports where this same type of effect occurs. For example, with the gigantic tennis racquets that have evolved over the years with enormous heads, there is more mass a long way away from the racquet handle. The inertia is higher, and so it's more stable. It does not reverberate on your hand; it doesn't move so easily.

If you look at cyclists, then you will see that cycle wheels don't look like the cycle wheels that you have on your bike. If you are trying to break the world sprint record indoors on the track, you may have wheels that are much more disc-like. So again, the moment of inertia of a fancy, high-tech racing cycle wheel is quite different to the moment of inertia of your ordinary cycle wheel.

If you are a runner, this comes into play at a rather simpler level. If you run with your arms rather raised, you behave rather differently than if you have your arms down. When your arms are down, your inertia is larger, there's more mass a long way away from the centre of movement; when you bring your arms up, then a very slight effort can make a small movement restore balance very easy. So a good running style optimises this inertial effect, reduces the inertia, makes you able to make small changes very quickly – the opposite to the tightrope walker, who wants lots of time to make small changes.

Let's move on from balancing to jumping. High-jump nowadays is a rather spectacular event, even more so if you don't have an airbed landing area! Of course, almost every high jumper that you see on the television or at a top class athletics event nowadays uses the so-called Fosbury Flop technique, which first appeared on the international athletics scene in 1968 when Dick Fosbury, the American high jumper, introduced this technique at the American Olympic trials and then he won the gold medal at the 1968 Olympics, as the only jumper using this technique. The interesting question to ask is what's going on here, so why is it effective to use such a technique?

Well, what are you doing when you high jump? You are simply transferring energy in one form to another. You have a certain energy of motion, and you convert that into lift and you do work against gravity. What's happening is that your incoming energy gets converted into your weight, your mass, times the acceleration due to gravity, times the height above the ground that it's raised – when I say "it" is raised, what we mean by "it" here is your centre of gravity. This is the key about high jumping: you use your incoming energy of motion to raise the centre of gravity of your body, your weight, by some height.

When you are at school playing around in the sandpit, you probably began by doing high jumping using the so-called scissors technique – so you basically just run in and step over the bar at high speed. This is a safe technique; it's very simple, but it's very inefficient, because not only is your body clearing the bar, but your centre of gravity is going over the bar, and it's going really rather far over the top of the bar. With a scissors jump, the centre of gravity of your body goes way up here over the bar; it clears the bar by a large amount. But what goes on with a technique like the Fosbury Flop, your body is bent over, rather like a banana, it's curved over in this way, and for a shape of that sort, the centre of gravity lies somewhere here. Using the Fosbury Flop technique, it is possible for your body to go over the bar but your centre of gravity to go underneath, so it's a much more efficient way of using your launch energy to go upwards.

Stefan Holm, who won the gold medal at the Athens Olympics, a remarkable high jumper, cleared 2.37m; that's not a world record, but he's only 1.80m tall, so he jumps way, way above his head height. One of the reasons he was so effective was he fantastic curvature of the body, which enabled his centre of gravity to go way below the bar, even though his body went over it.

This type of technique we see not just in high jumping but also in pole vaulting. A pole vault technique plays a similar type of game: you have an incoming energy of motion, horizontal motion, which you want to convert to vertical lift, vertical work against gravity. Exactly the same principle holds: you have to convert the kinetic energy that you come in with into height that your centre of gravity is raised multiplied by your weight. High jumping is relatively inefficient at transforming that incoming energy into vertical lift. It's technically complicated, you have to lose quite a lot of energy in the take-off process, but pole vaulting is rather efficient. There's an intermediate step. Your energy of motion gets converted into elastic energy by flexing the pole, and then the pole unfurls and launches you upwards, and if you are good, you will make use of the same technique as the high jumper: you'll curl your body over the top of the bar so that your centre of gravity goes underneath.

With rather simple mathematics, I'll show you a few equations in this talk, just to show you how you can understand what's going on in these events very easily. In pole vaulting, the kinetic energy that you lose when you take off goes into raising your potential energy or raising your weight. Well, here's your kinetic energy – the half times the mass times the velocity squared – that you come in with, and that gets turned into your mass times acceleration due to gravity times the height that your centre of gravity is raised. So your mass cancels out; that doesn't make any difference, and so the height just looks like the square of your speed divided by twice the acceleration due to gravity. Well, if you come in at ten metres per second, like a top class sprinter, acceleration due to gravity is about ten, you expect to raise your centre of gravity five metres, and that's almost exactly right for an Olympic winning vault. The Athens Olympics I think were won in 6.01m. So given that your centre of gravity starts a bit over a metre above the ground at the beginning and when you go over the bar it'll go a little below, this predicts that the height cleared will be about 6 metres.There's almost perfect transfer of the initial energy into the final result.

High jumping, the world record is 2.45m – fairly amazing, mark it out when you get home, it's over eight feet, nearly eight feet one, but you can see it's vastly less than this. It's about half this, and so high jumping is about half as efficient at transferring horizontal into vertical energy.

Well, one of the things that you have to have in those sorts of events is a lot of strength in order to launch yourself upwards. We're going to look at little about what we can learn about strength, and how it depends on how big we are. Strength is something that does not grow at the same rate as your mass or your weight, so as things get bigger, their strength grows in a very particular way, and you can understand that if you think about what you have to do to break something. In this case, suppose you want to break this piece of paper. We just tear it. So all we have to do is to break it along a little cross sectional surface, so all we have to do is to break all the atomic bonds along a line here, so we have to break an area. You can see that in this picture, that if you want to break this stick here, you've got to break an area of atomic bonds. So in general, strength is proportional to an area, it's proportional to the square of your size. On the other hand, your volume as you get bigger grows like the cube of some measure of your size; it grows more quickly.

If you have animals that just change in size but have the same basic body plan, like the cat and the kitten, as they are getting bigger, they are getting relatively weaker; their strength grows more slowly than their weight. I picked cats because you can see for yourself this effect. If you look at a big cat and it puts its tail in the air, you'll see the tail will curve over, but if you look at a kitten, you'll see that the tail sticks up rather like a sharp spike. The big cat is not strong enough to keep the tail upright, so as it's grown, its strength has not grown at the same rate as its size, whereas the kitten is still strong enough to do that.

Well, we want to think about that idea in connection with people like this: weight lifters. As you know, weight lifting is one of those sports where there are weight categories. So if you weight 350 pounds, you will not be competing against people that weigh120 pounds. Well, what does looking a little more closely at the strength against size variation tell us? Just to restate what we've seen: your mass looks like your density times your volume. All people have about the same density; it's fixed by the density of atoms and solids, and so your mass will look like your density times your volume, and your volume grows as the cube of some measure of your size. So as you get bigger, your mass will grow in proportion to the cube of your size.

As we've seen, as we change the size, your strength will grow just as the square of your size, because it depends on the cross section area of your muscles and so on. You can see here that there is a sort of relationship between your weight, going like R cubed, and your strength, going like R squared. I call it the "two-thirds rule", so your strength is proportional to your weight to the two-thirds power.

Strength is proportional to weight to the two-thirds power, or if you like, your strength cubed is proportional to your weight squared. If you're interested in a strength to weight ratio, then you would have to divide the strength by weight, and it would like one over weight to the third or one over the size. As with the cats, as you grow bigger, relatively speaking, your strength falls. A little dog can carry another little dog on its back, you can just about carry another person piggy back on your back, but a horse could not carry another horse on its back, and an elephant could not carry another elephant on its back. So as you get bigger, relatively speaking, you get weaker.

Well, I thought to try this out by looking at the world weight lifting records. Remember that what we're predicting is that your strength cubed is proportional to your weight squared. Well, the strength of a weight lifter is the weight that they lift, and their weight is tabulated rather meticulously by the World Weight Lifting Association. So if we plot the cube of the weight that's lifted by a weight lifter against their own weight, we look up all the world records in the different weight categories, plot them on this picture, what you see here is really rather beautiful agreement with this simple rule, that the world weight lifting records simply follow this two-thirds strength against weight rule. You can even see – so here's the two-thirds slop, here's the line giving that rule – you can even see who is the person who is really the strongest pound for pound lifter: that's the person who is most above this line. There's one person I haven't put on the picture, and that's the sort of super-duper heavyweight category, where the weight of the lifter can be absolutely unlimited, and the person who holds the record there has some stupefying weight of 350kg or something bizarre. If we were to put his weight lifted on this picture, it would be way below the line. He is easily the weakest pound for pound lifter. If he tried to lift the weight that this line would predict that he should lift for his weight, he would break, so the forces that would be exerted on his body would break the atomic bond.

By a similar type of analysis, you could think about other sports, like shot putting, discus throwing, where size and strength clearly play a role, but there are no weight categories, so we don't have weight categories for hammer throwers and other strength events. Of course in practice, only the biggest people tend to take on those events, but just as you have lightweight and non-lightweight crews in rowing, you could have weight categories in many other events, and the way you would figure out whether that was justified or needed would I think be simply by drawing a picture like this of what were the distances achieved by different competitors against their own weight.

Well, that's really a battle, as it were, against gravity of strength. I want to just point out a few other things about gravity that may not be so obvious. If you were awake just now when we talked about launching upwards to do the pole vault, you remember that the height being achieved depended on the launch speed squared divided by the acceleration due to gravity. It was actually V squared over 2G. Whenever you launch something upwards or you throw a projectile like a discus or a hammer, the height that you achieve or the distance that you go always depends on this combination: the square of the launch speed divided by the acceleration due to gravity. It couldn't be otherwise, because they're the only two things that affect the problem, and if what you're interested in is a distance, either upwards or along the ground, the only way that you can combine a velocity and an acceleration to get a quantity with units of a distance is like this It makes sense if you increase the launch speed, you will obviously throw or jump higher. If gravity is weaker, so if G is smaller, then again, it will be a smaller force that you're working against and you'll be able to throw or jump further or higher.

Well, the interesting thing about the Earth is that the acceleration due to gravity varies around the Earth's surface: it's not the same everywhere. That means of course that in some places it's smaller than it is in others. This variation occurs for two reasons. The first is that the Earth is not perfectly spherical, so it's slightly flattened, it's slightly squashed, and in some places there are mountains and there are high altitude venues. So there is an effect on the acceleration due to gravity just because of the shape of the Earth. But the more significant effect arises because of the rotation of the Earth. If you're standing at the North Pole or the South Pole, then the acceleration due to gravity is created by the force of gravity exerted by all of the mass of the Earth underneath you pulling you towards the centre; but if you move to the Equator, you have that same effect, countered by the centrifugal effect of the rotation of the Earth, pushing you outwards, so the net force of gravity, the difference between its intrinsic pull and rotation pushing you outwards, is smaller at the Equator than at the Poles, and obviously as you move from the Equator to the Poles, it varies steadily.

This is a simple, interesting effect, so the value of the acceleration due to gravity at the Equator is smaller than it is at the Poles, so if you weighed something with a spring balance, so if you had a mass from the market with one kilogram on it, there's one kilogram of atoms in it, and you weighed it with a spring balance, it would have a different weight at the Equator than what it has at the North Pole – tricky problem for market inspectors up in the Arctic climate.

Well, how big is this effect? Well, it's not very large, but it's significant. So for example, if you had a 200 kilogram mass, weightlifting barbell, then that 200 kilogram mass in Mexico City weighs 200.8 kilograms up in Helsinki. If you want to break the world weightlifting records, you should head towards the Equator. A place like Mexico City also happens to have high altitude, so there's the Earth's shape effect to help you as well. Similarly, if you are high jumping, just the effect on V squared over G, the effect on G, a two metre high jump in Helsinki is equivalent to 2 metres and 5 centimetres in Mexico City, just because of the change in acceleration due to gravity on the Earth's surface. That's very significant – easily enough to change places and medal positions in a major championship. An 8 metre long jump in Helsinki is worth 8 metres and 20 centimetres in Mexico City.

I want to move on to look at things to do with time, and timing, now. To begin with, a very simple example about why you might worry about timing. Here's an athletics event – equally it could be a swimming event – and if this is the parents' race at the school sports, which you're probably taking notes about to get some advantage, then you know how it works. You've got eight lanes on an international track here. The average, the maximum allowed size for a lane is 1.25m, the smallest is 1.22m, and there will be a starter over here. Now, suppose the starter just had an old-fashioned starting pistol, and says "On your marks, set," bang. What then happens is the sound travels at the speed of sound, so at sea level in ordinary temperature conditions, that's about 340m per second, 750 miles an hour. The first question is do you gain a significant advantage by being close to the starter as opposed to being far away? What you want to know is that what's the time it takes for the sound, for example, to go from one lane to the next, so does the person in 8 hearing it before 7 gain a significant advantage, and then you might ask, well, what about then between 8 and lane 1? The interesting thing is that, easy to work this out, you know distance equals speed times time, so the time difference is just the distance between the lanes divided by the speed of sound. Well, for say, two people in lanes next to each other, the time delay is 4000 th of a second, so you wouldn't worry too much about that. Electronic timing goes down to 100 th of a second, although that's 10cm at the end of an Olympic 100m final, so it's not a bizarrely small amount. But if you were to look between lane 1 and lane 8, so you multiply this by 7, then the advantage to the person in lane 8 over lane 1 is 300 th of a second, and that certainly is significant. That's almost the whole difference between coming first and fourth in last year's world championships. However, don't be alarmed: this is one of the reasons why, in top class athletics events, each athlete has a microphone behind their own lane, so you don't have the school sports starter in the infield firing the gun, so everybody hears the sound electronically at the same time. This shows you that that really matters. If you are running in the parents' race or the grandparents' race at school sports, make sure that you are in lane 1!

Another interesting matter of timing that one might think about, by using the same simple formula, that time is distance over speed, is to think about how quickly you have to react in different sports, so how dynamic, how alert do you have to be, how responsive. If you look at a football penalty kick, suppose you're the goalkeeper, (because it's football, there are strange irrational units like yards, so there are 12 yards from the penalty spot to the centre of the goal – it's only a little bit more if you were to fire at an angle to the corner), a top class player, someone like Alan Shearer, will probably hit the ball at about 80 miles an hour, the goalkeeper's got 0.3 of a second if he really doesn't move before the ball's kicked, so that's the reaction time in football.

If you're at the wicket with Flintoff coming in, bowling in about 95 miles an hour, as he was this summer, or McGraff, you've got a bit more time, about 0.5 of a second to react.

If you're playing at Wimbledon, and you're sort of a non-English player, that is, you know, near the final stage of the competition…you could be hitting, Ivanisevic I suspect probably serving 130, 140 miles an hour. Again, you have about 0.3 of a second if you're at the other end and position yourself return or take the new position to receive the next serve.

Table tennis, again, you have about 0.3 of a second to respond to the ball coming over.

In a game like ice hockey, if you're the goalkeeper in ice hockey, you have even less – well, you have no chance really. The goal is much smaller; you just have to fill it and hope the ball hits you.

But you see, this is quite interesting. It's presumably no accident here. I've looked at all these different sports, and there's this, even at the top level, there's this reaction time of about 0.3, 0.5 of a second, so presumably, this is challenging but it's realistic. You couldn't have sports where the reaction time was hundredths of a second, or it becomes rather boring if it's sort of 5 or 10 seconds. It becomes a bit like bowls or something like that.

Just for comparison, if you're competing in say swimming or athletics, where if you move too soon you'll be disqualified, so the computer will sense whether you make a deliberate or involuntary movement and press on your starting blocks, you'll be disqualified if you respond less than 100th of a second after the starting gun has gone. But the top flight sprinters, the very fastest starters in the world, people like Colin Jackson, they make legal starts with response times of about 14 hundredths of a second. These are among the fastest human responses to stimuli.

Well, these are ballpark reaction times to individual events. If you go to something like a team game, team games have a number of aspects. They have some pitch of a certain size, so it has an area. If it's football, you know, it might be as much as 100m long or something like that. There are a number of players who move around in a semi-random way. As you go down the divisions of the football league, it becomes more random…and they have some sort of average speed…and there is a certain density of players. I've labelled this by – so N here is the number of outfield players per team. In rugby, it would be 15, but in soccer, it would be 10. By treating this as a random process, you can work out what the average separation of the players is, divide the available area by the number of players, and you can work out the average time between encounters of players or encounters of players with the ball. The formula you get looks like this. It's characteristic of a diffusion-type process, depends inversely on the speed, so as the speed of the players gets greater, obviously you have less time to react. It depends on the square root of the area of the pitch: as the pitch gets bigger, the players get rather lost, and there's a long time between interactions; as the number of players gets bigger, things get crowded, you run into people, you receive passes more often. So if you take football, for example, the biggest pitch that you might get at Arsenal or Wembley Stadium, about 100m by 64m, 6400sq m, take the square root there, and you've got 80, and you've got 10 outfield players per team, average footballer probably moves around at about 5m per second when he's feeling in a hurry. If you put the numbers in here, the typical reaction time in a game taking place on this pitch, with this type of speed, is 8 divided by the square root of the number of players in seconds. When N is 10, you're looking at sort of 2 or 3 seconds as being the typical engagement time. If you have a game like rugby, where N is 15, then the time becomes a little shorter – there are more players on the pitch. Sn this way, you get a feel for how quickly you have to react in a game like football.

Well, the next thing I want to say about football is something probabilistic, so I want to look at the results of football matches over a season, and show you how some rather simple probability can be quite revealing about what's going on in a league championship. I just want to ask a provocative question: is the Premier Football League just a random process? So is there any skill element over and above what would be expected from a purely random statistical process?

Well, let's model the Premier League like a random process. There are 20 teams, and they each play everyone else home and away, so they play 19 times 2, 38 games. Just to simplify life, let's look at the history a bit. So if you look at the statistics, pretty constantly every season, one in four of those games is a draw, so a quarter of Premier League games are a draw. We'll assume that the probability of a game being a draw is one in four, 25%. Let's be democratic and assume that the probability of a home win or a home defeat are the same, so we won't pay attention to home advantage. Because one in four of the games are drawn, three out of four of them must be wins or losses, so it must be a three-eighths chance of a home win and a three-eighths chance of an away win. Three-eighths plus three-eighths plus a quarter is one. So these are the rules for our artificial league. Each game has a one in four chance of a draw, a three-eighths chance of a home win, and a three-eighths chance of an away win.

Now, you just play the random league. You could do this with an eight-sided spinner or an eight-sided dice, marking two of the sides "Draw", three "Home Win" and three "Away Win". You would have to be fairly patient: it's a lot of games to play. It's better to write a little computer programme, and see what the results are.

So you do this, and you just order the teams in terms of the number of points that they get, and here's the result of our random league. So the team that comes out top here is Number 1, going all the way down to Number 20, and you can see there's a fair spread. The top number of points - three points for a win, one for a draw, zero for a loss – has come out at 67, and down at the bottom of the table is 31. The average, which you can compute exactly from the rules, is actually 52.5, and sits about here.

How does this compare with the actual League? Here you have a bit of a shock, because here's what happened last season: so the bottom is Southampton, and at the top, Chelsea, and here you have Arsenal and Manchester United. What you notice here is that if you take out the top 3 teams, Chelsea, Arsenal and Manchester United, the rest of the League is really pretty completely mirrored by the purely random process, all the way down to the bottom, essentially the number of points, the spread between the top and the bottom. So what's this telling you? The top 3 teams, they have a much better chance than three-eighths of winning games, more like seven-tenths, so they do much better than this random process, but once you get below them, it's all completely consistent with a random process with the rules that I gave. This isn't a quirk of last season. I did exactly the same thing for the season before. The structure's the same, there's just a slight transposition of names: it's Arsenal at the top now, and Chelsea and Manchester United. Again, those three teams alone are doing better than the purely random model with a three-eighths chance of a win, but once you go below, look, 60 actual, 61 in the model, all the way down to 31, 33, the whole of the rest of the pattern of the League is consistent with the purely random process. So it's worse than you expected!

I want to move on to the last sort of main example I'm going to look at. It's something that's a little more complicated, so I'll go a bit more slowly. One of the fascinating things about sport are the scoring systems.

Sport, for various historical and cultural reasons, has a multitude of different scoring systems. So if you play tennis, the whole game, set and match structure, 15, 30, 40 love, these strange numbers inherited from the values of French coins when the game began in France. If you go to table tennis, you have a more simplified type of scoring. They changed the scoring system in table tennis a year or so ago, and they now just play to 11 rather than 21.

There are many games where there is an added ingredient, and you find this in squash and you find it in, say, volleyball, that you only score a point when you win a point on your own serve. If you're receiving the ball and you win a point, that simply gains you a serve; it doesn't gain you a point.

Let's think about the probability theory of a game like this. So suppose that your probability, your chance of winning a point that's being played is P. We can put in a number for that later on. Your chance of winning a point when you're serving is S for serving. But what's the probability of you winning a point, scoring a point, when you're a receiver? Well first of all, you've got to win a point with probability P to become the server, and then you've got to win your service. So, your probability R as a receiver winning a point is P, to gain the serve, multiplied by S, to win the point once you are serving. Well, if you then serve, what's your probability of winning the point when you're serving? Well, you could either win it straight off, with probability P, or you could lose it, with probability 1 minus P, but then gain service back and win the point for the second time. So your probability of winning it when you're serving is the probability that you win it immediately times the probability that you lose your serve multiplied by the probability that you win the point when you're receiving. So it's this combination of two terms, so there's a little bit of algebra here to solve, and the upshot is that you can work out what's your probability of winning the point when you're a server, what's your probability of winning it when you're a receiver, just in terms of P, your chance of winning the point in any rally. The answer is a sort of slightly messy algebraic formulae, and the interesting thing to note is that if you had 50:50 chance of winning any point, then your probability of scoring a point when you're receiving is one-third, but it's two-thirds when you're serving, so you have a big advantage under these rules of being the server. That's obvious because you can win, score a point immediately; you don't have to go through the double track process.

Now, squash has a strange rule – I don't play squash, so I just asked people about this and I discovered it. If the score reaches 8 all, then the receiver can choose to play first to 9 or first to 10. So the question I ask when someone tells me this is which option should they choose? What should you do? Should you play to 9 or should you play to 10? ell, the answer is not entirely trivial. Using what we've seen already, suppose that you decide you'll play to 9 when you're receiving, what's your chance of winning? Well, it's this thing that we called R just now, which is equal to P times S, so you'd have to gain serve and then it would be your chance of winning as a server. But what if you play to 10? There are 3 ways which you could win. You could either win the next 2 points, so as a receiver you would first win to gain serve and then while serving, you would win another point, so this would be R times S. Or you could win a point, then lose it, then win it again to re-gain serve. Or you could lose the point, but then win and win again. So you have these 3 routes, and these are the probabilities for you going along those 3 routes, so your chance of winning if you start by receiving and you're playing to 10, is the sum of these 3 probabilities. The question you've got to ask is that is this sum here bigger than R? Remember R and S we know just in terms of P. So the chance of winning if you play to 10 is this, winning to play to 9 is that If you look at the algebra here, there's a rather simple answer, that playing to 10 will be best if S is bigger than a half. Go back to your other formula, this occurs when P is bigger than 0.38. So if you have a 38% or better chance of winning any point in the squash game, you should elect to play to 10 not to 9. Why is that? Well, intuitively you can understand it. ou see, if you're a good player, so your chance of winning the point was 50 or 60%, you should play to 10; if you're a bad player, you might fluke one point and win playing to 9, but your chances of fluking 2 points to win to 10 are less. So the good players should always play more points because he's more likely to win out in the long run.

This is an interesting example, that in a game like squash, with this type of only winning points on your serve and this added rule, there's a rather complicated probability structure about what's going on. In some games, you play a sequence of games and sets, and similar type of analysis would be able to tell you that if you had the probability of winning a point in tennis, how would that translate into the probability of winning the game, or of winning the set? What you'll discover is that as the probability of winning a point gets, moves away from a half, so the players become more unequally matched, you really don't need to play many sets. You'll perhaps get away with just playing one. The probability of winning the set is very strongly determined by the probability of winning a point. But as the probability of winning a point gets close to a half, you really do need to play several sets to be sure that the better player in the long run is actually winning the match. The better the players get, the more closely matched they are, the closer P gets to a half, the more sets you have to play to reliably pick out the better player in the long run. That's why in the early rounds of a top tournament you might just play 3 sets of tennis, but at semi-final and final stage, you probably need to play 5 – unless Feder is playing, and then you don't need to play very many at all!

Finally, I want to mention something about judging. The Winter Olympics is of course the place for unusual judging, particularly skating. I don't think we fully even know what the judging structure is going to be of the skating events at the Turin Winter Olympics. But judging is a complicated and paradoxical thing to set up, and you have to be very careful that seemingly straightforward judging rules don't lead to logical paradoxes. So the Nobel Prize in Economics was given for discovering this in a very widespread way back in 1972, by Ken Arrow. Here's a very simple example.

Suppose that we've got 3 competitors, Alice, Bob and Chris, and they're performing in some subjective judging event, like skating or gymnastics, and there are 3 judges, and the 3 judges have the job of ordering them, ranking them, first, second and third. Well, what we discover is that the first judge on their scorecard has Alice in first place, Bob in second place, and Chris in third place, but the second judge, a Russian judge as it were, has Chris in first place, Alice in second, and Bob in third, and then the third judge, has Bob in first place, Chris in second place and Alice in third. Now, the problem is that they now have to combine those judgements to get final results, and that can produce a very strange situation, because you can see from these results that what happens is that Alice beats Bob here, Alice beats Bob, Bob loses to Alice, so Alice beats Bob by 2 votes to one, and Bob beats Chris by 2 votes to one, because Bob beats Chris here and Chris beats Bob, but Bob beats Chris. So Bob beats Chris by 2 to one, but Chris beats Alice by 2 votes to one, so you have a sort of circular paradoxical situation: A beats B, B beats C, but A does not beat C. This is not a sort of obscure little trick, as it were: it turns out that almost any voting system that you care to create which is subject to a small number of reasonable rules ends up producing paradoxes of this sort. What's happening here is something like preference, or liking something, is not something that has this transitive property as we call it. Something like being taller than does. So if Alice is taller than Bob, and Bob is taller than Chris, then Alice is necessarily taller than Chris – that's a property of being taller than. But preferring, or liking, so if Alice likes Bob, and Bob likes Chris, that doesn't ensure that Alice likes Chris. These preference voting systems have this in-built paradox.

I think probably the International Skating Union has to struggle with problems of this sort. When political scientists and economists set up rules for trying to produce best possible or fairest possible voting systems, they usually require certain properties to hold of the voting system, and one of them is, for example, that if you voted for one particular person, then it couldn't reverse the order of preference of 2 other candidates who have nothing to do with the one that you're voting for. This is called the exclusion of irrelevant alternatives. You would expect that if you gave a score for a certain competitor in gymnastics then it couldn't change the order of 2 other competitors. The one you gave the score for might go above both those competitors, but it couldn't reverse their order. But I think the new way of scoring in ice skating does unfortunately have this property, so the one that's usually excluded, that it's possible for judges to give a score to one particular skater which will invert the relative order of 2 other skaters, because the mark somehow determines the total amount of marks that's being distributed between all the competitors, and if this skater is given a low score, it lowers the total amount of marks available and lowers the relative weight for different types of performance. There's an interesting and possibly sort of controversial situation I think brewing in certain sports that have this unusual type of subjective judging.

Well, I hope I've at least given you some glimpse as to how it's possible by rather simple mechanics, simple mathematics, simple probability theory, to shed some light on what's going on perhaps in the minds of people scoring events, and what type of considerations might enable you to understand why human sporting performance has the approximate level that it does. Rather simple mechanical arguments, simple use of energetics, simple use of the study of projectiles and so forth, enables us to understand, in a fairly simple way, why the levels of human performance are roughly as they are, and perhaps what you would have to do if you want to improve your high jump or your sprint start at the parents' race next summer!