Introduction

Definition: Judgements are what underlie decisions.

Decisions are constrained by time, knowledge and computational power – humans have a limited capacity information processing capability.

People make mistakes, get bored, become anxious and may not really know what they want to achieve.

Consequently, what we do is not the same as what we would agree we should do.

Theories of decision making

Two types:

normative – the 'ought' – the 'ideal' decision

descriptive – the 'is' – 'actual' decision making

Differences between the ideal and the actual do not imply fundamental faults with the way judgements and decisions are made – just as the Muller-Lyer illusion doesn't lead to the conclusion that human vision is fundamentally flawed.

Supporting decision making

Prescriptive approaches attempt to bridge the gap between normative and descriptive – an attempt to help people make better decisions (closer to the normative). An example – decision analysis. A number of techniques are used – e.g. decision trees, to enable complex decision to be broken down into more manageable parts, determine values and beliefs and then to reintegrate these components into a normative framework, leading to a decision.

Chapter 11 – Judgement and decision making

Normative theory of choice under risk

In many situations, uncertainty exists about whether a choice we make will be good or bad – **risk** is therefore part of life. e.g. the choice to risk surgery or not; investment decisions that may yield a high return or nothing vs a lower risk but smaller, safer return.

Such decisions can be analysed as gambles.

Subjective expected utility – SEU – a normative model of risky choice (**Savage**). e.g. can be used to help decide between accepting an invitation to a picnic vs working on an OU essay. All of the relevant elements need to be quantified and combined in order to make the optimal decision – likelihood of rain; will x be there; do I need a good grade for this essay …

Prescriptive application of normative theory: decision analysis

A technology based on SEU developed in the 1960s. Classic framework (**von Winterfeldt and Edwards**) – numerical probabilities are assigned to all the events in a decision tree. (Subjective) utilities are applied to each of the outcomes of future events.

So, if the choice is between the actions of going on a picnic or writing an essay and each action has three possible future events (rainy; cloudy; sunny), each with their own probability and subjective utility, then the SEU of either action can be calculated – i.e.

SEU(event) = Σ P(robabilities) x U(tilities)
= P(rainy picnic) x U (rainy picnic) +
P(cloudy picnic) x U (cloudy picnic)
P(sunny picnic) x U (sunny picnic)

The best decision (for that individual as U is subjective) is therefore the one with the highest SEU – even if the difference between choices is numerically very small.

Phillips – role of decision analysis has changed since the 1960s so that decision trees are used as an aid to thinking – not as providing solutions.

Requisite decision modelling (**Phillips**) theory – models of decisions need only be sufficient in structure and content to resolve the matter at hand. An iterative process is used to construct such a model; when no further intuitions emerge the model is said to be requisite. 'Soul searching' (**Watson and Buede**) is used to construct representations and find values – rather than purely mechanical means. However, the key strength of decision analysis is it makes explicit the bases of a decision.

Axioms underlying SEU theory

Five axioms:

1. Comparability – if there are two alternatives (A or B) then you must be able to say if A is preferable to B; B preferable to A; or that you are indifferent.

2. Transitivity – if A is preferred to B and B to C, then A must be preferred to C (choices can be ordered)

3. Dominance – an option is dominant if it is better in at least one way (and no worse in any other way) than all other choices. A *dominated* choice should never be preferred.

4. Independence – if an outcome exists that is not affected by your choice, then this outcome should not affect your choice

5. Invariance – different representations of the same choice problem should result in the same choices being made.

These may appear to be intuitive – but psychologists have shown systematic violations of each occur in people's actual choices – **Tversky and Kahneman**. Implies there is something wrong with the choices and/or something wrong with normative theory.

SEU does therefore not provide a good description of actual human choices!

Violations of transitivity

Tversky – preferences of college applicants on three dimensions (intelligence – increasing A->E; emotional stability and social facility decreasing E->A but far more rapidly than intelligence increases) – most people prefer A->B,B->C,C->D,D->E but also E->A, violating transitivity.

Reveals information about the choice mechanism used – **Tversky** offered two hypotheses about the process:

1. People compare alternatives on each dimension in turn, rather than evaluating each option on all dimensions before making a decision

2. People ignore dimensions on which the alternatives are rated similarly.

On the intelligence scale in **Tversky**'s example, intelligence is rated similarly between each pair, until E->A is compared when the difference is too big to ignore. This explanation assumes we have limited information processing capacity.

Intransitivity is not uniquely human – **Shafir** – bees violate transitivity in foraging for nectar between artificial flowers with different amounts of nectar in them at different levels of accessibility. Costs outweighed by the gain made in less information processing being required.

Violations of transitivity are usually acknowledged when explained and people are often prepared to change their choice to preserve it. Transitivity therefore appears to be normative, even if it is not descriptive of behaviour.

Violations of the independence axiom

More problematic – challenge normative & descriptive status of SEU.

Allais paradox:

Ticket #		1	2-11	12-100
Situation 1:	Choice A	$1M	$1M	$1M
	Choice B	$0	$5M	$1M
Situation 2:	Choice C	$1M	$1M	$0
	Choice D	$0	$5M	$0

Given situation 1, people prefer A to B; situation 2 people prefer D to C.

Allais argued that intuition should overrule the independence axiom – i.e. the normative SEU theory is not valid. **Savage** argued the intuitions underlying the choices were wrong and that SEU is normatively correct.

Slovic and Tversky – found people don't always change their choice even after being made aware of the independence axiom. If they do alter there choice, it may be because they are intimidated into feeling foolish if they don't.

The debate between **Savage** and **Allais** shows that the normative status of SEU is more of an article of faith than being a demonstrable truth – it is a principle to live by.

Findings from behavioural decision research

Violations of the axioms of SEU imply it does not provide a valid description of human decision making.

There is also empirical evidence that SEU does not predict human decisions either.

Edwards - Given two gambles of the same expected utility, people prefer a long shot of winning a lot (provided they didn't lose very much) to a good chance of winning a small amount. Gambles which have even a low probability of losing a lot are avoided. **Edwards** therefore concluded SEU is not a good guide to a choice between gambles. **Lichtenstein et al** – expected value is irrelevant even when fully explained and understood by participants.

The preference reversal phenomenon

Slovic and Lichtenstein – choices between pairs of gambles are strongly influence by the probability of a win or loss, however, when asked how much they would pay to be able to take a particular gamble, the amount that could be won/lost is the assessment used.

=> if there is a different basis for choosing and valuing gambles, it ought to be possible to show people prefer gamble A to B but would pay more to take B than A. Evidence backs this up – including a study in a real casino.

e.g. p bet 11/12 chance of winning 12
1/12 chance of winning 24

$ bet 2/12 of winning 79
10/12 of losing 5

Of those choosing to take the p bet, 87% gave the $ bet a higher selling price (c.f. 88% overall). This is not rational!!!

A major threat to the normative status of SEU. – Even sceptical economists have been able to replicate the results controlling for all the confounding variables they could think of (**Grether and Plott**).

Causes of anomalies in choice

Slovic – summarised the evidence in favour of a 'scale compatibility' hypothesis - how much weight someone gives an attribute in a judgement/choice depends on how compatible that attribute is with the response mode.

Results of gambles are monetary – so therefore people

find it easier to assess the value of a gamble than say which they prefer – true even if there is no risk involved (**Tversky et al**).

The prominence effect

Tversky et al – the importance (prominence) of an attribute is weighed more heavily in a choice than in a matching task.

Currently 600 casualties.

Choice task – Option A – 570 casualties, costs $12M
Option B – 500 casualties, costs $55M

B is preferred over A. But when a matching task is used (same information, but participants asked how much they would pay for option B), typical value is < $55M (only above in 4% of respondents).

Tversky argues choice => qualitative reasoning (lives saved more important than money); matching => quantitative reasoning – task cannot be performed without paying attention to both attributes and their relative importance.

Choosing and rejecting options

Shafir – choosing 1of2 items != rejecting 1of 2 items. People focus on +ve characteristics when trying to select an option and –ve ones when trying to reject an option. People look for reasons for their decisions rather than rank order options – as SEU would imply.

Led to **Shafir**'s **reason-based theory of choice** – reasons for choosing have more influence when we select; reasons for rejecting are more influential when we reject.

Conflict in choices – **Schelling** – encyclopaedia purchase put off because he was unable to choose between the two – yet if only either had been available he would have bought. A clear reason to select either led to no selection.

The evaluability principle

Hsee – ***evaluability*** – a kind of preference reversal when items are evaluated separately or jointly.

e.g. shop for pianos – may compare several pianos in a shop. If at an auction, there may only be one piano to consider. Some attributes are difficult to evaluate when items are in isolation; some are difficult to evaluate when items are being compared to similar ones.

e.g. Two dictionaries differing in number of words and condition of cover – the one with more words and damaged cover is worth more when compared to the one with fewer words and perfect cover. If presented apart, the one with fewer words is worth more.

Hard to evaluate a dictionary with 10,000 words in it by itself, so focus is on things that are easier to understand – e.g. the condition of the cover when presented alone. If evaluated together, it is 'obvious' a dictionary with 20,000 words is better than one with 10,000.

Prospect theory

Anomalies in choice imply we don't have a set of stable values (preferences) we apply when we choose.

Decisions change die to the demands of the decision task and nature/context of information presented.

If different ways of obtaining preferences result in different choices, it begs the question how preferences can be defined and how they can be measured.

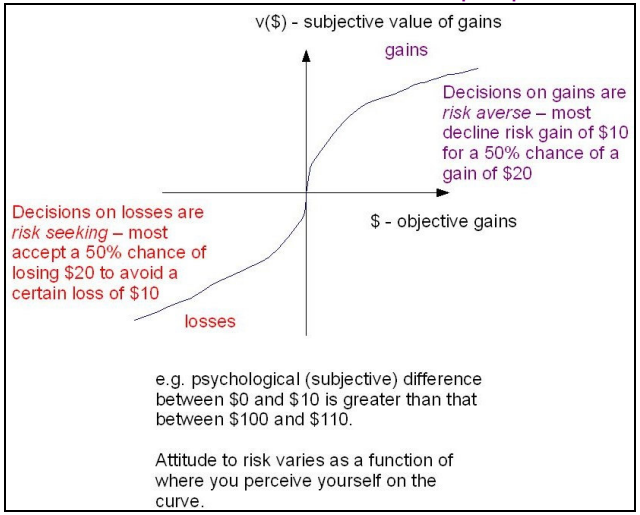
Kahneman and Tversky – descriptive model for decision making under risk – **prospect theory**.

Difference with SEU – does not define ideal choices; is descriptive, not normative … but is an adapted version of SEU.

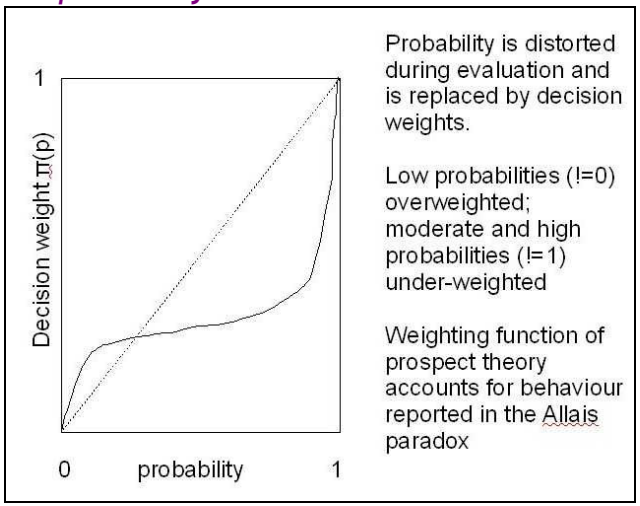
Identifies two phases in the act of making a choice:

1. Editing phase – decision problem is represented; 'negligible' components discarded; reference point used to enable outcomes of decisions to be determined as gains or losses.

2. Second phase – attitudes to risk involving gains and losses used to evaluate the identified prospects.



Prospect theory and loss aversion



Samuelson – colleague declined a $200 win chance vs $100 loss chance; but would play the bet if it was done 100 times. SEU says if you decline one bet, you should decline all that are the same.

Framing effects

This phenomenon is not predicted by SEU but is by prospect theory. **Kahneman and Tversky** – two identical programmes, but one framed in terms of lives saved and the other in terms of lives lost produces the opposite result – i.e. 200 saved preferred 72-28 to 1/3rd probability 600 saved vs 2/3rds probability no-one saved; yet 400 die not preferred 22-78 to 1/3rd chance no-one dies vs 2/3rds chance 600 die.

Relative overweighting of certainty also contributes to the relative attraction of '200 saved' option. In the 'loss' domain, no-one has yet died, so as 600 deaths not substantially worse than 400 it is preferred as it offers the chance that no-one dies.

Judgement under uncertainty

Normative theory requires value and likelihood to be used as inputs – therefore, being able evaluate likelihood is crucial for decision making where there is uncertainty.

Judging probabilities and Bayes' Theorem

Edwards et al – ***book bag*** and ***poker chip paradigms*** used to understand how good people are at judging probabilities. Bayes' Theorum provides an objective measure of probability – it can therefore be used to evaluate how good human estimation of probability on tasks such as determining which bag has most red or blue poker chips. It states that the odds of a hypothesis being right when new information is provided (posterior odds) is the product of the prior odds of the hypothesis being right (before new information is provided) x the likelihood ratio (ratio of probabilities that given the information we have the hypotheses is correct or incorrect.

Such studies allow the comparison of subjective probabilities with objective ones as calculated by Bayes' Theorum.

However – such a method does underplay the importance of subjective probabilities. It makes the assumption they should be the same for all participants provided with the same evidence.

Does Bayes' Theorum describe human judgement?

Book bag/poker chip experiments according to **Edwards et al** show:

1. Participants revise their probability judgements when new evidence is presented, but do not revise them sufficiently – **conservatism.**

2. Suggested people are good at assessing probabilities but are poor at combining them => should be possible to support judgements by taking such assessments and combining them using Bayes' Theorum to help people make better judgements. (If the opposite is true, there is no point in helping people combine probabilities more effectively – GIGO!)

Little research done into conservatism – as work on **heuristics** and **biases** (**Kahneman and Tversky**) overtook it.

Criticisms of **Edwards et al**:

1. Variability in conservatism is observed – for example, diagnosticity of data is important. If the poker chips are split 51/49 in colour (rather than 70/30) then 2 consecutive draws of the same colour is not very diagnostic. **Phillips and Edwards** – experiments showed the more diagnostic the information, the more conservatism occurs. Weakly diagnostic data results in human probability revisions being too extreme.

2. Bayes' Theorum is such that presenting all the information at once or one piece at a time should have no difference. **Peterson et al** found presenting one item and gaining revisions after each item produces less conservatism than providing all the information at once. Described as an inertia effect by **Pitz et al**.

3. **DuCharme and Peterson** – most experiments are not like the real world as they are discrete possibilities rather than on a continuum. Used a task that randomly sampled heights from male/female populations and participants asked to say which population was being sampled. Conservatism reduces by 50% - probably due to participants' familiarity with the height distributions underlying the task.

4. **Winkler and Murphy** – standard task (bag and chips) can be criticised as being different to the real world as:

(i) Pieces of evidence presented are conditionally independent – drawing one red chip from a bag and replacing it does not affect the chance of drawing another red chip. An aircraft flying off course and not responding to radio are not independent – both can be caused by equipment failure. After observing one thing, we should be less influenced by the other.

Participants may therefore have been using this expectation during the standard task – so an explanation for conservatism is that participants behave as they would in real world situations.

(ii) Fixed contents in the bags – but in reality, hypotheses are not constant.

(iii) Real information is often unreliable; so less diagnostic than the colours of poker chips – again, an argument for the conservatism seen.

(iv) Experiments usually offer very diagnostic information – in reality, weakly diagnostic information is the norm – so may result in conservatism in the lab.

'Conservatism may be an artefact caused by dissimilarities between the lab and the real world'

Heuristics and biases

Kahneman and Tversky's work show many errors in human judgement. They argued the cause was the operation of a number of mental heuristics. They argue the existence of these shows that judgment processes are not normative.

Limited mental processing capacity requires us to use simplification strategies – so that complex judgements can be made by humans – **Kahneman et al**.

The 'representativeness' heuristic

Used to work how likely an event is a category member by comparing its typicality to the category (c.f. similarity-based categorisation in ch.5)

E.g. a 'typical' librarian is used to judge if an individual is a librarian – but this neglects the **base rate** – i.e. how prevalent librarians are in the human population.

Kahneman and Tversky – experiments demonstrating base rate neglect.

Vignettes used. One group told the description was from a sample of 70 engineers and 30 lawyers; the other group was provided with the inverse frequency. Description had stereotypical 'engineer like' characteristics (i.e. male, no interest in politics/social issues, liked carpentry, maths puzzles). Mean estimates of the two groups very similar (50% vs 55%)

An uninformative description produced 50% vs 50% - so they concluded that if no specific information is given base rates are used; but even when information is uninformative, base rates are neglected.

Tversky and Kahneman – ***conjunction fallacy*** is explained by representativeness – studies that show probability of A or B exceeds the probability of A and B (logically, this cannot be true). Description provided of Linda judged to be more likely that she was a feminist **and** a bank clerk than just a bank clerk!

Judgements of representativeness therefore do not appear to respect the conjunction rule – even though probabilities do.

The 'availability' heuristic

Occurs when likelihood estimates are influenced by how easily other instances can be brought to mind. Frequent events easier to recall than less frequent ones so it can be a valid way of assessing likelihood.

However, emotionally salient / recent events also affect availability – e.g. perception of flying as risky increases after an air crash.

The 'anchor and adjust' heuristic

Random number (0-100) selected; participants asked to say if the % for a question is higher or lower than the random number and to say what % they thought the right answer was. Shows a correlation between entirely randomly generated numbers and a participant's belief as to the correct answer!

Evaluating the heuristics and biases account

1. It is possible to question if the **Kahneman and Tversky** demonstrations of biases would apply to experts in their domains – or just to the student samples used.

2. Tasks set to participants provide a misleading view of their competence

3. The standards used for the assessment of judgement are inappropriate.

Representativeness and base-rate neglect

Gigerenzer et al – just because the experimenter believes they've set a probability problem it doesn't mean the participants will see it in the same way. For example, they may have reasons not to accept the asserted base rate as their subjective prior probability. In **Kahneman and Tversky**'s experiments, participants were told the descriptions had been randomly sampled – but in reality, they had not been – they had been selected to be representative of the professions. It is reasonable for a participant to ignore the base rate offered if they suspected that this was the case!

Gigerenzer et al – participants sampled themselves by examining 10 descriptions marked 'lawyer' or 'engineer' (but could not see the mark.) Base-rate neglect was not seen. Where base rates were asserted, then base-rate neglect was replicated.

Kahneman and Tversky now argue base rates are underweighted – as **Gigerenzer et al** results still produce judgements that deviate from the Bayesian solution in the direction predicted by representativeness. **Koehler** – heuristic for representativeness is limited – as if judgements in these experiments reflect the use of base rates, it is difficult to account for the findings by assuming the operation of a simplifying representativeness heuristic.

Frequency and the conjunction fallacy

Tversky and Kahneman – violations of the conjunction rule largely disappear if participants requested to assess the relative frequency of events – rather than probability of a single event. e.g. Out of 100 men that had had a heart attack, participants asked to assess the number who were > 55y.o. and had had an attack. Only 25% of participants violated the rule. Results replicated frequently since.

Gigerenzer – we are naturally adapted to reason with frequencies as the fallacy is not observed under such conditions. Suggests difficulties people have with problems of probability can be diminished if they are framed in a way as to require frequency judgements for a class of events rather than the probability of a single instance. **Kahneman and Tversky** disagree – frequency format is providing a powerful cue to the

solution (set inclusion) and it is not the frequency information per-se that prompts people to alter their judgements.

Backed up by an expt presenting 1,000 'Lindas'. Asked on group to say how many are bank tellers; a second group how many were bank tellers and feminists; a third group made evaluations for both categories.

Those who evaluated both categories largely conform to the conjunction rule; but a between groups comparison of the other two shows bank teller and feminist higher than bank tellers.

Representativeness heuristic is therefore used to generate judgements, which are then edited if a cue is given that suggests class inclusion.

Overconfidence

If given a 50-50 chance on a question like is NY or Rome further south and asked to rate how confident they are their answer is correct, people tend to be overconfident – e.g. **McClelland and Bolger; Harvey**.

Has also been recorded in the judgement of experts. **Christensen-Szalanski and Bushyhead** – physicians diagnosing pneumonia give poorly calibrated probabilities and are overconfident. **Wagenaar and Keren** – observed overconfidence in lawyers' predictions of the outcome of court trials in which they represented one side.

This may be because feedback in both instances takes months and the ultimate truth may never be known.

Well-calibrated judgements found in tournament bridge players (**Keren**) – concerning the likelihood that their contracts would be made; Phillips – same for bookmakers forecasting horse racing results. Where feedback is this instant and explicit experts are unable to be insensitive to it – little scope for denial or neglecting the difference between forecast and outcome.

Gigerenzer – overconfidence observed in experimental setting as quiz questions used with misleading answers. When asked to select the larger city in randomly selected pairs of German cities, overconfidence disappears.

Erev et al – overconfidence may be partly explained by a random component of judgement as confidence and accuracy are not perfectly correlated. Will therefore result in a regression effect.

Overconfidence could be an artefact of inappropriate test items and regression effects – however, **Budescu et al** found 87% of their study that compensated for these effects were biased towards overconfidence.

Yet **Juslin et al** – meta-analysis of 35 studies with items randomly selected from a defined domain and 95 studies where items were selected by experimenters. Found overconfidence was close to 0 for randomly selected items – so suggesting overconfidence is one created inadvertently by researchers.

Fast and frugal theories of decision making

Testing the usefulness of heuristics outside of the laboratory – "in the wild". **Gigerenzer and Goldstein** – produced measures of efficacy of simple mental strategies for judgement by measuring the number of correct inferences made. Termed **fast and frugal**.

Frugal = one piece of information used to make decisions
Fast = no integration of different pieces of information as prescribed by SEU or Bayes' Theorum

[Human decision making research often focuses on its 'non-optimal' nature – but **Simon** suggested 'satisficing' methods are required as humans have bounded rationality – imposed by the limited capacity of our brain.]

'Take the best' heuristic tested on pairs of German cities with a set of properties on which to make the decision as to which was the larger. Participants work through a set of cues in order of predictive validity until one found that discriminated.

'Take the best' does as well as, if not better than, many other algorithms. It is faster than strategies requiring integration. Therefore, obeying normative rules may therefore be sufficient for good judgement, but it is not necessary.

Other examples – 100% Germans identify San Diego as being larger than San Antonio but only 62% of Americans do. Explained by the use of a recognition heuristic – Germans had heard of SD but not SA, so therefore SD must be larger.

Conclusions

Psychological studies show decision making cannot be described in idealised mathematical ways assumed by economists, for example.

Rejection of normative theory as a model for decision making does not mean it is poor or irrational – simply that the computational requirements of normative theory is beyond the capacity of human brains.