Resource Management:
INSTITUTIONS AND INSTITUTIONAL DESIGN
SOS3508
Erling Berge

Explanations and mechanisms,
The Mind,
Action,
Lessons from the Natural Sciences

NTNU, Trondheim
Fall 2010

Literature

- Elster, Jon. 2007. *Explaining Social Behaviour: More Nuts and Bolts for the Social Sciences*. Cambridge: Cambridge University Press.
  - Ch 1-17

Actors and Systems

Actor models must generate
- Distributions of actors
- Variation in motivations and actions

Systems models must comprise
- Institutions and actors (organisations)
- Mechanisms generating meanings and identities
- Levels of analysis
- Dynamics of change
- Mechanisms of non-ergodicity
  - A non-ergodic system do not repeat itself

Explaining social behaviour

Such as
- Why do preferences sometimes change through the sheer passage of time?
- Why are people unwilling to break self-imposed rules even when it makes little sense to follow them?
- Why do military commanders sometimes burn their bridges (or their ships)?

The aim is to inculcate scepticism to
- Functional explanations, and to
- Some kinds of rational choice explanations

Explanations (1)

- Agents perform actions
- Agents may be rational or irrational
  - If agents are irrational, one must take care in explaining the mechanisms involved in actions
  - If agents are rational, actions rely on choices informed by reasons, motives, desires, and/ or interests

Explanations (2)

- Explanation of actions is causal
  - Intentional explanations (including rational choice of means to obtain ends)
  - Explanations by consequences, rare in social science
  - Explanations by laws, strong laws rare in social science
    - Deterministic
    - Statistical explanations rely on intuitions about mechanisms
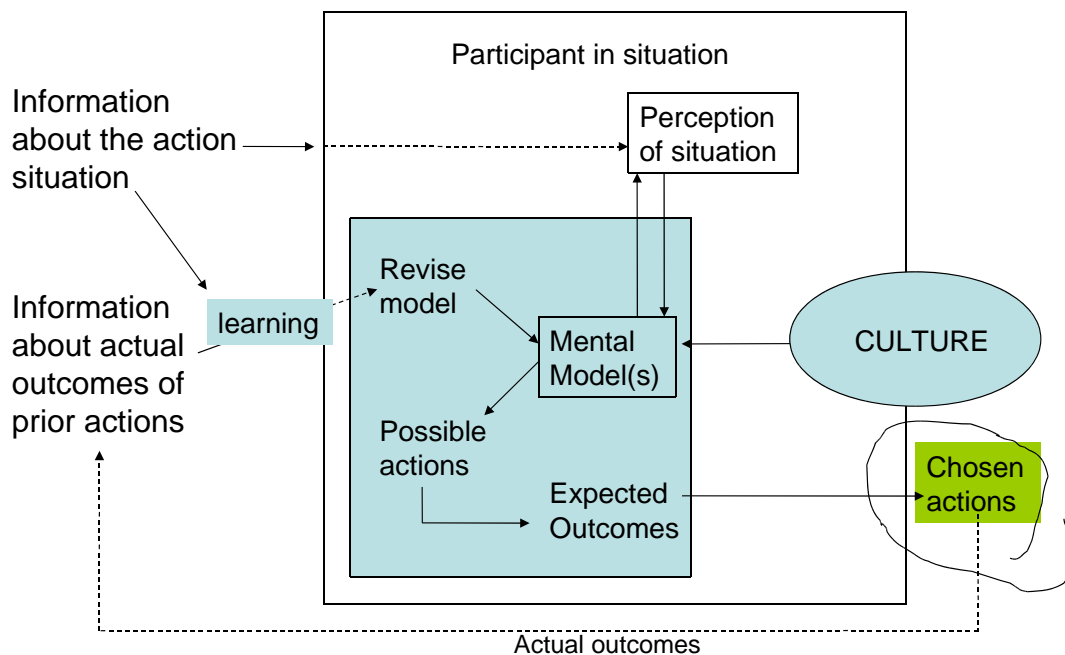
Illustrating causal explanations
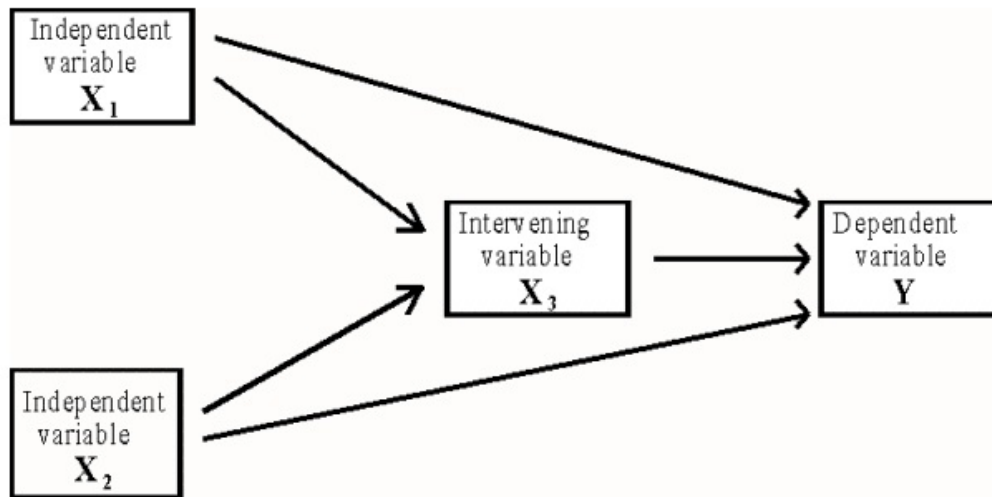
Information, action-outcome linkages, internal mental models
(adapted from Figure 4.1 page 105 in Ostrom 2005)

Explanations in general

- Explaining events by prior events
  - Give an account of why explanandum (event) happened by pointing to an earlier event as cause
  - Events vs facts {events – events, facts – facts, facts – events, events – facts}
  - Explaining differences and variation rather than "brute events" (absolute sizes or numbers)
  - Explaining variety
  - Explaining non-events (Kitty Genovese)

Ideal principles

- Event – event explanations
- Methodological individualism

- In practice
  - We use facts as explanandum and as explanans
  - We explain non-events and non-facts
  - We explain differences and variation rather than sizes and variety
  - We talk about families and communities and nations as if they were similar to individuals

Structure of explanations

- Using beneficial consequences as explanans is difficult. It requires that the loop linking consequences to event is established
- The usual structure of explanations
  1. Theory
  2. Hypothesis
  3. Derive consequences and rival explanations
  4. Refute rival consequences
  5. Strengthen the explanation by deducing novel facts and demonstrating their existence

Good explanations

- Support from below if more than the explanandum is observed and explained
- Support from above if the hypothesis is derived from a more general theory
- Lateral support if alternative explanations can be refuted (be the devil's advocate!)

Explanations are not

- True causal statements
- Correlations
- Necessitation
- Storytelling
- Statistical generalisations
- Answers to "why" questions
- Predictions


Strong and weak Laws

- In social science there are few if any good examples of strong causal laws
- The law of the relationship between income and demand is a weak law, it tells about the direction of a change, nothing about the magnitude

Mechanisms instead of laws

- Ideally we want to specify a causal chain
- Practically speaking we look for mechanisms:
  - Mechanisms are frequently occurring and easily recognizable causal patterns that are triggered under generally unknown conditions or with indeterminate consequences.
  - They allow us to explain but not predict
  - Mechanisms involving aggregates points to a need for methodological individualism

Molecular mechanisms

- Elementary psychic reactions as atomic mechanisms to build molecular mechanisms
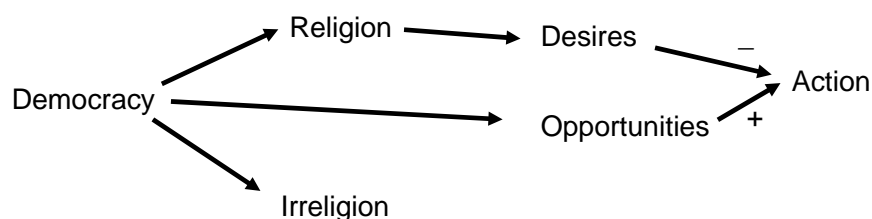
Case: impact of democracy on dangerous and licentious behaviour (from Tocqueville)

Mechanisms and laws

- If we are able to specify the triggering conditions of a mechanism we may be able to specify a law, usually a weak one
- Example: Groups asked to rank music records
  - Group 1 rank 4 records, reward get one picked at random
  - Group 2 rank 4 records, reward choose one yourself
  - Next day redo it based on the unavailability of the one ranked as no 3
  - Result: G1 displays "sour grape" reaction; G2 displays "forbidden fruit" reaction
  - The control group were not told it was unavailable and did not change its ranking
  - Triggering: G2's freedom of action encountered an impediment that G1 did not
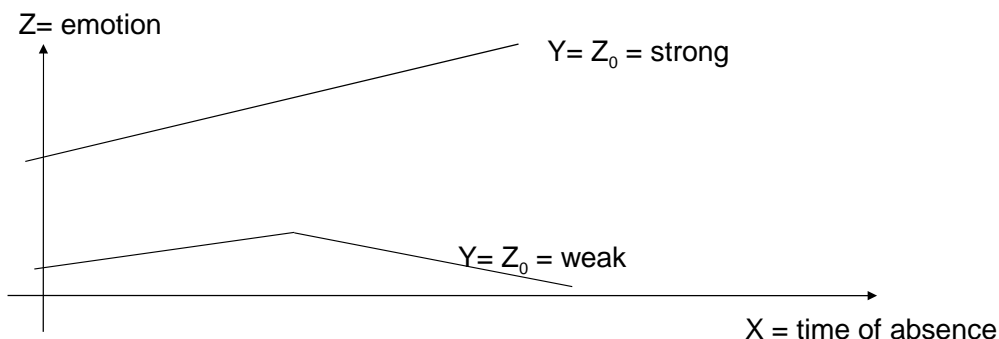
Interaction among causes

- Default assumption: additive effects
- Interactions: low values of z at time 0 may give decreasing value of y while high values of z at time 0 might give increasing levels of y as x (=time) increase

Z= emotion

$Y = Z_0$ = strong

$Y = Z_0$ = weak

X = time of absence

Interpretation (Verstehen)

- Interpretation is one kind of explanation
  - To decide among conflicting interpretations interpretative hunches or hypotheses needs to confront experience including novel facts
- Rationality and intelligibility (interpreting action)
  - What are the beliefs and desires (motivations)? Are they intelligible?
- Also irrational behaviour may be intelligible

Understand irrational behaviour

- If rational decision making is truncated for example by strong emotions
- If rational decision making is short-circuited by the agents desires
- If rational decision making is confounded by inconsistencies in the beliefs and desires of the agent

Unintelligible are actions based on phobias and obsessions, actions like anorexia, self-mutilation

Understanding Civil Wars

- Why are young Palestinians willing to give their lives in suicide missions?
- In general obtain or defend the homeland
  - Poverty and illiteracy?
  - Relative deprivation?
  - Comparisons and interactions inducing feeling of inferiority and humiliation
  - Induced religious and ideological fervour at the right moment for triggering the bomb

A hermeneutic dilemma

- How do we establish the desires and beliefs motivating action?
  - Oral and written professions by the persons?
    - Public or private context?
    - What is the cultural hierarchy of motives?
    - Self-serving bias in professed motives
  - Objective interests
    - Religion, power, and money may be involved
    - Investigate actual consequences
  - Look for sources least likely to be biased: letters, diaries, conversations, drafts, etc.
  - Asking questions in a way that creates an artificial "veil of ignorance" to bolster the promise of anonymity
  - Do agents put their money where their moth is?
- Sometimes 'always telling the truth' is the greatest subtlety

A short summary of institutions seen as

- Social facts by agreement (Searle)
- Thought worlds/ subjective models (Douglas)
- Rules of the economic game (North) shaped by
  - Transaction and information costs
  - Subjective preferences and learning
  - Increasing returns and political processes

Institutions are

Humanly devised rules with some

Humans mandated to monitor and sanction rules

- Created to aide in collective actions problems to safeguard life and livelihoods
  - Avoid conflicts, create justice
  - Allocate legitimate benefits and duties, profits and costs
  - Economize on transaction costs
- **Not** created to achieve efficiency or optimise economic performance(of the neo-classical model)

Property rights institution

Tells that some person(s) have legitimate
- Rights and duties to be exercised in relation to
- Particular goods and services subject to possible
- Limitations on times and durations,
- Limitations of technology, and
- Limitations on organisation of exploitation

The construction of social institutions

For example
- Property rights regimes
  - Public property
  - Common property
  - Private property
- Regulations regimes
  - Governing externalities
  - Protecting unitary/ universal values
- Bureaucracies

The construction of social institutions

Informal institutions

- Conventions
- Customs
- Values, Preferences
- Norms, Standards of conduct
- Beliefs, Ideologies, Morals

The dynamic of Institutions and organisations

- Rules of the game   (the law)
- Guardians of the rules (the judge)
- Players   (organisations)
  - Owners,
  - Local users,
  - Workers,
  - Professional managers, and
  - Firms of resource industries

Strategies of the players

Our theory requires by assumption that players

- Optimise their returns from resource use activities by conforming to and **exploiting the existing institutional environment**, or to
- **Change the resource policy** in a desired direction if the expected outcome of a political effort is seen as cost effective.

Lock-in of institutions and organisations

- Mutual interdependence institution-organisations
- Institutional changes by public initiative or revolution creates counter-forces
- Economic performance is **PATH DEPENDENT**
- Change occurs at the margins

The Mind

- Insights from
  - Introspection
  - Folk psychology
  - Psychology
  - Behavioural economics
- Needed for understanding
  - Praise
  - Blame
  - Punishment

Intentions, beliefs, desires

- Current models imputing mental states to others are fragile
  - To maximise expected utility we need to know
    - Values attached to each possible outcome
    - Probability assigned to the occurrence of the outcome
    - Answers will depend on how questions are asked
  - Asking about preferences also reveals context dependence
  - Trade-offs among values are highly unstable
  - Statements about beliefs and mental states are often questionable
  - Beliefs in afterlife and martyrdom
  - Experiences of quasi-emotions (no implication for action)
  - Powers of autosuggestion

Most mental states are unstable

- What about unconscious mental states?
  - If causally effective they can be identified by their effects
  - Unconscious prejudice
  - Unconscious emotions show up in body language probably
  - Self-deception is probably not unconscious

Motivations

- From visceral to rational
  - Acting on emotions
    - Visceral fear vs prudential fear
  - Acting on good reasons
    - Cost-benefit considerations
- Interest, Reason, and Passion
  - Interest is the pursuit of personal advantage
  - Reason is about impartiality and the public good
  - Passion is about emotions and the visceral urges, maybe also forms of madness

Freud

- Id, Ego, Superego as subsystems of the mind
  - Id is similar to passions and corresponds to the pleasure principle
  - Ego is the active I and corresponds to the reality principle
  - Superego is the conscience and corresponds to impulse control
- As ego navigates the external world it also has to fight a two-front war against the impulses from the id (pleasure principle) and the punitively severe impulse control exercised by the superego (conscience)

Taking account of consequences

- Consequentialist motivations
  - The consequences following the actions are their sole motivation
  - E.g.: Most economic behaviour
- Non-consequentialist motivations
  - Consequences are irrelevant, the motivation is the action itself
  - E.g.: conscientious objectionist to military service
  - Kantianism: always do what will be best if all did the same
  - Sanctioning of social norms

Four approaches to motivation

	Biology/emotion	Institutions?	Cognition/ rationality
1	Visceral Pleasure principle		Rational Rationality principle
2	Passions	Reason	Interest
3	Id	Superego	Ego
4		Non-consequentialism	Consequentialism

Wants and wishes

- Wanting and wishing
  - Wants require actions by the one with wants
  - Wishes requires some state of the world to obtain no matter how
- States that are essential by-products
  - Some wants cannot be obtained by direct action, or only accidentally so
  - E.g.: desires to forget, to believe, …
- Push or pull?
  - Rational: comparisons of two situations
  - Emotional: running away from something (fear, shame) or moving towards goals of glory or honour

Conflicts of motives

- Winner takes all (strongest motive wins)
- If compromise is possible the stronger motive has strongest impact
  - What determine strength here?
- Metamotives
  - Cultural hierarchy of values determine strength of motive
  - Passion and interest will often defer to reason, justice, and fairness
  - Sometimes this is after the fact dressing
- Cognitive dissonance theory suggest small differences in motives will be transformed into large
  - This may result in a kind of path dependence in motivations
- Alliances of motives will determine strength of motivation

Self-interest and altruism

- Altruistic motivation is the desire to enhance the welfare of others even at a net welfare loss to oneself
- Altruistic behaviour is no good indicator of altruistic motives
  - Reason or passion may mimic altruism
  - Love of virtue is different from virtue
- Approbativeness (wish to be thought well about by other people) will often have to stand in for altruism
- Shamefulness (wish to not be thought badly about by others) may also do so
- Social norms encouraging may affect behaviour but to obtain approbation behaviour needs to be supererogatory: in excess of the norm

Transmutation and reciprocity

- Transmutation
  - From interest to reason because of self-love (love of esteem and self-esteem)
  - Finding plausible reasons for self-interested behaviour is easy
- Reciprocity
  - From dyadic to generalised reciprocity
  - Applies both positive and negative (punishment)
  - Applies to building reputations
  - Applies in experiments of trust game

Norms

- Moral norms (proactive)
  - Help others in distress, equal sharing, the everyday Kantian norm
  - Unconditional, but affected by what others do
- Social norms
  - Etiquette, revenge, regulating the use of money,
  - Conditional, Triggered by presence and behaviour of other people
- Quasi-moral norms (reactive)
  - Reciprocity, conditional cooperation,
  - Conditional, Triggered by presence and behaviour of other people
- Identifying altruistic motives
  - Action needs to be proactive, not reactive
  - Action is anonymous
- Imputing motives is often tainted by malice
  - Hermeneutic dilemma
  - Conspiracies occurs
  - Public figures do (sometimes) act on good reasons

Myopia and foresight

- Myopia: Scanning the nearby options to choose the one with largest immediate gratification compared to status quo
  - This leads to a local maximum
- Humans can do better by planning ahead
  - Deferred gratification
  - Choosing the fast road rather than the short
  - Time discounting, high rate means future rewards have low value
    - Involves both cognitive and motivational elements

Pure time discounting

- Usually modelled as exponential discounting: present value of one unit t periods into the future is k^t where k<1 is the per period discount factor. This allows consistent planning
- Empirically this is questionable. To allow for everyday changes of mind and many other phenomena hyperbolic discounting is used. For example with a discount factor of $1/(1+kt)$ for the present value of one unit t periods into the future
- Choosing the highest present value in a choice between 10 at t=5 and 30 at t=10 will force a switch some time between t=3 and t=4 (figure 6.3)

Pascal's wager

- You should bet on God's existence since even the smallest chance of being right will give eternal bliss thus trumping all earthly pleasure
- Is the present value of eternal bliss finite or infinite?
  - Exponential discounting gives a finite value
    - Hedonistic life style will be preferred
  - Hyperbolic discounting gives an infinite value
    - Even a small probability of being right will suggest a godly life
    - But at any moment in time hyperbolic discounting will indicate that pleasure today is to be preferred
    - "Give me chastity and continence, but no yet." (St.Augustine)

Weakness of will

- A weak-willed (akratic) person
  - Has a reason to do X
  - Has a reason to do Y
  - In the person's own judgement the reason for doing X is weightier than the reason to do Y
  - The person does Y
- What is the causal mechanism?
  - Preference reversal
    - Temporal proximity (hyperbolic discounting)
    - Spatial proximity (cue dependence)
    - Passions
- These ideas may also apply to temporary changes in beliefs

Beliefs

- Belief as certainty: absolute certainty of being right
- Belief as knowing: justified true belief
- Belief as risk assessment: knowing that one may be wrong with some probability
- Belief as resolution to fundamental uncertainty (at most a ranking of probabilities)
- Belief as resolution to fundamental ignorance

There is a strong tendency for intolerance of uncertainty and ignorance flowing from both pride and a universal propensity to impute meaning, pattern, and agency to the world

  - Ignorance and motivation leading to certainty begets errors
  - Increasing levels of knowledge may make us more confident than data warrants

Assessing probabilities

- Observing frequencies
  - Do not ignore absolute sizes
  - Think about selection biases
  - Heuristics
    - Availability of an event for the mind
    - Representativeness of a small sequence of events
- Subjective evaluations of available information
  - Few possess good judgement,
    - Also few experts
  - Most should learn to distrust intuitions

Magical thinking

- "Cold" unmotivated mistakes (believing in causal links that cannot exist)
- "Hot" motivated mistakes (unduly influenced by desires)
  - Wanting a belief because it is believed to have good consequences. Getting it is difficult
- Beliefs from content
  - Rationalisation (behaviour first then belief)
  - Wishful thinking (belief first then behaviour)
  - Self-deception (belief first then behaviour)

Emotions

- Source of happiness (love) and misery (shame)
- Impact on action
- Impact on belief
- No agreed definition
- Will focus on
  - Cognitive antecedents (triggering by beliefs, often new)
  - Physiological arousal
  - Physiological expression
  - Action tendencies (there is at least a form of incipient behaviour)
  - Intentional objects (emotion is about something)
  - Valence (strength of happiness or misery)
  - There may exist qualitative differences to feelings of emotions like shame vs guilt not reducible to their valence

What emotions are there?

- Focus on those important for social life
  - Evaluative emotions
    - Shame
    - Contempt and hatred
    - Guilt
    - Anger
    - Cartesian indignation
    - Pridefulness (own character)
    - Liking (another's character)
    - Pride (own actions)
    - Gratitude
    - Admiration

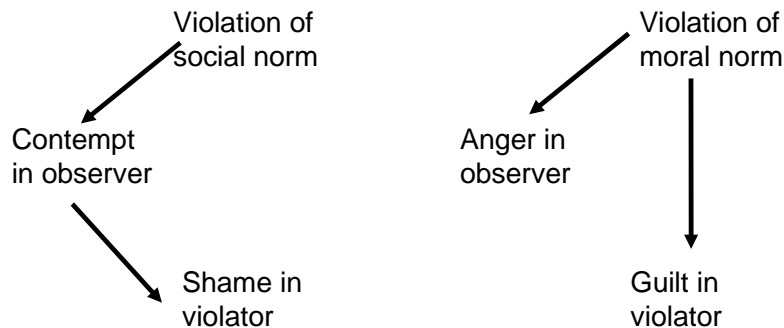# More emotions

  - From a state of affairs
    - Envy (another's deserved good)
    - Aristotelian indignation (another's undeserved good)
    - Sympathy (another's deserved good)
    - Pity (another's undeserved bad)
    - Malice (another's undeserved bad)
    - Gloating (another's deserved bad)
  - Joy and grief from what have or will happen to oneself (known with certainty)
  - Hope, fear, love, jealousy based on beliefs that have uncertainty as a component
  - Disappointments, regrets, [elation, rejoicing] (relief) caused by some kind of counterfactual reasoning

Happiness

- A gross national happiness product?
  - Economy as a means to achieve happiness
  - Happiness is subjective
  - Will emotional ups and downs go together?
- Emotion generated action tendency as a temporary preference
- Action is then wanted to occur sooner rather than later
  - Impatience: reward sooner rather than later
  - Urgency: action sooner rather than later

Table 8.1 Emotion	Action tendency
Anger or Cartesian indignation	Cause the object of the emotion to suffer
Hatred	Cause the object of the hatred to cease to exist
Contempt	Ostracize; avoid
Shame	"Sink through the floor"; run away; commit suicide
Guilt	Confess; make repairs; hurt oneself
Envy	Destroy the envied object or its posessor
Fear	Flight; fight
Love	Approach and touch the other; help the other; please the other
Pity	Console of alleviate the distress of the other
Gratitude	Help the other

Restoring the moral balance of the universe

Violation of social norm → Contempt in observer → Shame in violator

Violation of moral norm → Anger in observer

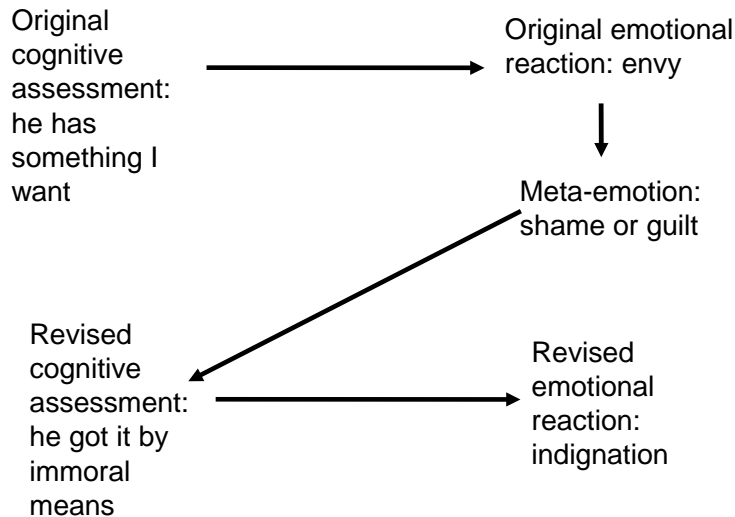Violation of moral norm → Guilt in violator

- Duration of emotions (short half-life?)
  - Situation dependent, memory dependent
    - But decay may be difficult to anticipate
  - Some are enduring

Emotion and belief

- Direct effects create biased belief
  - Love bias the perception
  - We believe what we fear
- Indirect effects create low-quality beliefs by acting on prior information gathering
- The two often go together reinforcing each other

Transmutation:
cognitive rewriting may be necessary

Original cognitive assessment: he has something I want → Original emotional reaction: envy

↓

Meta-emotion: shame or guilt

Revised cognitive assessment: he got it by immoral means → Revised emotional reaction: indignation

Culture

- Are there universal emotions? Yes, probably all.
- Happiness, surprise, fear, sadness, disgust, and anger have facial expressions recognized across cultures
- The way people think about emotions may be culture specific even if emotions themselves are not
- However, the emotional concepts and ways of thinking about them may also affect behavioural manifestations

Action

- Behaviour (including action (intentional), decision, choice)
  - Not all decisions lead to action
- In choices and decisions we have rationality as our ideal
- This leads to rational choice theory, but not necessarily to more rational actions …

# Desires and opportunities

- Doing one's best
  - Desires define "best"
  - The opportunities frames the choices and give the means (or the beliefs about these do)
  - What is the opportunity set?
  - How can we find the best action in this set?
 - Are opportunities more important than desires?

Opportunities

- They are more easily observed
- Differences in opportunities create variation (indifference curves and budget constraints)
- Are there irresistible desires?
- Opportunities are more easily manipulated than their desires (case: suicide)

Interactions: simultaneous impacts

Four cases

A: +desire +opportunity

B: -desire -opportunity

C: +desire -opportunity

D: -desire +opportunity
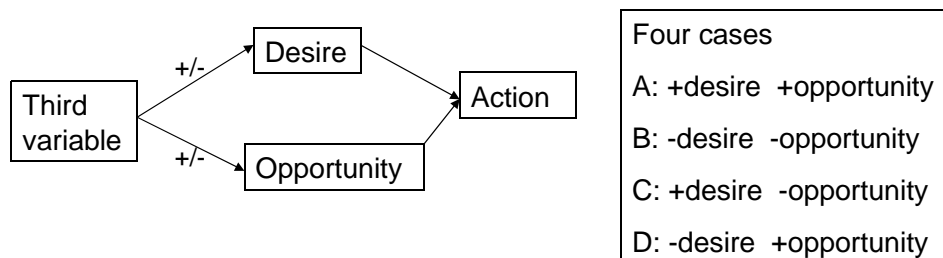
- Explaining action in case B
- Tocqueville:
  - Slavery is unprofitable compared to free labour (negative impact: slavery is not the best choice)
  - Slavery affects the owner's desire to work for his fortune (negative impact)

Interactions opportunity-desire

Two cases

E: opportunity⟶desire

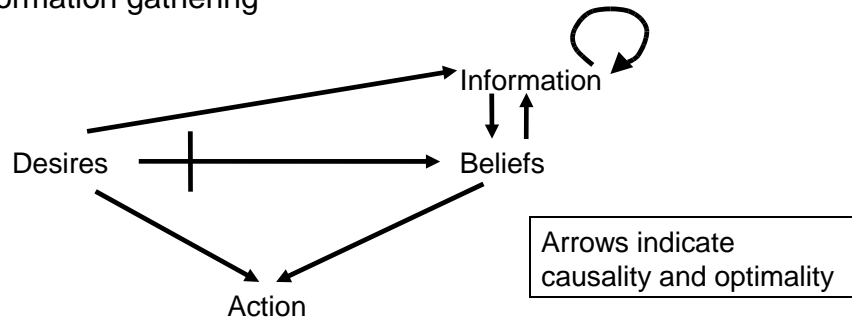F: desire⟶opportunity

Desire

Opportunity
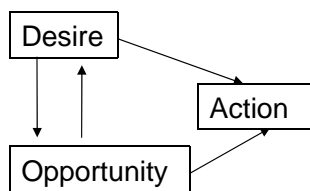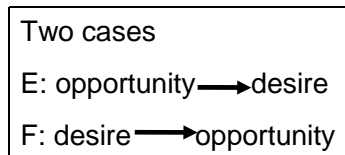
Action

- E: People end up desiring what the opportunities offer

- F: desiring states that are essential by-products may interfere with the opportunity to get them

Persons and situations

- Can character explain action?
  - Folk psychology assumes stable character and clustering of good traits
    - To some extent this is self-fulfilling
  - To a large extent it is demonstrably false
- The power of the situation
  - Very low consistency of behaviour across situations
  - Consistency of situations leads to disposition of character as explanation: the fundamental attribution error
- No common mechanism can be found in how situations affect behaviour, but both person and situation matter

The structure of rational action

- Assuming agents are rational
- Rational action requires
  - The action must be optimal given the beliefs
  - The beliefs must be as well supported as possible given the evidence
  - The evidence must result from an optimal investment in information gathering

Desires → Information

Information ↓↑ Beliefs

Desires ─┤─→ Beliefs

Desires → Action ← Beliefs

Arrows indicate causality and optimality

Weber was mistaken

- Weber: Departure from rational action is a sufficient indicator to identify irrationality

- But seemingly rational action is not sufficient to certify rationality

- Adaptive reactions (fear and flight) are not rational in the sense used here even if a rational decision making agent would come to the same conclusion

Preferences and ordinal utility

- The desire for the best or optimal outcome
  - Preferences define a rank order of alternative outcomes in terms of "betterness"
- Desires do not have to be "selfish"
  - Rationality does not mean egoism
- Desires do not have to be stable
- Assumptions about preferences (disregarding that the set of options be compact and closed)
  - Preferences have to be transitive
  - Preferences have to be complete
  - To get utilities preferences must be continuous
- Then choosing the best feasible option means maximising utility

Cardinal utility and risk

- Risky choice: action with more than one outcome
  - Choice: option with maximum expected utility (utility times probability)
  - Does not work with ordinal utilities
  - Needs "cardinal utility" (utility measured on an interval or ratio scale)
  - Cardinal utility functions are linear in probability
- Defining cardinal utility
  - A is best option, utility is set to 1
  - B is worst option, utility is set to 0
  - C is some option in between
  - There is a probability p(C) such that an agent is indifferent between getting C with certainty and a lottery getting A with probability p(C) and B with probability [1- p(C)]
  - The cardinal utility u(C) is defined as equal to p(C)

Risk aversion and decreasing marginal utility

- Risk aversion will conceptually be embedded in the utility function
- Risk aversion may sometimes be confounded with decreasing marginal utility
- Intrinsic utility (impossible to measure so far) is defined by the subjective intensity of enjoyment of some good (intensity of a preference)
- Decreasing marginal utility occurs as the intensity of enjoyment decrease with each additional unit of the good beyond some threshold
- One may also have increasing marginal utility
- Cardinal utility measures the combined effect of risk aversion and intrinsic utility

Rational beliefs. Optimal information gathering.

- Rational beliefs are those formed by processing available evidence by procedures that in the long run, and on average, are most likely to yield true beliefs
    - Bayesian learning (see p203-204)
- How much time and money should we allocate to acquire new information?
    - Depends on desires (wishful thinking is irrational)
    - Depends on prior beliefs and expected utilities of available options
        - If information costs are above possible gains in utility it is not rational to collect the information
    - Optimal search may depend on the information gathered

Indeterminacy

- Agents may be unable to identify the best element in the feasible set
  - Indifferent between two options
  - Incomplete preferences
  - Unable to determine optimal information gathering
  - Forming beliefs about actions of other agents involved in strategic interactions
    - Reward structure may prevent convergence of beliefs

Rationality is subjective

- Rationality cannot mean good consequences as determined by an external observer. That would be explanation by consequences
- Choices must be seen through the eyes of the agent
  - Choosing a "utility discounting pill" is irrational, the possible actions are already available
- Rational beliefs is different from true beliefs
  - Opportunity costs may discourage investment in information
  - Beliefs about direct costs of information may make investment irrational

Rationality and Behaviour

- Hyper-rationality: optimisation of decisions disregarding the costs of deciding
  - Costs of the means of deciding
  - Costs of side effects of deciding
  - Opportunity costs (value of goods forgone by spending time and resources on deciding)
- Canonical principles of rationality are frequently violated (18 examples discussed)
  - Case 1: Choose acting over non-acting if expected utility of acting is higher
    - Voting can be seen as violating this principle

Alternatives to rational choice theory

- Eleven mechanisms explaining deviations
  - **Loss aversion** (prospect theory)
  - Non-probabilistic weighting of outcomes (prospect theory, utility is non-linear in probability)
  - **Hyperbolic discounting**
  - Heuristics
  - Wishful thinking
  - Inability to project
  - The desire to act for a reason (not the same as acting according to reason)
  - Magical thinking
  - The categorical imperative
  - **Emotions**
  - Social norms

Emotions

Emotions may also be involved in

•Magical Thinking and in

•Social Norms
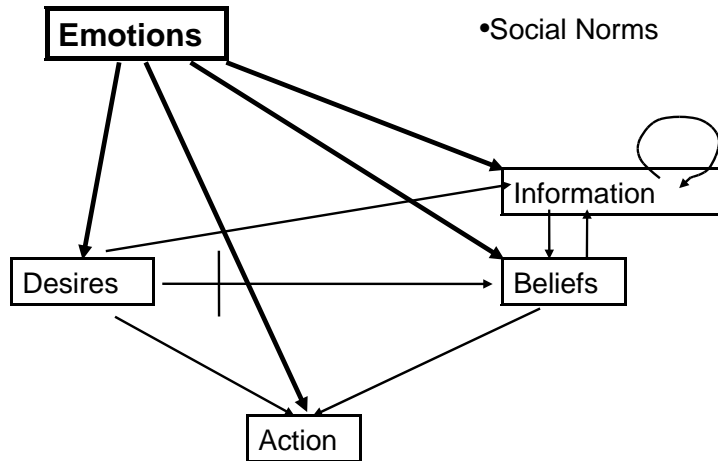
Responding to irrationality

- Second best rationality: taking precautions to avoid irrationalities such as hyperbolic discounting
  - Different from learning
  - Intrapsychic or Extrapsychic (precommitment)
- Agents knowing themselves to be subject to hyperbolic discounting are sophisticated
  - Bundling (or bunching) choices by reframing
  - Acting strategically against a future self
- Extrapsychic devices: precommitment
  - Eliminating a choice of early reward from the feasible set
  - Imposing a penalty on choice of early reward
  - Adding a premium on choice of early reward
  - Imposing a delay between choice and reward
  - Avoiding cues triggering preference reversal
- Precommitment may involve help from other individuals, organisations or public authorities
  - Sometimes this has the form of state paternalism not state assisted self-paternalism
- State constitutions as collective self-paternalism

Implications for textual interpretation

- Interpretations of art (e.g. literature) and explanation are closely related
- A successful work of art can be given a rational choice explanation
  - Explanation by consequences is not allowed
  - Internal (unfolding of plot) vs external explanations (author intentions and dramaturgical know-how): good works can be explained twice over: causally and teleologically
- Rationality requirements for authors
  - Intelligibility of acts and utterances of characters (absolute or relative; global or local)
  - Fullness and parsimony of story (all parts are necessary, non superfluous)
  - The work has to flow "downhill" (plot intelligibility in the minds of readers)

Lessons from the natural sciences

- Between scientific disciplines there are relations of reduction or analogy
  - Reductions may be an important engine for scientific progress (some current links between psychology and biology), or it may be premature, crude, or speculative
  - Analogies may provide valuable hypotheses but cannot provide evidence. Sometimes the analogy may provide thought worlds inimical to scientific progress (the organism analogy of society)

Physiology and neuroscience

- Basis for human behaviour
  - Fear: two different pathways from perception to emotional reaction, one going by way of cognition (taking longer providing more detail)
  - Memory affected by levels of stress induced hormones (too high gives no (conscious) memories)
  - Trust: trust game experiments shows
    - When reciprocity is above what pure rational choice suggests it is affected by the hormone oxytocin making people less "betrayal averse" (rather than "risk averse")
    - An emotional foundation for punishing of unfair behaviour. Punishing provided stimulus for pleasure centre of the brain
  - Filling in capacity of the brain explains the quest for meaning in all information gathered

Explanation by consequences

- Reinforcement
  - If we notice positive consequences we may choose to repeat our action
  - Learning by reinforcements
    - Rewards by fixed or variable ratio schedules
    - Rewards by fixed or variable intervals
  - Learning may be extinguished by removing the reward.
    - This happens faster if the reward has been on fixed frequent ratios
  - To explain behaviour reward schedules needs to occur naturally and be opaque
    - This do not happen often with fixed schedules
  - Response patterns generated by reinforcements will seldom conform to rational choice theory
    - Responses will maximise average reward rather than marginal as rational choice would dictate

Differential reproductive fitness

- Selection
  - Agents may be selected by the environment rather than adapt to it
    - Human's consciously decide in breeding animals or plants based on their behavioural characteristics
  - Differential survival of organisms in natural settings will, across generations, increase the frequency of behaviour that increase reproductive success
    - This is called natural selection

Natural selection

- Optimize fitness measured by number of offspring
- Determined by environment and physiological and behavioural properties (phenotype)
- Genotype determined by DNA, a code written by the nucleotides T, A, G, C
  - Each "word" (or codon) of the code is a triplet of these
- Gene is a segment of DNA coding for one particular protein
- Small mutations will delete, insert or substitute one of the nucleotides
  - Evolution can be seen as an analogy to the substitution of letters in a word or sentence: will it still be meaningful? Can meaning be achieved in more than one step?
- Natural selection will generate local maxima

Why local maxima?

Small marginal mutations
- Cannot use indirect strategies, descendants have to survive
- Cannot wait for the mutation be become an improvement
- Cannot anticipate what will be an improvement in a changing environment

Caveats
- Large mutations occur
- Change across generations do not eliminate all "sub-optimal" adaptations providing for possible new starts

The units of selection

- Natural selection is opportunistic and myopic and usually fiercely individualistic
- Exceptions
  - Kin selection when seemingly altruistic behaviour increase the survival of kin carrying the same genes
  - Group selection can occur if punishment of non-cooperators is feasible (requires identification of non-cooperators)
- Kin and group selection may explain cooperative behaviour
  - Reciprocal altruism is a third mechanism, but would seem implausible in large groups since it requires the "grim trigger" strategy

Selection and human behaviour

	Intentional source of variation	Non-intentional source of variation
Intentional selection	Plant and animal husbandry	Gradual improvement of boats Eugenics Selective abortion and infanticide
Non-intentional selection	Firms in market competition	Natural selection

Natural selection

Non-intentional variation and selection

- Emotions may provide for a mechanism where genes affect certain types of behaviour (jealousy, anger against defectors, contempt against breaking moral norms, self deception, …)
  - Murder of wives and stepchildren
  - Propensity to punish non-cooperators
  - ???

Deliberate selection

Intentional variation and selection

- Animal and plant breeders, GMOs

Non-intentional variation -"-

- Sundt's example of improvements in boat design as arising from imperfect copying of earlier good boats and seamen noticing improvements
  - Leads to local maxima
  - Leads to intentional variations as builders start experimenting

Market competition

Intentional variation, non-intentional selection

- Type 1: all firms try to maximise profits, imperfect copies of more successful firms may provide improvements
- Type 2: firms are satisficers engaging in search for improvement only when profits fall below a threshold

Selection models in social science

- With high rates of change in the environment, firms needs to anticipate change to keep even with competitors
- Large firms and lobbying groups may be able to shape the environment (cpr.: path dependence mechanism)
- Modelling markets: There is a vast space between "improving efficiency" and "maximizing returns"
  - Compare: Adaptive efficiency (North 2005)
- Electoral market models do not do justice to variations in motives among politicians (opportunists, reformers, activists). Not all politicians are votre maximizers
- Outside arenas of competition the selection model of "as-if" rationality is even less plausible
- Constraints (before the fact) and selection (after the fact) contribute to explain behaviour. But **choice** is the core concept to understand

Next

The next and last section of Elster's book, chapters 18-26, are the rew material for your class presentation. Pick one chapter and prepare your presentation!