

Chapter 14: Reasoning and hypothesis testing

Reasoning is related to problem solving, because people trying to solve a reasoning task have a definite goal and the solution is not obvious. However, problem solving and reasoning are typically treated separately. Reasoning problems differ from other kinds of problems in that they often owe their origins to systems of formal logic. However, as we will see, there are clear overlaps between the two areas, which may differ less than one might initially suppose.

Inductive reasoning involves making a generalised conclusion from premises referring to particular instances. Hypotheses can never be shown to be logically true by simply generalising from confirming instances (i.e., induction). Generalisations provide no certainty for future events. *Deductive reasoning* allows us to draw conclusions that are definitely valid provided that other statements are assumed to be true. For example, if we assume that Tom is taller than Dick, and Dick is taller than Harry, then the conclusion that Tom is taller than Harry is necessarily true.

- **Hypothesis testing**

Karl Popper (1968) argued for a distinction between confirmation and falsification:

- *Confirmation* involves obtaining evidence to confirm the correctness of one's hypothesis.
- *Falsification* involves attempting to falsify hypotheses by experimental tests.

Popper argued that it is impossible to achieve confirmation via hypothesis testing. Rather, scientists should focus on falsification.

WEBLINK: [Karl Popper](#)

Wason (1960) devised the 2–4–6 task, in which participants have to discover a relational rule underlying a set of three numbers. Performance was poor on the task because people tended to show confirmation bias – they generated numbers conforming to their original hypothesis rather than trying hypothesis disconfirmation:

- A positive test is when numbers produced are an instance of your hypothesis.
- A negative test is when numbers produced do not conform to your hypothesis.

WEBLINK: [Video of Wason 2–4–6 task](#)

Wason's theoretical position predicts that people should perform better when instructed to engage in disconfirmatory testing. The evidence is mixed. Cowley and Byrne (2005) argued that people show confirmation bias because they are loath to abandon their own initial hypothesis. Tweney et al. (1980) found that performance on the 2–4–6 task was enhanced when participants were told to discover two rules, one the complement of the other. Gale and Ball (2009) argued that it was important for participants to identify the crucial dimensions of ascending vs. descending numbers.

Performance on the 2–4–6 task involves separable processes of:

- hypothesis *generation*;
- hypothesis *testing*.

Cherubini et al. (2005) argued that participants try to preserve as much of the information contained in the example triple (i.e., 2–4–6) as possible in their initial hypothesis. As a result, this hypothesis is typically much more *specific* than the correct rule. Most hypotheses are *sparse* or narrow in that they apply to less than half the possible entities in any given domain (Navarro & Perfors, 2011).

The 2–4–6 problem is a valuable source of information about inductive reasoning. The findings from the 2–4–6 task may not be generalisable because, in the real world, positive testing is not penalised. Additional factors come into play in the real world.

Hypothesis testing in simulated and real-world settings

There is a popular view that “scientific discovery is the result of genius, inspiration, and sudden insight” (Trickett & Trafton, 2007, p. 868). That view is largely incorrect. Scientists typically use what Klahr and Simon (2001) described as weak methods. Kulkarni and Simon (1988) found scientists make extensive use of the unusualness heuristic, or rule of thumb. This involves focusing on unusual or unexpected findings and then using them to guide future theorising and research. Trickett and Trafton (2007) argued that scientists make much use of “what if” reasoning in which they work out what would happen in various imaginary circumstances. Dunbar (1993) used a simulated research environment. He found that participants who simply tried to find data consistent with their hypothesis failed to solve the problem.

It is believed that scientists should focus on falsifying their hypotheses. However this does not tend to happen.

Nearly all reasoning in everyday life is inductive rather than deductive. Hypothesis testing is a form of inductive reasoning. Karl Popper argued that it is impossible to confirm a hypothesis via hypothesis testing. Rather, scientists should focus on falsification. However, it is now accepted that Popper’s views were oversimplified and confirmation is often appropriate in real scientific research. Wason (1960) devised the 2–4–6 task and found people tended to show confirmation bias, producing sequences that confirmed their hypotheses rather than seeking negative evidence. However, later studies demonstrated that people’s behaviour is often more accurately described as confirmatory or positive testing.

- **Deductive reasoning**

Conditional reasoning is basically reasoning with “if”. It has been studied to decide if human reasoning is logical. In propositional logic, meanings are different from those in natural language. There are different types of logical reasoning statements:

- “Affirmation of the consequent”:
 - Premise (if P then Q), (Q), Conclusion (P).
 - Invalid form of argument.
- “Denial of the antecedent”:
 - Premise (if P then Q), (not P), Conclusion (not P).
 - Invalid form of argument.
- “Modus tollens”:
 - Premise (If P, then Q), (not Q), Conclusion (not P).
 - Valid form of argument.
- “Modus ponens”:
 - Premise (If P, the Q), (P), Conclusion (Q).
 - Valid form of argument.

INTERACTIVE EXERCISE: Conditional reasoning

Invalid inferences (denial of the consequent, affirmation of the consequent) are accepted much of the time, the former typically more often (Evans et al., 1993). De Neys et al. (2005) found evidence that conditional reasoning is strongly influenced by the availability of knowledge in the form of counterexamples appearing to invalidate a given conclusion. They also found performance on conditional reasoning tasks depends on individual differences. Bonnefon et al. (2012) argued that reasoners draw *inferences* when presented with conditional reasoning problems. According to Markovits et al. (2013), there are two strategies people can use with problems:

- a statistical strategy;

- a counterexample strategy.

WEBLINK: [Common fallacies](#)

Various findings suggest many people fail to think logically on conditional reasoning tasks. Conditional reasoning is closer to decision making than to classical logic (Bonnefon et al., 2012).

RESEARCH ACTIVITY: Deductive reasoning

The Wason selection task has four cards, each with a number on one side and a letter on the other. Participants are told a rule and asked to select only those cards that must be turned over to decide if the rule is correct. The correct answer is only given by 5–10% of university students.

WEBLINK: [Wason selection task](#)

Many attempts have been made to account for performance on this task. Evans (e.g., 1998) identified *matching bias* as an important factor. This is the tendency for participants to select cards matching items named in the rule regardless of whether the matched items are correct. Stenning and van Lambalgen (2004) argued that people have difficulties interpreting precisely what the selection problem is all about. Oaksford (1997) argued that the logical answer to the Wason selection task conflicts with what typically makes most sense in everyday life.

Performance in the Wason selection task can be improved by:

- making the underlying structure of the problem more explicit (Sperber & Girotto, 2002);
- motivating participants to disprove the rule (Dawson et al., 2002).

Syllogistic reasoning

A *syllogism* consists of two premises or statements followed by a conclusion. The validity of the conclusion depends only on whether it follows logically from the premises. *Belief bias* is when people accept believable conclusions and reject unbelievable conclusions, irrespective of their logical validity or invalidity. Klauer et al. (2000) found various biases in syllogistic reasoning, including a base-rate effect, in which performance is influenced by the perceived probability of syllogisms being valid. Stuppel and Ball (2008) found with syllogistic reasoning that people took longer to process unbelievable premises than believable ones. Stuppel et al. (2013) found participants were more likely to accept conclusions that matched the premises in surface features than those not matching.

Conditional reasoning has its origins in a system of logic known as propositional logic. Performance on conditional reasoning problems is typically better for the modus ponens inference than for other inferences (e.g., modus tollens). Conditional reasoning is influenced by context effects (e.g., the inclusion of additional premises). Performance on the Wason selection task is generally very poor, but is markedly better when the rule is deontic rather than indicative. Performance on syllogistic reasoning tasks is affected by various biases, including belief bias and the base rate. The fact that performance on deductive reasoning tasks is prone to error and bias suggests people often fail to reason logically.

- **Theories of deductive reasoning**

The *mental models* approach is one of the most influential approaches and was proposed by Johnson-Laird (e.g., 1983, 1999). A mental model represents a possibility, capturing what is common to the different ways in which the possibility could occur. People use the information contained in the premises to construct a mental model. Here are the main assumptions of mental model theory:

- A mental model describing the given situation is constructed and the conclusions that follow are generated. The model is iconic (its structure corresponds to what it represents).

- An attempt is made to construct alternative models to falsify the conclusion by finding *counterexamples* to the conclusion. If a counterexample model is not found, the conclusion is assumed to be valid.
- The construction of mental models involves the limited resources of working memory.
- Reasoning problems requiring the construction of several mental models are harder to solve than those requiring only one mental model because of increased demands on working memory.
- The principle of truth: “Individuals minimise the load on working memory by tending to construct mental models that represent explicitly only what is true, and not what is false” (Johnson-Laird, 1999, p. 116).

Successful thinking results from the use of appropriate mental models. Unsuccessful thinking occurs when we use inappropriate mental models.

Knauff et al. (2003) found deductive reasoning was slower when it involved visual imagery. Copeland and Radvansky (2004) tested this assumption. They found a moderate correlation between working memory capacity and syllogistic reasoning. They also found that problems requiring more mental models had longer response times. Legrenzi et al. (2003) tested the principle of truth. They found performance was high on problems when adherence to the principle of truth was sufficient. In contrast, there were illusory inferences when the principle of truth did not permit correct inferences to be drawn. People are less susceptible to such inferences if explicitly instructed to falsify the premises of reasoning problems (Newsome & Johnson-Laird, 2006). Newstead et al. (1999; see Eysenck & Keane, 2015, p. 605) also found participants consistently failed to produce more mental models for multiple-model syllogisms than for single-model ones.

Most predictions of mental model theory have been confirmed experimentally. In particular, evidence shows that people make errors by using the principle of truth and ignoring what is false. Limitations with the theory are that it assumes that people engage in deductive reasoning to a greater extent than is actually the case. The processes involved in forming mental models are underspecified.

Dual-process theories

There are two process involved in human reasoning:

- One system involves unconscious processes and parallel processing, and is independent of intelligence.
- The other system involves conscious processes and rule-based serial processing, has limited capacity and is linked to intelligence.

Evans (2006) proposed the heuristic–analytic theory of reasoning, which distinguishes between heuristic processes (System 1) and analytic processes (System 2). Initially, heuristic processes use task features and knowledge to construct a single mental model. Later, effortful analytic processes may intervene to revise this model. This is more likely when:

- task instructions tell participants to use abstract or logical reasoning;
- participants are highly intelligent;
- sufficient time is available for effortful analytic processing;
- participants need to justify their reasoning.

Human reasoning is based on the use of three principles:

- Singularity principle
- Relevance principle
- Satisficing principle.

In contrast to the mental model theory, the heuristic–analytic theory predicts that people initially use their world knowledge and immediate context in reasoning. Deductive reasoning is regarded as less important.

CASE STUDY: Exploring dual-system theories of deductive reasoning

Belief bias is a useful phenomenon for distinguishing between heuristic and analytic processes. Evans (2000) found less evidence of belief bias when instructions emphasised logical reasoning. Stuppel et al. (2011) compared groups of participants who showed much evidence of belief bias and those showing little belief bias. Those with high levels of belief bias responded faster on syllogistic reasoning problems. De Neys (2006) found high working memory capacity was an advantage only on problems requiring the use of analytic processes. A secondary task impaired performance only on problems requiring analytic processes. Evans and Curtis-Holmes (2005) found belief bias was stronger when time was strictly limited.

INTERACTIVE EXERCISE: Heuristic–analytic theory

Thompson et al. (2011) suggests two processes are used in syllogistic reasoning:

- Participants provided an intuitive answer.
- This was followed by an assessment of that answer's correctness (feeling of rightness). After which, participants had unlimited time to reconsider their initial answer and provide a final answer (analytic or deliberate answer).

Thompson et al. (2011) argued that we possess a *monitoring system* (assessed by the feeling-of-rightness ratings) that evaluates the output of heuristic or intuitive processes. Evidence that people are more responsive to the logical structure of reasoning problems than suggested by performance accuracy was reported by De Neys et al. (2010).

The heuristic–analytic theory of reasoning has several successes:

- The notion that cognitive processes used by individuals to solve reasoning problems are the same as those used in other cognitive tasks seems correct.
- Evidence supports the notion that thinking is based on singularity, relevance and satisficing principles.
- There is convincing evidence for the distinction between heuristic and analytic processes.
- The theory accounts for individual differences, for example in working memory capacity.

Limitations with the approach are:

- It is an oversimplification to distinguish between implicit heuristic and explicit analytic processes.
- The distinction between heuristic and analytic processes poses the problem of working out exactly *how* these two different kinds of processes interact.
- It is not very clear precisely what the analytic processes are or how individuals decide which ones to use.
- Logical processing can involve heuristic or intuitive processes occurring below the conscious level.
- The assumption that heuristic processing is followed by analytic processing in a serial fashion may not be entirely correct.

According to mental model theory, people construct one or more mental models, mainly representing explicitly what is true. Mental model theory fails to specify in detail how the initial mental models are constructed, and people often form fewer mental models than expected. Dual-system theories answer the two main limitations of most other research into human reasoning because they take account of individual differences in performance and processes. There is now convincing evidence for a distinction between relatively automatic, heuristic-based processes and more effortful analytic-based processes. However, it is unlikely that we can capture all the richness of human reasoning simply by assuming the existence of two cognitive systems.

- **Brain systems in reasoning**

Prado et al. (2011) found the brain system for deductive reasoning is centred in the left hemisphere involving frontal and parietal areas. Specific brain areas activated during deductive reasoning included:

- inferior frontal gyrus;

- middle frontal gyrus;
- medial frontal gyrus;
- precentral gyrus;
- basal ganglia.

Goel et al. (2007) studied patients having damage to left or right parietal cortex. Those with left-side damage performed worse than those with right-side damage on reasoning tasks in which complete information was provided.

Prado et al. (2011) found the precise brain areas associated with deductive reasoning depended to a large extent on the nature of the task. Prado et al. (2011) also found the left inferior frontal gyrus (BA9/44) was more activated during the processing of categorical arguments. Prado et al. (2011) found the left precentral gyrus (BA6) was more activated with propositional reasoning than with categorical or relational reasoning. Language seems to play little or no role in processing of reasoning tasks post-reading (Monti & Osherson, 2012).

Reverberi et al. (2012) identified three strategies used in categorical reasoning:

- sensitivity to the logical form of problems:
 - the left inferior lateral frontal (BA44/45) and superior medial frontal (BA6/8) areas;
- sensitivity to the validity of conclusions (i.e., accurate performance):
 - the left ventro-lateral frontal (BA47) area;
- use of heuristic strategies:
 - no specific pattern of brain activation.

More intelligent individuals exhibit less belief bias because they make more use of analytic processing strategies (De Neys, 2006). Individual differences in performance accuracy (and thus low belief bias) were strongly associated with activation in the right inferior frontal cortex under low and high cognitive load conditions (Tsuji & Watanabe, 2009).

Fangmeier et al. (2006) used mental model theory as the basis for assuming the existence of three stages of processing in relational reasoning. Different brain areas were associated with each stage:

- *Premise processing*: temporo-occipital activation reflecting the use of visuo-spatial processing.
- *Premise integration*: anterior prefrontal cortex (e.g., BA10), an area associated with executive processing.
- *Validation*: the posterior parietal cortex was activated, as were areas within the prefrontal cortex (BA6, BA8) and the dorsal cingulate cortex.

Bonnefond et al. (2013) studied the brain processes associated with conditional reasoning focusing on modus ponens: There is enhanced brain activity when premises and conclusions do not match and *anticipatory* processing before the second premise occurs when they match.

Limited progress has been made in identifying the brain systems involved in deductive reasoning. This is because of simple task differences and individual differences that affect the results.

- **Informal reasoning**

Informal reasoning is a form of reasoning based on one's knowledge and experience. People make extensive use of informal reasoning processes such as heuristics in formal deductive reasoning tasks. However, there are also differences between processes in formal and informal reasoning:

- content;
- contextual factors;
- informal reasoning concerns probabilities;
- motivation.

Ricco (2007) identified common informal fallacies:

- *Irrelevance* (seeking to support a claim with an irrelevant reason).
- *Slippery slope*.

The *myside bias* is the tendency to evaluate statements with respect to one's own beliefs rather than solely on their merits (Stanovich & West, 2007).

Support for the probabilistic approach was reported by Hahn and Oaksford (2007). They identified several factors influencing the perceived strength of a conclusion:

- degree of previous conviction or belief;
- positive arguments have more impact than negative arguments;
- strength of the evidence.

Hahn and Oaksford (2007) found a Bayesian model predicted informal reasoning performance very well. However, Bowers and Davis (2012) argued that the Bayesian approach is too flexible and thus hard to falsify. Sá et al. (2005) found unsophisticated reasoning was more common among those of lower cognitive ability.

Informal reasoning is more important in everyday life than deductive reasoning. However, most reasoning research is far removed from everyday life. Hahn and Oaksford (2007) proposed a framework for research on informal reasoning based on probabilistic principles. There is reasonable support for their model, particularly for the role of prior belief and new evidence on strength of argument. In future, it will be important to establish the similarities and differences in processes underlying performance on informal and deductive reasoning tasks.

- **Are humans rational?**

Much evidence seems to indicate that our thinking and reasoning are often inadequate, suggesting that we are not rational. Human performance on deductive reasoning tasks does seem very prone to error.

RESEARCH ACTIVITY: Cognitive Reflection Test

RESEARCH ACTIVITY: Are humans rational?

Most people cope well with problems in everyday life, yet seem irrational and illogical when given reasoning problems in the laboratory. However, it may well be that our everyday thinking is less rational than we believe.

Heuristics allow us to make rapid, reasonably accurate, judgements and decisions, as Maule and Hodgkinson (2002) pointed out. Laboratory research findings suggest people can think rationally when problems are presented in a readily understandable form. Many of the apparent "errors" on deductive reasoning tasks may also be less serious than they seem. There is reasonable support for the notion that factors such as participants' misinterpretation of problems, or lack of motivation, explain only a fraction of errors in thinking and reasoning (e.g., Camerer & Hogarth, 1999).

Individual differences in intelligence and working memory also influence performance on conditional reasoning tasks. Some researchers have found inadequacies in performance even when steps are taken to ensure that participants fully understand the problem (e.g., Tversky & Kahneman's (1983) conjunction fallacy study). Interestingly, those who are incompetent have little insight into their reasoning failures; this is the Dunning-Kruger effect (Dunning, 2011).

Deciding whether humans are rational depends on how we define "rationality". Sternberg (2011, p. 270) pointed out that "Few problems of consequence in our lives had a deductive or even any meaningful kind of 'correct' solution." *Normativism* "is the idea that human thinking reflects a normative system [one conforming to norms or standards] against which it should be measured and judged" (Elqayam & Evans, 2011, p.233). An alternative

approach is that human rationality involves effective use of probabilities rather than logic. Oaksford and Chater (e.g., 2009) put forward an influential probabilistic approach to human reasoning.

Simon (1945) suggested the notion of *bounded rationality* should be considered in human reasoning. This means an individual's informal reasoning is rational if it achieves his/her goal of arguing persuasively. Many "errors" in human thinking are due to limited processing capacity rather than irrationality.

Toplak et al. (2011) reported a correlation of +0.32 between cognitive ability and performance across 15 judgement and decision tasks. Stanovich (2012) developed the tripartite model with two levels of processing:

- Type 1 processing (e.g., use of heuristics) within the autonomous mind is rapid and fairly automatic.
- Type 2 processing (also called System 2 processing), which is slow and effortful.

There are *three* different reasons why individuals produce incorrect responses when confronted by problems:

- The individual lacks the mindware (e.g., rules, strategies) to override the heuristic response.
- The individual has the necessary mindware but fails to realise the need to override the heuristic response.
- The individual has the necessary mindware and realises that the heuristic response should be overridden, but doesn't have sufficient decoupling capacity.

Stanovich (2009, p. 35) used the term *dysrationalia* to refer to "the inability to think and behave rationally despite having adequate intelligence". Most people (including those with high IQs) are cognitive misers, preferring to solve problems with fast, easy strategies than with more accurate effortful ones.

Humans can be considered rational because:

- Errors are caused by limited processing capacity (Simon, 1957).
- Classical logic is almost totally irrelevant to our everyday lives because it deals in certainties.
- Our thinking and reasoning are rational when used to achieve our goals

However, humans can be considered irrational because:

- Many humans are cognitive misers.
- There is a widespread tendency on judgement tasks to de-emphasise base-rate information.
- They fail to think rationally because they are unaware of limitations and errors in their thinking.

Apparently poor performance by most people on deductive reasoning tasks does not mean we are illogical and irrational because of the existence of the normative system problem, the interpretation problem and the external validity problem.

Additional reference

Gale, M. & Ball, L.J. (2009). Exploring the determinants of dual goal facilitation in a rule discovery task. *Thinking & Reasoning*, 15: 294–315.