INVESTIGATION · FOUNDATIONS

Hypothesis Generation and Elimination in Investigations

Investigations succeed or fail on how well investigators generate possible explanations and how rigorously they eliminate them. This article explains the logic of hypotheses in investigation, the psychological traps involved, and structured methods such as the analysis of competing hypotheses.

ModQu research article · 2,831 words · 27 references (APA 7th edition)

Every investigation is, at heart, a contest between explanations. Money is missing: was it stolen, mislaid, misposted or never there? A fire broke out: was it arson, an electrical fault or an accident with a cigarette? A respondent pool produced strange survey results: is it fraud, a programming error or a real shift in opinion? The investigator's task is to identify the plausible explanations, work out what evidence would tell them apart, gather that evidence, and see which explanations survive. How well this is done determines whether the investigation reaches the truth or merely reaches a conclusion.

This article explains the role of hypotheses in investigation, describes how investigators generate and eliminate them, reviews the research on where the process goes wrong, and sets out structured techniques that make it more reliable.

What a hypothesis is in an investigation

In an investigative setting, a hypothesis is a proposed explanation of the events under inquiry that can be tested against evidence. It need not be elaborate. "The cashier took the money" is a hypothesis, as is "the till was miscounted at the start of the shift". What makes it a hypothesis rather than a belief is that the investigator treats it as provisional and asks what would have to be true if it were correct.

Philosophers describe the reasoning that generates hypotheses as abduction: moving from a surprising observation to an explanation that would, if true, make the observation unsurprising (Josephson & Josephson, 1994). Lipton (2004) calls the related process "inference to the best explanation", in which we accept the explanation that would, if correct, provide the most satisfying account of the evidence. Abduction is creative and fallible. It generates candidates; it does not establish them. Establishing them requires testing, and here investigation borrows from the logic of science. Popper (1959) argued that no number of confirming observations can prove a general theory, but a single clear counter-example can refute it. Investigators do not deal in general laws, but the asymmetry still matters: a hypothesis that predicts something that turns out to be false is in serious trouble, whereas a hypothesis that is consistent with the evidence may simply be one of several that are.

Generating hypotheses

Why generation matters

An investigation can only find an explanation that someone thinks of. If the true explanation is never considered, no amount of careful testing will identify it. Research on diagnosis in medicine illustrates the point. Elstein et al. (1978) found that physicians generated a small number of diagnostic hypotheses very early in a consultation, often within the first few minutes, and that the accuracy of the final diagnosis depended heavily on whether the correct one was among those early candidates. Croskerry (2003) lists premature closure, accepting a diagnosis before it has been fully verified, among the most common causes of diagnostic error.

The same pattern appears in investigation. Interview studies with detectives show that they form working theories early and that certain pieces of information act as tipping points, after which the inquiry narrows around a suspect (Fahsing & Ask, 2013). Once that happens, explanations that were never on the list receive no attention at all.

How people generate hypotheses

Psychological models of hypothesis generation suggest that people retrieve candidate explanations from memory, cued by the features of the current situation, and that the number they can hold in mind at once is small (Thomas et al., 2008). Two consequences follow. First, investigators tend to generate the explanations that are most familiar or most recently encountered, not necessarily the most likely. Second, they tend to overestimate how complete their list is. In a classic study, Fischhoff et al. (1978) showed people fault trees listing possible reasons why a car might fail to start. When some branches were removed, participants did not notice the omission and assigned little probability to an "all other problems" category, even though the missing causes were common. What is out of sight is out of mind.

Methods for better generation

Investigators can counter these limits with deliberate techniques:

- **Brainstorm before evaluating.** List every explanation the team can think of, including innocent and unlikely ones, before discussing which is most probable. Mixing generation and evaluation tends to kill unusual ideas early.
- **Use structured prompts.** Ask systematically about people (who had access, motive and opportunity?), processes (could a system, procedure or error produce this?), timing (when could it have happened?) and external factors (could someone outside be responsible?).
- **Include the null hypothesis.** Always consider that nothing improper happened, and that the apparent anomaly has an ordinary explanation such as a recording error, a timing difference or chance.
- **Involve outsiders.** People with different expertise and no investment in the case notice explanations that the core team misses.
- **Revisit the list.** New evidence often suggests explanations that were not considered at the start. Hypothesis generation is not a single step but a recurring one.

Experience helps, but in complex ways. Fahsing and Ask (2016) found that experienced detectives in England and Norway generated more relevant hypotheses and lines of inquiry about a case than novices, but the advantage depended on the quality of training and the investigative culture in which experience was gained, not simply on years of service.

Testing and eliminating hypotheses

The logic of diagnostic evidence

The key question when testing hypotheses is not "is this evidence consistent with my theory?" but "how much better does my theory explain this evidence than the alternatives do?" Evidence that is equally consistent with every hypothesis on the list has no diagnostic value, however striking it seems. If a suspect was near the scene, but so were hundreds of other people for innocent reasons, the fact that the suspect was there does little to distinguish between explanations.

Evidence scholars express this as the inferential force of evidence (Schum, 1994). In probabilistic terms, the force of an item of evidence depends on the ratio between the probability of seeing it if one hypothesis is true and the probability of seeing it if another is true (Anderson et al., 2005). Items with a ratio close to one are of little use, however much they seem to "fit". Items that would be very unlikely under one hypothesis and likely under another are highly diagnostic.

A simple example shows the difference. Suppose an investigator is trying to decide whether a disputed payment was authorised by a manager or by someone using the manager's login. The payment was approved from the manager's account: that fact is expected under both hypotheses, so it does not help. The approval was made at 2:14 a.m. from an address outside the company network. If the manager never works at night and never logs in remotely, that observation is much more likely under the second hypothesis than the first, and it shifts the balance. If the manager regularly approves payments late at night from home, the same observation tells the investigator very little. The value of evidence depends on what the alternatives predict, which is why listing alternatives is not a formality but the basis of the whole assessment.

Investigators rarely put numbers on these judgments, and they do not need to. What matters is the habit of asking, for every significant item, "how likely would I be to see this if my main theory were wrong?" Evidence that the investigator would expect to see whatever the truth is should be recorded but given little weight.

The positive test strategy

People tend to test hypotheses by looking for cases where the hypothesised feature should be present, a habit Klayman and Ha (1987) called the positive test strategy. In many everyday situations this is a reasonable heuristic. But it becomes a trap when the hypothesis is too narrow or when evidence that would disconfirm it is never sought. Wason (1960) famously asked people to discover a rule governing a sequence of three numbers. Most proposed a hypothesis and then tested only sequences that fitted it, receiving confirmation each time, and announced a rule that was wrong. They never tested a sequence that their hypothesis predicted would fail.

In investigation, the equivalent is gathering evidence that the main suspect could have committed the act, without checking whether others could have done so too, or whether there is evidence that the suspect did not. Rassin et al. (2010) found that participants in an investigative scenario preferred to seek incriminating rather than exonerating

evidence. O'Brien (2009) found that naming a suspect early made participants less receptive to alternatives and more likely to interpret ambiguous information as incriminating. Nickerson (1998) reviews the many forms that such confirmation bias takes.

Elimination rather than confirmation

The remedy is to focus on elimination. Rather than asking which hypothesis has the most evidence for it, the investigator asks which hypotheses can be ruled out, and what evidence would rule out each of the survivors. A hypothesis that has survived serious attempts to disprove it deserves more confidence than one that has merely accumulated consistent evidence. This approach also has a practical advantage: it directs effort towards the evidence that will actually change the conclusion.

Elimination should be firm only when the evidence is strong. Many investigative hypotheses are not ruled out by a single fact but become steadily less likely as evidence accumulates against them. Investigators should record the strength of each elimination, so that if a key piece of evidence later proves unreliable, they know which hypotheses need to be reopened.

The analysis of competing hypotheses

The best-known structured method for hypothesis testing is the **analysis of competing hypotheses (ACH)**, developed by Richards Heuer for intelligence analysts at the Central Intelligence Agency (Heuer, 1999). It is now widely taught to analysts and investigators (Heuer & Pherson, 2015). The method has eight steps, which can be summarised as follows:

1. Identify all the reasonable hypotheses, using a group with diverse perspectives.
2. List the significant evidence and arguments for and against each hypothesis, including assumptions and the absence of expected evidence.
3. Build a matrix with hypotheses across the top and evidence down the side, and assess whether each item is consistent, inconsistent or not applicable to each hypothesis.
4. Refine the matrix: add hypotheses that emerge, and remove evidence that has no diagnostic value because it is consistent with everything.
5. Draw tentative conclusions by working down the columns, focusing on how much evidence is inconsistent with each hypothesis rather than how much is consistent.
6. Analyse how sensitive the conclusion is to a few critical items of evidence, and consider what would happen if they were wrong or deceptive.
7. Report the conclusions, including the relative likelihood of all the hypotheses, not just the most likely.
8. Identify indicators that would show if events are taking a different course than expected.

An analysis of competing hypotheses matrix

Illustrative case: a stock shortfall at a warehouse. The focus is on inconsistent evidence, read down each column.

	H1: Theft by night supervisor	H2: Theft by another employee	H3: Stock system error	H4: Supplier short delivery
Losses began the week the new system went live	Neutral	Neutral	Consistent	Neutral
Losses occur at the same rate on every shift	Inconsistent	Consistent	Consistent	Consistent
Losses continue during supervisor's leave	Inconsistent	Consistent	Consistent	Consistent
Delivery notes match purchase orders	Neutral	Neutral	Neutral	Inconsistent
Supervisor has prior dismissal	Consistent	Neutral	Neutral	Neutral
Inconsistent items	2	0	0	1

Illustrative data. Method based on Heuer (1999) and Heuer & Pherson (2015) ModQu.com

Figure 1. A simplified analysis of competing hypotheses matrix for a hypothetical stock shortfall. The prior dismissal is consistent with H1 but has little diagnostic value; the shift pattern and losses during leave are inconsistent with H1. H2 and H3 survive, and the next step is to find evidence that distinguishes them. Diagram by ModQu.

ACH has two main strengths. It forces analysts to consider several hypotheses at once rather than evaluating one at a time, and it focuses attention on disconfirming evidence. It also produces a record that others can examine and challenge. Its evidence base is more mixed than its popularity suggests. Experimental studies have found that ACH does not consistently improve the accuracy of judgments and may not remove some biases, particularly when analysts are given the evidence in a particular order or have a strong initial view (Dhami et al., 2019; Whitesmith, 2020). The method's value probably lies less in the matrix itself than in the habits it encourages: listing alternatives, thinking about diagnosticity and documenting the reasoning.

Other structured approaches

ACH is one of a family of techniques designed to make reasoning more explicit (Heuer & Pherson, 2015):

- **Key assumptions check.** List the assumptions on which the current view depends, and ask how confident the team is in each and what would happen if it were wrong.
- **Devil's advocacy and red teams.** Assign a person or group to argue the strongest case against the prevailing view.

- **Consider the opposite.** Ask explicitly how the evidence might support a different conclusion. Lord et al. (1984) found that this simple instruction reduced biased interpretation of evidence more effectively than asking people to be unbiased.
- **Premortem.** Imagine that the investigation's conclusion has later proved wrong, and ask what the most likely reason would be.
- **Timeline and link analysis.** Laying out events in time order or mapping relationships between people, accounts and places often reveals inconsistencies that suggest new hypotheses or rule out old ones.

Journalists use a related approach. The story-based inquiry method treats the story as a hypothesis to be verified, and encourages reporters to state at the outset what they would need to find to prove it and what would disprove it (Hunter et al., 2011).

Why investigations narrow too early

Several pressures push investigators towards early commitment to one hypothesis:

- **Time and resources.** Keeping several hypotheses open requires more work. When resources are short, there is pressure to focus.
- **External expectations.** Victims, managers, the media and the public want answers. A named suspect relieves that pressure.
- **Organisational culture.** In some teams, questioning the lead theory is seen as disloyal or obstructive.
- **The appeal of a coherent story.** A narrative that ties the evidence together feels convincing, even if parts of it are weak. Klein et al. (2006) describe sensemaking as fitting data into a frame, and note that once a frame is established, data that do not fit tend to be discounted or reinterpreted.
- **Expert interpretation.** Forensic and other experts can be influenced by knowing which hypothesis the investigators favour, so that the evidence presented as independent support for a theory may in part reflect it (Kassin et al., 2013).

Reviews of failed investigations show these forces at work. Findley and Scott (2006) and Rossmo (2009) describe cases in which investigators focused on an early suspect, interpreted ambiguous evidence as incriminating, dismissed or failed to pursue evidence pointing elsewhere, and continued to defend their theory after contrary evidence emerged. In many cases the investigators were experienced and conscientious; the failures were systemic rather than personal.

Practical guidance

The research and practice discussed above can be distilled into a set of working rules for any investigation:

1. **Write the hypotheses down** at the start, including innocent explanations and "something we have not thought of".
2. **For each hypothesis, state what evidence should exist** if it is true and what evidence would rule it out.
3. **Prioritise diagnostic evidence**, meaning evidence that would distinguish between hypotheses, over evidence that merely fits the favoured one.
4. **Delay focusing on an individual** until the evidence justifies it, and record when and why the focus narrowed.

5. **Seek disconfirmation of the leading hypothesis**, not just confirmation.
6. **Record the strength of each elimination**, so that hypotheses can be reopened if evidence changes.
7. **Review periodically with people outside the team**, using techniques such as a key assumptions check or devil's advocacy.
8. **Report residual uncertainty.** If more than one hypothesis remains plausible, say so, and explain what further evidence could resolve the question.

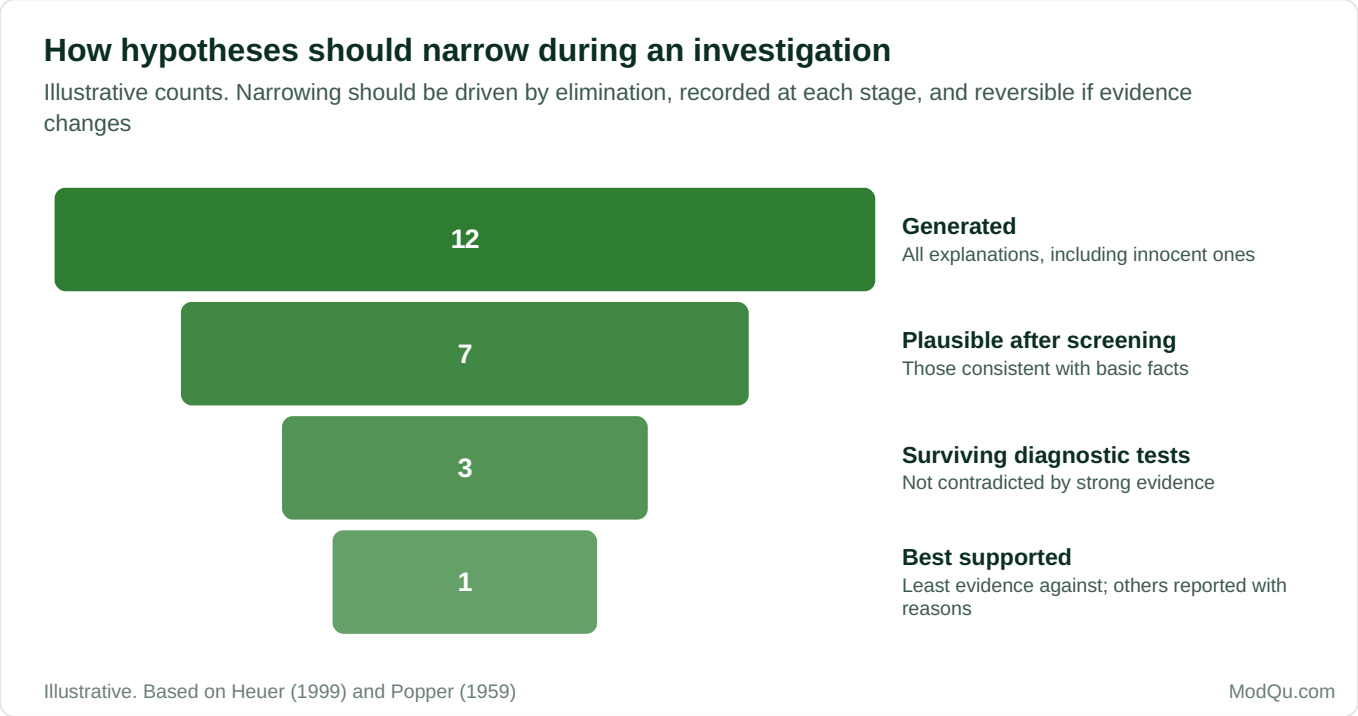


Figure 2. An idealised narrowing of hypotheses. In a well-run investigation the reduction happens through documented elimination, not through attention drifting to a favoured explanation. Diagram by ModQu.

Hypotheses and the standard of proof

Eliminating hypotheses does not end with a single survivor in every case. Sometimes the evidence leaves two or more explanations standing. In that case, what the investigator can conclude depends on the standard of proof required. A criminal prosecution requires that every reasonable alternative to guilt be excluded. A civil or disciplinary finding may rest on the more probable explanation. A safety investigation may recommend action to address several possible causes at once. Being clear about which standard applies prevents investigators either overstating their findings or failing to act when action is justified.

Innes (2003), in his study of homicide investigations, observed that detectives construct a narrative of what happened that must then survive the scrutiny of prosecutors and courts. Hypothesis testing during the investigation is a rehearsal for that scrutiny. An investigator who has already asked how the defence might explain the evidence, and has looked for evidence that would rule out those explanations, presents a stronger case and is less likely to have charged the wrong person.

Beyond criminal investigation

The same logic applies wherever people must explain unexpected results. A market researcher who finds a sudden drop in brand awareness should consider fieldwork changes, sample differences, questionnaire edits and data processing errors before concluding that public opinion has shifted. An analyst investigating declining sales should test explanations about pricing, competition, distribution and measurement, not just the explanation favoured by the sales director. In each case, the discipline is the same: generate a wide set of explanations, look for evidence that distinguishes them, eliminate deliberately, and report honestly what remains uncertain. That discipline is what separates investigation from the search for confirmation.

References

Anderson, T., Schum, D., & Twining, W. (2005). *Analysis of evidence* (2nd ed.). Cambridge University Press. https://doi.org/10.1017/CBO9780511610585

Croskerry, P. (2003). The importance of cognitive errors in diagnosis and strategies to minimize them. *Academic Medicine*, *78*(8), 775–780. https://doi.org/10.1097/00001888-200308000-00003

Dhami, M. K., Belton, I. K., & Mandel, D. R. (2019). The "analysis of competing hypotheses" in intelligence analysis. *Applied Cognitive Psychology*, *33*(6), 1080–1090. https://doi.org/10.1002/acp.3550

Elstein, A. S., Shulman, L. S., & Sprafka, S. A. (1978). *Medical problem solving: An analysis of clinical reasoning*. Harvard University Press.

Fahsing, I., & Ask, K. (2013). Decision making and decisional tipping points in homicide investigations: An interview study of British and Norwegian detectives. *Journal of Investigative Psychology and Offender Profiling*, *10*(2), 155–165. https://doi.org/10.1002/jip.1384

Fahsing, I., & Ask, K. (2016). The making of an expert detective: The role of experience in English and Norwegian police officers' investigative decision-making. *Psychology, Crime & Law*, *22*(3), 203–223. https://doi.org/10.1080/1068316X.2015.1077249

Findley, K. A., & Scott, M. S. (2006). The multiple dimensions of tunnel vision in criminal cases. *Wisconsin Law Review*, *2006*(2), 291–397.

Fischhoff, B., Slovic, P., & Lichtenstein, S. (1978). Fault trees: Sensitivity of estimated failure probabilities to problem representation. *Journal of Experimental Psychology: Human Perception and Performance*, *4*(2), 330–344. https://doi.org/10.1037/0096-1523.4.2.330

Heuer, R. J., Jr., & Pherson, R. H. (2015). *Structured analytic techniques for intelligence analysis* (2nd ed.). CQ Press.

Heuer, R. J., Jr. (1999). *Psychology of intelligence analysis*. Center for the Study of Intelligence, Central Intelligence Agency.

Hunter, M. L., Hanson, N., Sabbagh, R., Sengers, L., Sullivan, D., & Svith, F. T. (2011). *Story-based inquiry: A manual for investigative journalists*. UNESCO.

Innes, M. (2003). *Investigating murder: Detective work and the police response to criminal homicide*. Oxford University Press.

Josephson, J. R., & Josephson, S. G. (Eds.). (1994). *Abductive inference: Computation, philosophy, technology*. Cambridge University Press. https://doi.org/10.1017/CBO9780511530128

Kassin, S. M., Dror, I. E., & Kukucka, J. (2013). The forensic confirmation bias: Problems, perspectives, and proposed solutions. *Journal of Applied Research in Memory and Cognition, 2*(1), 42–52. https://doi.org/10.1016/j.jarmac.2013.01.001

Klayman, J., & Ha, Y.-W. (1987). Confirmation, disconfirmation, and information in hypothesis testing. *Psychological Review, 94*(2), 211–228. https://doi.org/10.1037/0033-295X.94.2.211

Klein, G., Moon, B., & Hoffman, R. R. (2006). Making sense of sensemaking 2: A macrocognitive model. *IEEE Intelligent Systems, 21*(5), 88–92. https://doi.org/10.1109/MIS.2006.100

Lipton, P. (2004). *Inference to the best explanation* (2nd ed.). Routledge.

Lord, C. G., Lepper, M. R., & Preston, E. (1984). Considering the opposite: A corrective strategy for social judgment. *Journal of Personality and Social Psychology, 47*(6), 1231–1243. https://doi.org/10.1037/0022-3514.47.6.1231

Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology, 2*(2), 175–220. https://doi.org/10.1037/1089-2680.2.2.175

O'Brien, B. (2009). Prime suspect: An examination of factors that aggravate and counteract confirmation bias in criminal investigations. *Psychology, Public Policy, and Law, 15*(4), 315–334. https://doi.org/10.1037/a0017881

Popper, K. R. (1959). *The logic of scientific discovery*. Hutchinson.

Rassin, E., Eerland, A., & Kuijpers, I. (2010). Let's find the evidence: An analogue study of confirmation bias in criminal investigations. *Journal of Investigative Psychology and Offender Profiling, 7*(3), 231–246. https://doi.org/10.1002/jip.126

Rossmo, D. K. (2009). *Criminal investigative failures*. CRC Press.

Schum, D. A. (1994). *The evidential foundations of probabilistic reasoning*. Wiley.

Thomas, R. P., Dougherty, M. R., Sprenger, A. M., & Harbison, J. I. (2008). Diagnostic hypothesis generation and human judgment. *Psychological Review, 115*(1), 155–185. https://doi.org/10.1037/0033-295X.115.1.155

Wason, P. C. (1960). On the failure to eliminate hypotheses in a conceptual task. *Quarterly Journal of Experimental Psychology, 12*(3), 129–140. https://doi.org/10.1080/17470216008416717

Whitesmith, M. (2020). *Cognitive bias in intelligence analysis: Testing the analysis of competing hypotheses method*. Edinburgh University Press.