Our dubious decisions

How to improve decision strategies in ophthalmology

Atle Einar Østern, MD,
Oslo University Hospital HF (Ullevål)

In August 2021, external ophthalmologists referred a patient with fungal septicemia to the Eye Department in Oslo for a second opinion. Did the patient have fungal endophthalmitis, needing intraocular vitrectomy and treatment? The Optomap image was presented at the morning meeting, where our experts agreed that the visible ocular findings could indicate vitritis and chorioretinitis. However, the senior consultant in charge returned at the following session and informed us that a proper eye examination had revealed only vitreous opacities and myelinated nerve fibers, with no pathological consequences.

This case was a real eye-opener. How were we misled? The patient was severely ill, and another ophthalmological department suspected an associated eye infection; this information influenced our first perception. In short, we suffered from a bias (Figure 1). There are many different biases, which include the following:[1]

- **Anchoring** occurs when we rely too much on the first piece of information about something without adjusting our beliefs when we learn more later.
- **Availability bias (recent case bias or posterior probability error)** is the tendency to think about a particular diagnosis because of the influence of an easily remembered event or a recent experience.
- **Base rate neglect** is the inclination to ignore the true prevalence of a (rare) disease during diagnostic judgments (skewed Bayesian reasoning), resulting in false positives and adverse events.
- **Confirmation bias** is the premature closure because of the tendency to search for, interpret, favor, and recall information that supports our initial diagnosis.
- **The gambler's fallacy** means that the frequency of preceding cases influences our estimation of the probability of a random event.
- **Loss aversion (and the sunk cost fallacy)** is the tendency to avoid losses more than acquiring the same gains and for a physician not to abandon a diagnosis once mental energy and time have been invested.
- **The overconfidence effect** is when the subjective certainty of internal judgments is greater than the objective accuracy.
- **Representativeness heuristic bias (restraint)** occurs when we misjudge the probability of something because we overestimate its resemblance to a mental stereotype.

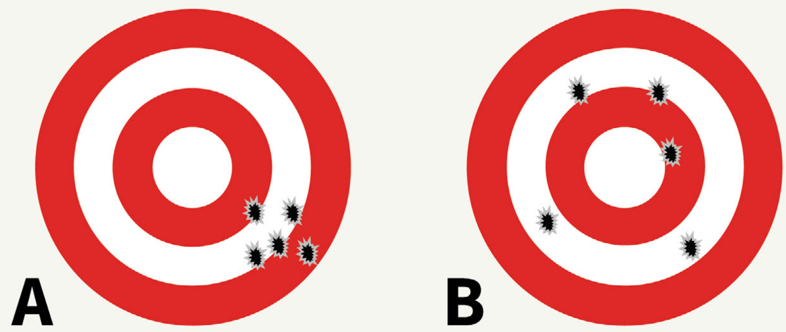


Figure 1. Visualizing the terminology. Bias (A) in psychology occurs when judgment errors diverge in one direction (invalid results). Noise (B) is the variable and random distribution of judgments that should not differ—for instance, among clinical experts (unreliable results). A fallacy is incorrect logical reasoning. Heuristics are simplified, practical problem-solving methods that include mental shortcuts. Illustration: Atle Østern

This story also illustrates another point. Group discussions frequently generate better outcomes after exchanging ideas and regression to the mean, especially when all opinions are respected and the participants have diverse areas of experience. However, there is also a risk that, due to group dynamics, the conclusion may be more extreme than the average opinion of the individual participants. This outcome is attributable to various cascade effects and group polarization, which increase noise (**Figure 1**).

Our flaws

Most of us aspire to become better at diagnosing and treating patients through lifelong learning founded on updated evidence-based medicine. Knowledge is vital and matters a great deal, as might be expected. The more experienced a doctor is, the more likely their decisions are to be correct. For instance, variation in skills can explain 44% of discrepancies in diagnostic decisions about pneumonia. The best evaluations come from professionals with procedure-specific expertise, good comprehension, and an open-minded cognitive attitude. However, doctors are only human—we are not immune to bias and noise, as the above case demonstrates.

Many factors influence decision-making. Before a decision, there will be a judgment. Judgment in this context refers to a measurement where the human mind provides a qualitative or quantitative assessment of something. Most medical judgments are predictive and later verifiable during follow-ups. However, as a clinician, you usually make judgments independently without the simultaneous, systematic evaluation of other experts.

Judgments can vary over time, and they can be shaped by factors as diverse as the doctor's mood (including fatigue and stress) or even the time of the day (for instance, when one is behind schedule). Hypotheses are often generated prematurely, only a few are actively considered, new information is discarded, distractions lead us astray, prior experiences are not fully taken advantage of, and our competence varies for different topics. In addition, reliance on memory alone is risky because it is selective, inaccurate, and changeable.

A clinical examination involves numerous decisions, both conscious and unconscious. Whether small or large, each decision opens up potentially divergent pathways for different outcomes. Even though the purpose of continuing education programs and revised guidelines is to minimize mistakes, research shows an astonishing level of disagreement among medical experts due to noise.[2] The resulting amount of damage can be devastating. For example, if two doctors give different diagnoses, at least one of them is wrong. Depending on which physician a patient meets, this could mean the difference between life and death.

Review your last day at work. Were all your judgments and clinical decisions of the "highest standard," utterly free of mistakes—or unacceptable risk thereof? Remember that the outcome—for example, an improved disease status in a patient—is not always proof positive that the prior treatment decision was sound. Perhaps the change coincided with a random variation. How systematic were you during the decision-making process in each case? Have you reflected on the decision strategies you apply at all times? Most of us will honestly realize that there is significant potential for improvement. Let us now dwell on the answers to these questions.

Figure 2. Consequence analysis. Chess is the ultimate decision-making board game. Our limited short-term memory makes foreseeing all possible outcomes challenging, especially in situations with many uncertainties and possibilities. Photo: Atle Østern

The two systems

The brain is not like a computer that can calculate the consequences of all potential moves (**Figure 2**). Instead, our mind is the overall manifestation of the network of integrated neurons interacting on all levels. Dreams and conversations show how we seamlessly drift from one topic to another associated theme. Furthermore, we are selective and tend to apply categories to make sense of all the sensory information we receive. Consequently, we are often right but also prone to errors.

The 2002 Nobel Prize winner in Economics, psychologist Daniel Kahneman, describes the workings of the human brain in his bestselling book: *Thinking, Fast and Slow.*[3] As Kahneman tells it, the mental process is divided into two parts, which he labels "System 1" and "System 2." System 1 is the simple, automatic and effortless, intuitive approach based on gut feeling. Intuition is deeply rooted in experiences and feelings from a lifetime; it is lightning-quick—and, most of the time, it is correct. In our distant past, this decision style helped us survive. For example, if a lion threatened you, you had to decide immediately: fight, flight, or freeze. Intuition differs from instinct, which is the genetically preprogrammed behavior found in all members of a particular species (**Figure 3**). Yet, the remarkable fact is that our brain decides what to do as early as several seconds before we become consciously aware of it. Thus, we may ask questions about our assumptions of free will.

The alternative is System 2, the analytical decision strategy, wherein a systematic, logical, rational, and calculated

Figure 3. Animal instincts and decisions. Decisive responses exhibited by animals are primarily instinctive. Birds, modern descendants of the relatively small-brained dinosaurs, have evolved remarkably complex brains. As a result, some crows can flexibly make profitable decisions in tool-using tasks, similar to humans.[4] Diorama and photo: Atle Østern

assessment is made that is time-intensive and deliberate. Even so, emotions remain crucial for analytical decision-making as well. It is generally good to keep a cool head, but even minor decisions are difficult to reach without an emotional push. The link is in the brain's ventromedial prefrontal cortex, which is involved in both decision processes and the regulation of feelings.

In reality, the two systems are not exclusive. Rather, they are distinct yet related styles used by all people, depending on the situation and personality traits.

How to proceed

Aviation safety was famously revolutionized by the introduction of checklists, a powerful tool in many situations. They are now implemented in healthcare—for instance, to ensure "safe surgery."[5] A procedure, similar to a checklist in some ways, is more extensive. In my opinion, such schemes, whether written or informal, are fundamental to all good decisions and task performances in one way or another.

There are many decision strategies we can draw upon.[6] Consider the following examples:

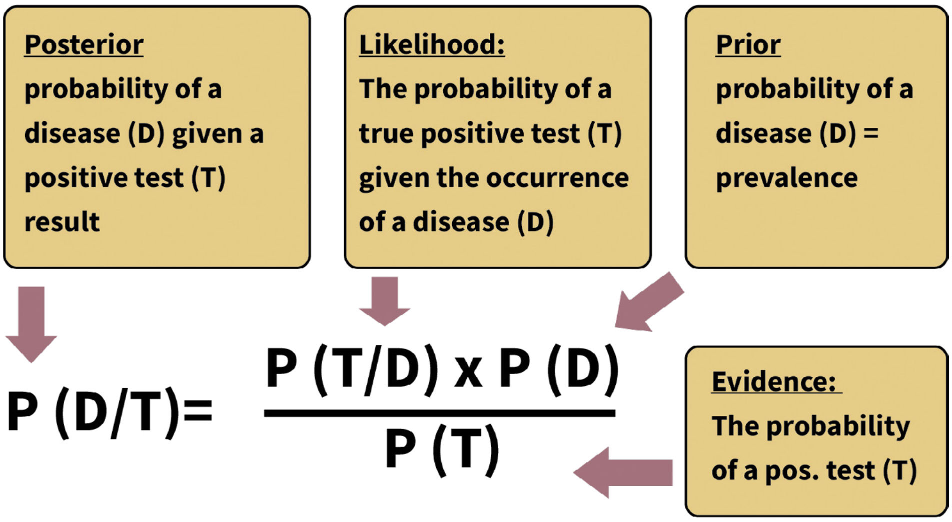
- **Exclusion method**: The least likely hypotheses are eliminated through pairwise comparisons.
- **Pattern recognition (or associative model)**: The automatic identification of perceived inputs that match internalized pre-existing patterns.
- **Multiple criteria decision analysis (MCDA)**: An additive sum of weighted scores for different criteria.
- **Flow charts (algorithmic model)**: A diagrammatic representation of a step-by-step approach to solve a task.
- **Even swaps**: A method for making trade offs through sequential elimination of alternative solutions by equalizing the values of one selected measure in exchange for adjusting another.
- **Pro et contra**: The well-known list of arguments for and against an option.

Each of these decision strategies has advantages and limitations. Although we might not know the procedure by name, pattern recognition is likely the most commonly used method among ophthalmologists. It is fast, and its accuracy increases with experience—but the risk of bias or noise is relatively high. Measurable tests, calculations, and quantitative scores reduce or erase noise. However, a problem with MCDA is that pattern noise may emerge if the criteria are not interpreted equally by all users. Flow charts visualize the decision-making process, and a great many exist for various eye conditions. Flow charts can be a great aid in decision-making, but they are difficult to recall and you may not always have access to them.[7]

I will now describe my strategy, as of 2022, for adopting System 2. It is certainly not the only or necessarily the best solution, and you may disagree with it. My aim is to inspire (or provoke) you to think through your strategy.

A suggested strategy for ophthalmology

1. **Medical interview (a thorough history-taking):** What is the (real) problem (root cause analysis)? Why has the patient asked for help (chief complaint)? A systematic template for potential questions is of great assistance regarding possible vision loss/disturbance or pain/discomfort. Ask when the symptoms started (chronology), whether the onset was sudden or gradual, whether the symptoms have varied over time, and what their relationships are to other factors. Screen, then elaborate. What does the patient understand, feel, and expect?
2. **Systematic physical examination:** Gather diagnostic clues—for instance, by asking myself, "is the cornea transparent (yes or no)?" I use the abbreviation "ABC" to remember to determine the Anatomical location, Be descriptive of and Classify the pathological conditions (e.g., infiltrates, exudates).
3. **Diagnostic hypothesis generation:** What are the alternative explanations? A starting point can be a crude division of conditions and anatomical sites that fit in with voiced pain/discomfort (inflammation, infection, ischemia, injury, increased intraocular pressure), photophobia (keratopathy, uveitis), and vision loss (refractive error, opaque media, retinal disease, optic neuropathy, brain disorders), but also different pathophysiological categories. I also ask, "what is the most likely (expressed as a quantified probability when possible; see **Figure 4**), severe and treatable diagnosis I should consider?" Resist premature intuitions before balanced information has been obtained and the evidence considered.
4. **Diagnostic tests:** Decide on supplementary workups to gather appropriate data if the results can change the management, based on an estimate of the prior probability of a particular disease. Assess their accuracy (predictive values, sensitivity, and specificity). Stop when you have enough relevant information.
5. **Rank the different alternatives in order of plausibility:** I often apply a combination of the exclusion method, pattern recognition, multiple criteria analysis, flow charts (occasionally), and finally, pro et contra until one or a few diagnoses remain. Pathognomonic signs (such as jaw claudication in giant cell arteritis) are rare. Then, combine diagnoses with the same consequences and judge whether one diagnosis may explain everything (rule of parsimony). Regarding treatments, I try to calculate the predictive outcomes. For example, what are the benefits of treatment compared to the harms involved? Secondly, what is the risk (calculated as the multiplication of the probability of an event by its severity) of doing nothing compared to the risk of doing something? Establish the purpose or goals (these should be SMART—i.e., specific, measurable, assignable/achievable, relevant and time-bound). Prioritize and determine the sequence of actions (**Figure 5**). Finally, design a structured and feasible plan A (and B).
6. **Evaluation of the main hypothesis:** Are the symptoms, findings, and tests compatible with the diagnosis and treatment? Consider fallacies. Have I checked all clues and interpreted them correctly (constructive "Socratic questioning")? Repeat, if necessary, some of steps 1–5. Review reliable textbooks, guidelines, or meta-analyses. Discuss cases with trustworthy colleagues (second opinions). At last, revise your decisions and plan.
7. **Shared decision-making:** Explain to the patient the medical problem and treatment options at their level of understanding. Assess the patient's attitude regarding the risks and benefits (utility) and the treatment threshold probability (the level of certainty of a disease for starting treatment). Make individual adjustments. Get informed consent. Treat, test more, or observe. Agree on a follow-up.



	Sick (D)	Not sick	Sum
T+	170	7,830	8,000
T-	30	91,970	92,000
Sum	200	99,800	100,000

P (D) = 200/100.000 = 0.002
P (T/D) = 170/200 = 0.85
P (T) = 8,000/100,000 = 0.08
P (D/T) = 0.021 = 2%

Figure 4. Bayes' theorem (or its derived odds ratio form) is a widely used method to calculate the conditional probability of events on a scale of 0 to 1. The interpretation of a test result, as new information becomes available, depends on the probability of a disease before the test.[8] Brown boxes explain the formula; blue boxes show an example.
Figure: Atle Østern

Medical decision-making is supremely complex and challenging because of the inherent uncertainty and difficulty in making accurate predictions in real time. The key is to estimate the probability before and after acquiring new information (**Figure 4**). In addition, the patient may have two or more diseases with conflicting and confusing findings—or one condition may be connected to another. For example, chickenpox during childhood can later manifest as herpes zoster ophthalmicus, possibly due to another triggering illness, with simultaneous or time-delayed neurotrophic keratopathy, secondary bacterial keratitis, uveitis, and acute retinal necrosis. To make matters worse, some clinical cases diverge from the typical presentations or prior experiences. The clinician must sort out all these possibilities, acting like a "medical detective."[10]

As shown, the decision-making process involves many choices, and it is helpful to take the outside view of the problems and break them down into several smaller, manageable tasks. Use your creativity as well to organize and plan ahead. The PDSA cycle refers to the first letters in Plan (choose a method), Do (intervene), Study (reassess), and Act (modify based on reassessment). It

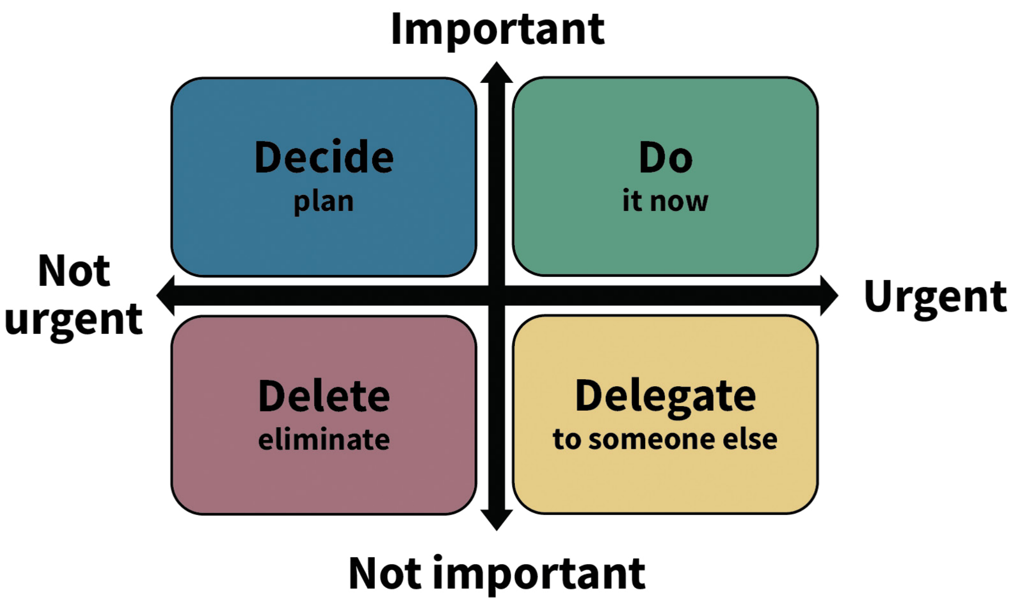


Figure 5. Prioritization: General (and later President) Eisenhower introduced the above matrix, now named after him.[9] Another solution is to assign letters to tasks: A (absolutely urgent), B (becomes necessary), C (can do), D (delegate), E (eliminate), and F (file). Scoring systems are increasingly applied to sort patients according to their treatment needs (like triage).
Figure: Atle Østern

is a problem-solving model for testing and executing change that is also employable during follow-ups. In the end, you should adhere to your values and principles and make sure that your conclusions fall within relevant frameworks (e.g., laws and financial restraints).

Critical thinking improves the quality of healthcare. For example, noise in a clinic can be measured by an audit set up by many professionals who independently judge the same cases (actual or fictitious). Thus, this evaluation can disclose deficiencies in skill and training at the workplace, necessitating altered guidelines and educational activity. Finally, a growth mindset in the doctor can enable a change of habits. One way is to recall an event, review whether the proper procedure was pursued, mentally replay a revised version, and think (procedure selection) before the task (process).

What will the future bring?

A radical solution is the replacement of judgment with rules or algorithms that eliminate noise. Artificial intelligence (AI) heralds a future where computers can assist and even out-compete humans (**Figure 6**). Even today, they are better than ophthalmologists at assessing diabetes retinopathy. Thus, the quality of ophthalmologists' decision-making strategies may become even more critical in the future. As a result, the role of doctors may change. Nevertheless, doctors can detect stupid mistakes made by AI, inspire more trust and satisfaction, resolve misunderstandings, and adjust treatment plans following feedback from patients. For the foreseeable future, most vulnerable patients will probably prefer—and need—empathetic experts to advise and help them.[11]

Let us not forget who we are. We are *Homo sapiens*, meaning "wise man" in Latin, with the means to make good decisions. Above all, the human eyes and brain are still the best tools to observe and perceive the eyes and brains of others.

Still, somehow, we failed in this respect in August 2021.

References

1. Hesselberg J-O & Reinholdt R. *Bedre beslutninger*. Kagge forlag. 2022
2. Kahneman D et al. *Noise*. 2021
3. Kahneman D. *Thinking, Fast and Slow*. 2011
4. Miller R et al. Decision-making flexibility in New Caledonian crows, young children and adult humans in a multi-dimensional tool-use task. *PLOS*. 2020 Mar. https://doi.org/10.1371/journal.pone.0219874
5. Gawande A. *The Checklist Manifesto*. Profile Books. 2011
6. Samset K & Wenstøp F. *Smarte valg*. NKI forlaget. 2016
7. Zwaan J. *Decision Making in Ophthalmology*. Jaypee Brothers Medical Publishers. 2014
8. Sox H et al. *Medical Decision Making*. Wiley-Blackwell. 2013
9. Krogerus M & Tschäppeler R. The Decision Book. Profile Books. 2011
10. Trimble M & Hamilton P. The thinking doctor: Clinical decision making in contemporary medicine. *Clinical Medicine* 2016 Vol 16; No 4: p. 343-6
11. Østern A. Fremtidens oftalmologer: To be or not to be? *Oftalmolog*. 2017

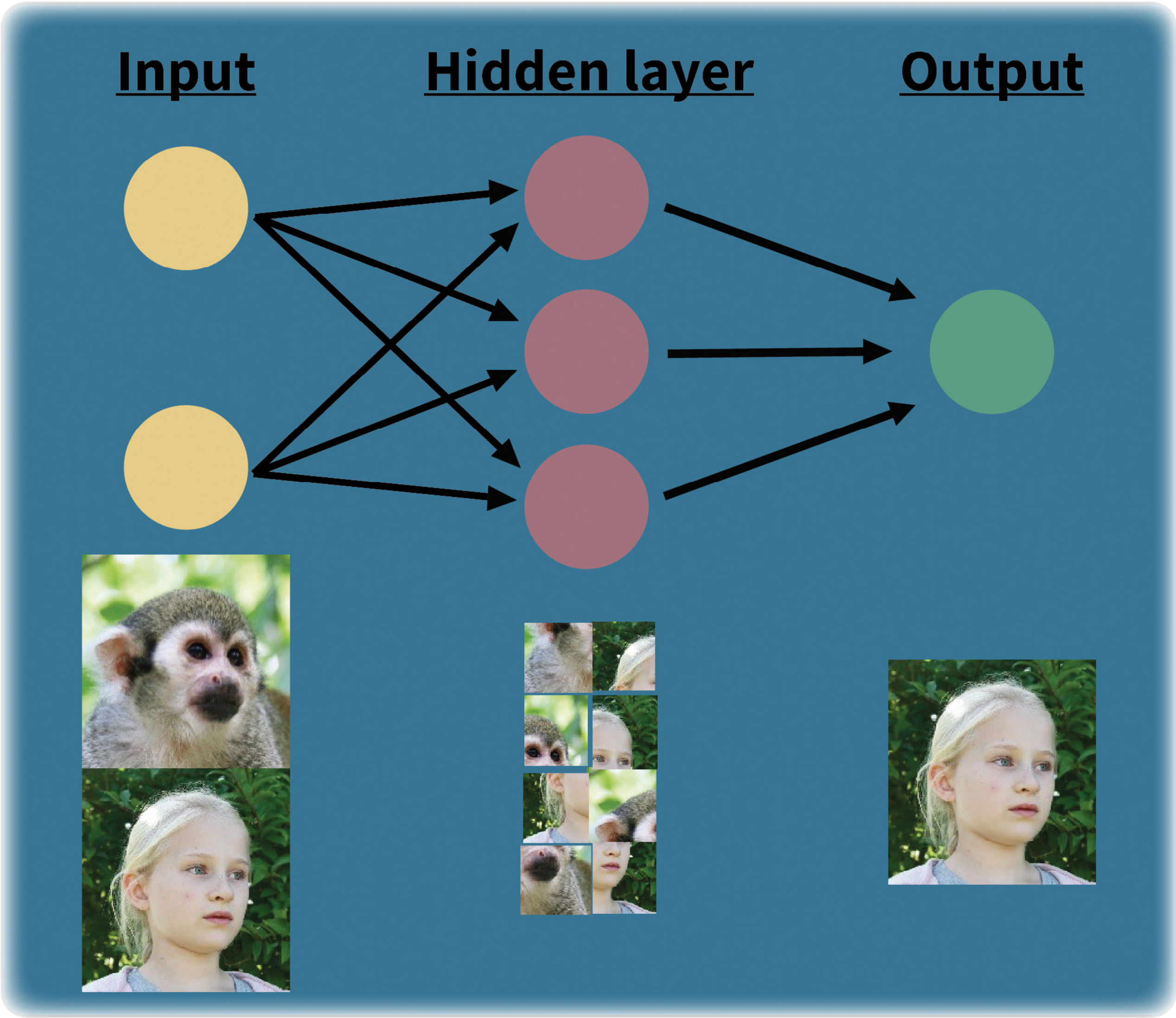


Figure 6. Artificial intelligence and machine learning. Through a hierarchical structure of neural networks with coupled nodes, the software will, over time, find patterns as a result of training and deep learning. It can, in the end, distinguish images of different categories.
Illustration and photos: Atle Østern