2024 ■ Vol. 20 ■ no. 1

# A Compendium of Common Heuristics, Misconceptions, and Biased Reasoning used in Statistical Thinking

Tracy Husereau [a], Denis Cousineau [a], Shuo Zang [a] & Rose-Marie Gibeau [a]

[a]Université d'Ottawa

**Abstract** ■ Over the past decades, many researchers have identified ways of reasoning in the domain of statistics and probabilities that do not match statistics and probabilities results. Some of these inadequate conceptualizations are reviewed herein. They include, among others, the gambler's fallacy, the law of small numbers, the misunderstanding of randomness, and they touch various aspects of statistics reasoning (sampling procedures, probability estimation, mean estimation, variance estimation, and inference). A classification is put forward.

**Keywords** ■ Statistics reasoning; heuristics; cognitive biases; misconceptions.

denis.cousineau@uottawa.ca

10.20982/tqmp.20.1.p057

**Acting Editor** ■ Denis Cousineau (Université d'Ottawa)

*This editorial presents current challenges with regards to statistics reasoning.*

> Statistical thinking will one day be as necessary for efficient citizenship as the ability to read and write.
>
> *Wilks, 1951 (cited in Snee, 1990, where it is incorrectly attributed to H. G. Wells)*

## Introduction

When it comes to making decisions, individuals can rely on heuristic processes, lean on misconceptions, or be affected by cognitive biases. These shortcuts in decision-making are acquired through exposure and learning, or they might be indicators of innate biases; in any case, they become embedded in thinking mechanisms (Tversky & Kahneman, 1974; Bronner, 2023). **Heuristics** are practical and quick rules of thumb for solving a problem. Heuristics may be easy to use but often leads to erroneous outcomes (Kahneman, 2011; Liu, 2019). **Misconceptions** comprise adequate reasoning. As these conceptions are incorrect or incomplete, the outcome is likely to be erroneous. **Cognitive biases** influence the reasoning process used to reach a response. All three of these limitations in reasoning significantly affect our ability to take proper course of actions which can result in various errors (Gilovich et al., 2002). They also tint individuals' learning processes and as a consequence hinder their learning trajectory.

Although these limitations exist for all aspects of decision-making and reasoning, the present paper looks at situations where the information given has a certain degree of uncertainty. This is specifically the case for *statistics reasoning*, where decision-makers have to gauge a situation based on limited information. Through statistics reasoning processes, one has to I) examine information only accessible through *samples*, which are II) entrenched by *variability*, from which one III) *estimates* an underlying trend that leads one to IV) *infer* about a future issue or a proper course of action. These four categories will be used to organize the subsequent sections (see also Table 1 at the end).

What is called hereafter *limitations* are not to be seen as something negative. Indeed, they can be quite beneficial. Heuristics are an important time and energy saver (Kahneman, 2011); misconceptions have sociological utility (Bronner, 2023); and cognitive biases have evolutionary relevance (Pinker, 2021). However, in training professionals and citizens alike, it is necessary to identify these limitations so that trainees are aware of these internal representations and processes that may influence the adequate resolution of a problem. Also, this awareness may allow trainers to devise activities that foster the desired ways of reasoning. By understanding learners' incorrect or simplistic ways of reasoning, instructors can adopt a more adequate approach to teach decision-making by integrating more formal statistical reasoning and critical thinking with regards to limited information (Gigerenzer & Gaissmaier, 2011; Pohl, 2017; Watson et al., 2003).

This compendium has been developed with multiple subsidiary objectives in mind as well: (i) this compendium can be used to minimize the impact of misconceptions, biases, and heuristics on decision-making processes to meet the standards of scientific research; (ii) it can also be used as a roadmap to develop new statistics vignettes aimed specifically at debunking these limitations in trainees (Béland, 2020).

A *compendium* of inadequate ways of reasoning in statistics is presented herein. It is a short but hopefully fairly complete enumeration of the known limitations. A two-way classification is proposed in the discussion. Note that the purpose of this document is not to offer causal explanations of these limitations. However, it might be useful in future research on statistics understanding and for the development of statistics *assessments* (Garfield et al., 2002). Understanding the common reasoning limitations will help with devising tests and questionnaires to better evaluate the acquired abilities of test takers (Garfield & Ben-Zvi, 2004; Garfield, 2003; Allen, 2006; DelMas et al., 2007; Gibeau & Cousineau, 2024).

This paper could have included limitations related to numeracy, but these will not be discussed for two reasons. Firstly, it is believed that numeracy is a different set of skills related to, but distinguishable from, statistics reasoning (the latter could be coined *statisticy*). Indeed, statistics reasoning can make fairly good predictions without formal computations. If we only want to know an approximate estimation, little numeracy skills are required, as long as the concepts needed to reach a solution are well understood. Secondly, let's not forget that statistics are often confused with mathematics (Stuart, 1995) but are, in fact, different (Zeidner, 1991). This confusion is mostly apparent in *statistics anxiety* which is partly inherited from preexisting *mathematics anxiety* (Gibeau et al., 2023). Mathematics is often a predictor or a precursor of statistics just like it is for physics: they both end up being packed with equations predicting the outcomes of systems, yet no one confuses physics with mathematics (Cousineau & Harding, 2017).

## I: Sampling

Sampling is the process of observing a part of a population. It is through this process that observations become available on which subsequent judgments are based. Sampling can be formally defined and systematically achieved (e. g., *simple randomized sampling*, *clustered randomized sampling*, or *stratified sampling*; Kish, 1965). It can also be loosely performed (using techniques called *convenience sampling*, *snowball sampling*, etc.; Bhardwaj, 2019), or be based on one's own observations or word of mouth without varying the sources or worse, be based on echo chambers (Chavalarias, 2022). As listed below, many limitations affect how a sample is understood, undermining the primary source of information on which statistical reasoning is built.

### *The Law of Small Numbers*

The *law of small numbers* is a misconception regarding the representativeness of a small sample with respect to the whole population. It is based on the implicit belief that a small sample closely resembles the population (Tversky & Kahneman, 1971). Considering that individuals are influenced by this misconception from the start, it will propagate to and contaminate any subsequent decisions, such as inductions and generalizations (see *Representativeness heuristic*in Section IIIa). In a sense, decision-makers mistakenly map the implications of the *law of large numbers* onto small samples, and as a result demonstrate unjustified confidence in the validity of conclusions drawn from small samples (see Section IV). This misconception gives exaggerated representativeness to cases that could be anecdotal. In other words, decision-makers tend to ignore the influence of sample size when it is small (see Section II; Reyna & Brainerd, 2008). It is therefore quite close to the *Base rate neglect* effect (see Section IV).

In one problem, Tversky and Kahneman (1971, p. 105) asked the attendants of two psychology conferences the following question:

*Suppose you have run an experiment on 20 subjects, and have obtained a significant result which confirms your theory ($z = 2.23$, $p < .05$, two-tailed). You now have cause to run an additional group of 10 subjects. What do you think the probability is that the results will be significant, by a one-tailed test, separately for this group?*

The median estimate given by the attendants was 85% whereas the correct answer is closer to 48%. The authors argue that the participants had an exaggerated belief in the probability of replicating the findings because they had an exaggerated belief in the first sample's representativeness.

### *The Expectancy of Local Representativeness*

The law of small numbers can be characterized by the *expectancy of local representativeness* whereby short runs of events should closely espouse characteristics of large runs. As an example, Konold et al. (1993) presented the following problem:

*Which of the following sequences is least likely to result from flipping a fair coin 5 times?*

*a) H H H T T*
*b) T H H T H*
*c) T H T T T*
*d) H T H T H*
*e) All four sequences are equally likely.*

The correct answer is *e* (selected by 38% of the participants). The response choice that was the least selected (hence, the sequence that was judged the most likely; selected by only 2 participants out of 79, or 2.5%) was sequence *b* because it seems *random*. In other words, this sequence has the expected characteristics of a long run of heads and tails, that is, *i*) close to 50% of heads (compared to option *c* selected by 22.8% of the participants), ii) fairly frequent alternations between heads and tails (compared to option $d$ selected by 29.1% of the participants and option a selected by 8% of the participants). Tversky and Kahneman (1974) speculate that people have a mental representation of what randomness looks like, and this representation determines their understanding of this concept (also see Alberoni, 1962; Tune, 1964; Wagenaar, 1970; R. S. Nickerson, 2002).

### *The Fairness of Randomness*

This misconception assumes that the sampling process is fair or "self-correcting". If you obtained an unexpected run of events, then randomness will compensate in the subsequent few runs (as if randomness had memory). Tversky and Kahneman (1971, p. 106), presented this problem to participants:

*The mean IQ of the population of eighth graders in a city is known to be 100. You have selected a random sample of 50 children for a study of educational achievements. The first child tested has an IQ of 150. What do you expect the mean IQ to be for the whole sample?*

The correct response is 101 but "a surprisingly large number of people" answered 100 (p. 106). This incorrect answer highlights the hidden assumption that the remaining observations in the sample will "correct" for the unexpected first observation. As noted by R. S. Nickerson (2002), samples do not self-correct; as the sample grows, the unexpected early observations will only become diluted in the subsequent ones. The problem's correct answer, 101, is an illustration of this dilution effect (the 50 points of IQ in excess of the first participant is diluted over the sample).

Instead of attributing an exaggerated representativeness to the first observation (here 150), the participants simply do not take it into account. One possible reason for this neglect is that they have expectations about the observations (they expect the mean to be 100). This is in opposition to the law of small numbers where the observations, taken from an unspecified population, are weighted very strongly.

After interrogating a large sample of community-based participants ($n = 1559$), Bronner (2023) noted large response differences based on the educational level achieved in the two-hospital problem (see section *Representativeness heuristic* presented later in Section IIIa). He noted that participants with post-secondary diplomas were more likely to select a response congruent with the fairness of randomness assumption. Hence, the notion that randomness has some regularity might be overgeneralized in this educated sub-group of participants. In French, a well-known expression captures this overgeneralized conception: "Le hasard fait bien les choses". It loosely translates to “randomness does things right”.

### *The Gambler’s Fallacy*

The *gambler’s fallacy*, which may be related to the fairness of randomness, results from the misconception that past outcomes of a random event influence future outcomes. For example, if a particular outcome has occurred frequently in the recent past, it is less likely to occur in the future. If someone flips a coin and gets "head" several times in a row, they may believe that "tail" is now "due" and more likely to occur (although each time the probability of getting heads remains 50% because each flip is independent; Tversky and Kahneman, 1974, p. 1130).

A related effect is the *negative recency* effect (Jarvik, 1951). When a participant is exposed to a random sequence of A’s and B’s, they must guess the next letter. The results of such a problem show that after a single occurrence of A, participants rate its probability of occurring a second time higher. The probability of occurrence decreases the more A occurs. The initial increase in estimated probability might be an instance of the law of small numbers (i.e., *believe that a small sample closely resembles the population*) and the subsequent decrease might be an instance of the fairness of randomness (i.e., *assuming that assumes that the sampling process is fair or "self-correcting"*).

### *Sampling Processes Preferences*

It is generally acknowledged that simple randomized sampling is the gold standard in order to get unbiased estimations (Kish, 1965; Thompson, 2012). Most statistical procedures are based on this assumption. Clustered sampling provides weaker estimates (i.e., they have reduced precision and wider confidence intervals) and the effect of stratified sampling on the precision of estimates is unknown.

To see how these techniques are appraised, Jacobs (1997) interviewed 69 children aged between 10 to 12, using two scenarios. Here is the first one:

*The school is an elementary school with grades 1 through 6 and 100 students in each grade. A fifth grade class is trying to raise some money to go on a field trip to Great America (an amusement park). They are considering several options to raise money and decide to do a survey to help them determine the best way to raise the most money. One option is to sell raffle tickets for a SEGA video-game system. Consequently, nine different students each conducted a survey to estimate how many students in the school would buy a raffle ticket to win a SEGA. Each survey asked 60 students but each sampling method and results were different. The nine surveys and their results were as follows:*

*(a) Raffi asked 60 friends (75% yes, 25% no);*
*(b) Shannon got the names of all 600 kids in the school, put them in a hat, and pulled out 60 of them (35% yes, 65% no);*
*(c) Spence had blond hair, so he asked the first 60 kids he found who had blond hair too (55% yes, 45% no);*
*(d) Jake asked 60 kids at an after-school meeting of the Games Club. The Games Club met once a week and played different games – especially computerized ones. Anyone who was interested in games could join (90% yes, 10% no);*
*(e) Abby sent out a questionnaire to every kid in the school and then used the first 60 that were returned to her (50% yes, 50% no);*
*(f) Claire set up a booth outside the lunchroom and anyone who wanted to could stop by and fill out her survey. To advertise her survey she had a sign that said "WIN A SEGA". She stopped collecting surveys when she got 60 completed (100% yes);*
*(g) Brooke asked the first 60 kids she found whose telephone number ended in a 3 because 3 was her favorite number (25% yes, 75% no);*
*(h) Kyle wanted the same number of boys and girls and some kids from each grade. So he asked 5 boys and 5 girls from each grade to get his total of 60 kids (30% yes, 70% no);*
*(i) Courtney didn’t know too many boys so she decided to ask 60 girls. But she wanted to make sure she got some young girls and some older ones so she asked 10 girls from each grade (10% yes, 90% no);*

Results show that about one third of the participants were aware of potential bias in the non-random samples (found in all except option 2) and recognized the advantages of random sampling procedures. Yet, having to choose a preferred method, almost 40% chose stratified sampling (e.g., survey 8). One-quarter of those seemed to prefer stratified sampling because they were concerned about equity issues ("it's not nice to the people who are not his friends. They want to answer the survey too but they aren't allowed") or mistrusted the overreliance on randomness to get observations. Furthermore, about 10% preferred samples that were congruent with their expectations (see *Confirmation bias* in section IV) or the ones that were decisive. Decisiveness may be a form of *Need for structure* whereas congruence with expectations may be a form of *Confirmation bias* (see section IV).

### *The Selection Issue*

Gigerenzer (1991) argues that when an observation is selected from a sample for further illustration, it goes against the assumption that this observation was randomly selected. The author considers the following problem from Kahneman and Tversky (1973, pp. 241-242):

> *A panel of psychologists have interviewed and administered personality tests to 30 engineers and 70 lawyers, all successful in their respective fields. On the basis of this information, thumbnail descriptions of the 30 engineers and 70 lawyers have been written. You will find on your forms five descriptions, chosen at random from the 100 available descriptions. For each description, please indicate your probability that the person described is an engineer, on a scale from 0 to 100.*
> *Jack is a 45-year-old man. He is married and has four children. He is generally conservative, careful, and ambitious. He shows no interest in political and social issues and spends most of his free time on his many hobbies which include home carpentry, sailing, and mathematical puzzles. The probability that Jack is one of the 30 engineers in the sample of 100 is ______ %.*

In the original study, a second group of participants was told that the base rates were 70 engineers and 30 lawyers (as opposed to 30 engineers and 70 lawyers). They found nearly identical mean responses in both groups (71% and 81% respectively) irrespective of the base rate. Tversky and Kahneman instead suggested that this behavior reflect a base rate neglect bias (discussed in section IV).

For Gigerenzer (1991), the critical assumption in this problem is *random selection*. If the description presented in the problem was not selected randomly, but was selected because of its peculiarities, the base rates of engineers and lawyers would become irrelevant. In a follow-up study, Gigerenzer (1991) examined whether the presence of a single word “random” (line 5 of the problem) was sufficient to make the participant think that the case selected was really selected at random. In a control group, he made participants fully aware of random sampling in this problem by inviting them to draw by themselves from an urn a description, read it, and express their thoughts. This condition reduced the base-rate neglect effect (mean responses of 64% when the urn was said to contain 30 engineers and 70 lawyers vs. 80% when the urn was said to contain 70 engineers and 30 lawyers).

### II: Variability

Variability has been identified as a foundational concept on which to build statistical reasoning (e.g., Wild & Pfannkuch, 1999). Many studies examining the concept of variability presented frequency distribution plots (i.e., histograms). Meletiou-Mavrotheris and Lee (2005) noted that "bumpier" plots were rated as having more variability. Kaplan et al. (2014) observed common misunderstandings about these plots, in particular that "flatter" histograms were rated as having less variability, congruent with Meletiou-Mavrotheris and Lee (2005). Studies using plots (such as Watson et al., 2003) were not included herein since plot literacy goes beyond the scope of the present paper.

### *Population Variability*

Landwehr (as reported in Shaughnessy, 1992, p. 35, the actual report is unavailable so that it is not possible to know how these claims were substantiated) proposed the following two misconceptions related to variability: "(1) any difference in the means between two groups is significant; (2) there is no variability in the 'real world'" (1 is actually a direct consequence of 2). Points (1) and (2) are essentially synonymous and suggest that variability is severely underestimated. These misconceptions would be related to population variability. However, no study on this subject was found.

### *Sampling variability*

Shaughnessy et al. (2004, pp. 178-179) presented the following problem to children in grades 6 to 12:

*Suppose you have a container with 100 candies in it. 60 are red, and 40 are yellow. The candies are all mixed up in the container. You pull out a handful of 10 candies.*
*How many reds do you expect to get? ____*
*Would you expect the same number every time? ____*
*How many reds would surprise you in a handful of ten? ____*

Most respondents (82%) answered 6 for the first question, which matches the mathematical expectation. Only 8 participants (3%) provided a range of possibilities. The second question explicitly tests if participant could envision a range response; 78% said that they expected the same number every time (Shaughnessy et al., 2004, interpreted this as a misconception caused by statistics classes and its insistence on *expectation* which is a single number). For the final question, 18% gave a response between 4 and 8 inclusively, indicating that few sampling variation around 6 is anticipated by these participants.

Based on participants' written justifications, the authors propose three types of statistics reasoning: *additive*, *proportional*, and *distributional* justifications. Additive justifications tend to rely on absolute numbers or frequencies of red candies (e.g., "*Because there are more reds*"). Proportional justifications are based implicitly or explicitly on proportions, probabilities, or percentages (e.g.,"*Most samples of 10 should contain around 6 reds, but I just can't explain why*"). Finally, distributional justifications rely on centers and the variation around them (e,g,, "*There should be approximately 6 but not precisely 6 all the time as 4 is very possible as well*"). The last two categories of justifications were observed in approximately a quarter of the participants only.

### Insensitivity to Sample Size in Sampling Distribution

The sampling distribution of a statistic is influenced by sample size. For example, the mean of a handful of individuals is less reliable than the mean of a large collection of individuals taken in similar conditions. (Kahneman & Tversky, 1972) examined how participants estimate sampling distribution with problems such as:

*On what percentage of days will the number of boys among 1000 babies be as follows:*

- *Up to 50 boys*
- *50 to 150 boys*
- *150 to 250 boys*
- *... [we abbreviate the whole list of options]*
- *More than 950 boys*

*Note that the categories include all possibilities, so your answers should add up to about 100%.*

The above problem is framed using a total $N$ of 1000 babies, but this number $N$ was varied between groups (it could be 100 or 10). The response options were always divided into 11 segments. The present problem assumes a binomial population; another problem uses the average height of $N$ males which is based on a normal population. The authors estimated the frequency distribution from the participant's judgments. For example the participants reported that 2% of the days had only one boy among a sample of 10. Notably, the authors found that the frequency distribution was unaffected by the problem's sample size $N$: Observing a single boy in a sample of 10, 5-to-15 boys in a sample of 100, or 50-to-150 boys in a sample of 1000 had the same probability whereas the correct probabilities are actually 1%, 10-11% and 10-117% for $N$ of 10, 100 and 1000, respectively.

### Probabilistic and Deterministic Reasoning

Participants in Pfannkuch and Brown's (1996) study seemed to forget probabilistic thinking when facing problems with real-world contexts, justifying their response by deterministic causes related to the context. However, when context was removed, participants used probabilistic thinking to solve the problem.

As an example of a real-world context, ten New Zealand psychology undergraduates where shown the map problem:

*Every year in New Zealand, approximately seven children are born with a limb missing. Last year the children born with this abnormality were located in New Zealand as shown on the map [A map with the country's six regions is shown with the number of limb-less new born indicated in each, varying from 0 to 3]. What do you think?*

These numbers are very small as approximately 10,000 babies are born each year in each region. Interpretations given by the participants were based on farming, pollution, genetics, but not on random sampling. To interpret this deterministic explanation preference, instead of random variability explanation, the authors suggest that participants were unaware of the random dimension of such phenomena. This may be a case of the law of small numbers (i.e., exaggerated representativeness given to cases that could be anecdotal).

**IIIa: Probability and Frequency Assessment**

Sections IIIa and IIIb are dedicated to problems of estimation. Estimation concerns probabilities and frequencies, central tendencies (e.g., the mean) or higher statistical moments (e.g., variance, skew and kurtosis). To our knowledge, this last estimation category has never been examined, which is why the present paper only looks at estimation of probabilities and frequency (in this section) and central tendencies (in the next section). As probability estimation plays a central role in Tversky and Kahneman's seminal work, a whole section is dedicated to it.

Probabilities can be estimated by observations, using a mental model or a mathematical model (here, we distinguish the mental model from the mathematical model, assuming that the former is less formal than the latter; Heath & Tversky, 1991). Computing probabilities is a well-developed branch of mathematics; its teaching involves, for example, urn problems (urns containing red and black balls and the problem is to compute the probability of certains draws) which are a nightmare to many students in social sciences programs (Liu, 2019). The present paper focuses on probabilities estimated by observations using mental models.

***The Outcome Approach***

Konold (1989, 1995) explains that students may see probabilities as something binary (i.e., Yes/No, 0%/100%, Never/Always, i.e., equiprobability for two classes only). Also, when "asked for a probability of some event, people reasoning according to the outcome approach do not see their goal as specifying probabilities that reflect the distribution of occurrences in a sample, but as predicting results of a single trial" (Konold, 1995, p. 2). This so-called "outcome approach" first put forth a single outcome then considers its chance of occurrence using a binary judgment. For example, consider the weather problem that he presented to 16 undergraduates (Konold, 1995, p. 3):

*The Springfield Meteorological Center wanted to determine the accuracy of their weather forecasts. They searched their records for those days when the forecaster had reported a 70% chance of rain. They compared these forecasts to records of whether or not it actually rained on those particular days.*
*The forecast of 70% chance of rain can be considered very accurate if it rained on:*
*a) 95% - 100% of those days.*
*b) 85% - 94% of those days.*
*c) 75% - 84% of those days.*
*d) 65% - 74% of those days.*
*e) 55% - 64% of those days.*

In this study, approximately one third of the participant gave the correct answer ($d$). However, nearly the same number of participants responded ($a$). Konold (1995) hypothesized a grey zone around 50% where events are seen as undetermined. "Given the desire for predictions, outcome-oriented individuals translate probability values into yes/no decisions. A value of 50% is interpreted as total lack of knowledge about the outcome, leaving one no justification for making a prediction" (Konold, 1995, p. 2). This is reminiscent of the Need for structure described later.

Cousineau and Harding (2017) considered an alternative explanation whereby probabilities are understood correctly, but what is misunderstood is the notion of randomness. Using the inverse weather problem, the authors argue that what is difficult is making a future prediction using probabilities, not estimating probabilities in such simple problems.

*The Canadian Meteorological Center observes parameters (humidity, wind, etc.) in order to predict tomorrow's weather. For tomorrow, however, the computers cannot reach a prediction as to whether it will rain or not. Consulting the 2013 to 2023 archives, a meteorologist determines that on 70% of the occurrences with identical parameters, it rained on the subsequent day.*
*According to you, what should the Meteorological Center announce for tomorrow:*
*a) a 95% to 100% risk of rain?*
*b) a 85% to 95% risk of rain?*
*c) a 75% to 85% risk of rain?*
*d) a 65% to 75% risk of rain?*
*e) a 55% to 65% risk of rain?*

(the years in the problem were changed to match the date this text has been written). The results showed more accurate responses in the inverse problem relative to the original formulation (using a slider ranging from 0% to 100%, the choices were 30% closer to the right answer; Husereau, 2021).

### *Representativeness Heuristic*

The representativeness heuristic is one of the major three heuristics identified by Tversky and Kahneman (1974, 1973, 1974). It is the tendency to judge the probability of a specific outcome based on past events (Tversky & Kahneman, 1974). Indeed, representativeness heuristic uses the similarity of a situation, population or individual to imagine an "ideal type", and then judge the probability of occurrence according to the occurrence's proximity to that ideal type. In other words, people estimate the probability of an event on the basis of its similarity to past events, rather than relying on statistical data and reference values. People predict the outcome that appears most representative in the circumstances (Kahneman & Tversky, 1973; Tversky & Kahneman, 1983). In a study, Kahneman and Tversky (1972, p. 443) show the hospital problem to participants:

*A certain town is served by two hospitals. In the larger hospital about 45 babies are born each day, and in the smaller hospital about 15 babies are born each day. As you know, about 50 percent of all babies are boys. However, the exact percentage varies from day to day. Sometimes it may be higher than 50 percent, sometimes lower.*
*For a period of 1 year, each hospital recorded the days on which more than 60 percent of the babies born were boys. Which hospital do you think recorded more such days?*
*- The larger hospital*
*- The smaller hospital*
*- About the same (that is, within 5 percent of each other)*

Twenty-three participants out of 95 answered the large hospital whereas 21 answered the small hospital (and 53 answered that it was the same in both hospitals), with the correct answer being the smaller hospital. This result is coherent with the insensitivity to sample size noted earlier.

### *Availability Heuristic*

This heuristic is defined as estimating the frequencies or probabilities of events based on the ease with which one thinks about these events or can find instances in memory (Tversky & Kahneman, 1974). The individual tends to over-estimate the probability of an event occurring if they can easily recall an instance of it happening and to underestimate it if instances of it happening are not easily recalled.

In Tversky and Kahneman (1973, pp. 211-212), the following problem is presented:

*The frequency of appearance of letters in the English language was studied. A typical text was selected, and the relative frequency with which various letters of the alphabet appeared in the first and third positions in words was recorded. Words of less than three letters were excluded from the count.*
*You will be given several letters of the alphabet, and you will be asked to judge whether these letters appear more often in the first or in the third position, and to estimate the ratio of the frequency with which they appear in these positions.*
*Consider the letter R. Is R more likely to appear in*
*- the first position?*
*- the third position?*

As it is easier to think about words beginning with "R", the authors argue that the participants' estimation is higher in this case. This argument is supported by their data (69% of participants chose the first option) but goes against actual frequencies.

***The Equiprobability Bias***

The weather problem given above suggests that there is a point where predictions switch from a *certain yes* to a *certain no*. This point seems to be located arbitrarily at 50%. Thus, when asked to make predictions based on probabilities, individuals look at percentages as yes/no statements on a certain outcome, both having equiprobability. R. S. Nickerson (2002, p. 332), indicates that "[p]erhaps the most widely accepted assumption of randomness is that of constant equiprobability of the possible outcomes. In the case of the rolling of a dice, it is assumed that each of the six possible outcomes has the same chance of occurring as all the others on each roll".

**IIIb: Descriptive Statistics Assessment**

This section considers more broadly the estimation of summary statistics.

***Anchoring Bias***

The anchoring bias is the tendency to rely on the first piece of information encountered or on an already known value when making decisions. For example, with statistics, people may anchor their decisions on a reference statistic without taking into account subsequent relevant data. This bias can distort statistical analysis, as people will stick to their initial impression, even when facing evidence that suggests otherwise.

Within the domain of mathematics, Tversky and Kahneman (1973, p. 215) used the following problem:

*We asked subjects to estimate, within 5 sec, a numerical expression that was written on the blackboard, One group of subjects ($N = 87$) estimated the product 8 × 7 × 6 × 5 × 4 × 3 × 2 × 1, while another group ($N = 114$) estimated the product 1 × 2 × 3 × 4 × 5 × 6 × 7 × 8.*

The median estimate for the descending sequence was 2,250 whereas the median estimate for the ascending sequence was 512. The correct answer of both sets is 40,320. The results were explained by the fact that within 5 seconds, only a few terms had been multiplied, and that partial multiplication was used by participants to extrapolate a response.

A similar problem in Tversky and Kahneman (1974, p. 1128) was presented. In this problem, participants must estimate the percentage of African countries that are members of the United Nations:

*A number between 0 and 100 was determined by spinning a wheel of fortune in the subjects' presence. The subjects were instructed to indicate first whether that number was higher or lower than the value of the spinner, and then to estimate the percentage of African countries by moving upward or downward from the given number.*

The median estimates of the percentage of African countries in the United Nations were 25% and 45% for groups that saw 10 and 65 on the spinner, respectively (the correct answer in 2023 is 30% of sovereign countries). Tversky and Kahneman (1973) later included this bias as part of the availability heuristics.

***Confusion between precision and accuracy***

The statistical term "*accuracy*" is for many students indistinguishable from "*precision*". Bar-Hillel and Falk (1982) found in their experiments that students described as "accurate" samples whose descriptive statistics exactly matched the population parameter. The precision of a statistic describes its expected variation from the population parameter; its magnitude is a function of the variation in the population and the sample size. Students would tend to confuse the meaning of precision with the everyday meaning of the word. Thus, they consider as "precise" or "accurate" samples whose statistics correspond exactly to the population parameter.

***Failure to Notice Regression to the Mean***

Regression to the mean in a phenomenon whereby, following an extreme event, the next random event is likely to be less extreme. Tversky and Kahneman (1982b) provided the following example:

*An athlete has an excellent season - the commentators congratulate him, he is voted athlete of the year - but his next season is bad. Critics and commentators put forward all sorts of reasons to explain this performance, but it may be that the last season was a fluke and that the athlete has now regressed to a level close to the average.*

This problem shows that critics and commentators may miss the notion of regression to the mean and so have no way to recognize when it may be a plausible explanation. Instead, they invoke deterministic reasoning to identify "causes" to the observed changes (as seen above in Section II).

Kahneman, during the Nobel Memorial prize in Economic sciences 2002 noticed ironically "I understood an important truth about the world: because we tend to reward others when they do well and punish them when they do badly, and because there is regression to the mean, it is part of the human condition that we are statistically punished for rewarding others and rewarded for punishing them".

According to the gambler's fallacy, the next event should "compensate for" or "even out" the previous events. Thus, this fallacious belief might have evolved in the humain species to implement regression to the mean in our reasoning tools.

***Illusory Correlation***

Chapman (1967) uses the term "illusory correlation" to explain superstition, belief in primitive magic, errors in clinical observations, etc. This bias happens when one perceives a high rate of co-occurrence between two classes of events that share some similarities but are, in reality, not co-occurring at a higher than chance rate. Chapman's (1967) original study showed 20 word pairs that were either related (e.g., *lion* and *tiger*) or unrelated (e.g., *head* and *paper*). Participants were given the first word of each pair and were asked to provide the second word as well as to report how often the first word was paired with the second word by estimating their pairing percentage. Related word-pairs were more correctly completed and the two words were estimated as more paired together. Tversky and Kahneman (1973, p. 224) reproduced the finding using words describing personality traits (some being thought as naturally associated, e.g., alert and witty), where traits that are believed associated were judged as being co-occurring more frequently. The authors interpreted this finding as another example of the availability heuristic: as such pairs are more easily thought jointly and retrieved more accurately, their prevalence is judged higher.

**IV: Inference**

Inference can be seen as the process through which one makes a decision after data has been examined. Decisions in the dominant statistics paradigm used in social sciences are performed by considering the data under an assumed model; the Bayesian framework adds pre-existing knowledge under the form of priors. In decision making, priors can be seen as a mental model of a situation that may support or interfere with the evaluation of the observations.

***Non-symmetrical Inferences***

Tversky and Kahneman (1977, pp. 2-3) considered the following problem:

*Which of the following events is more probable?*
*a) That an athlete won the decathlon, if he won the first event in the decathlon*

*b) That an athlete won the first event in the decathlon, if he won the decathlon*
*c) The two events are equally probable*

The correct answer is ($c$). The prior probability that an unspecified athlete wins the decathlon and that an unspecified athlete wins the first event is both $1/N$, where $N$ is the number of competitors. Consequently, the two conditional probabilities are equal. More participants chose ($b$) (63%) compared to response choices ($a$) or ($c$) (13% and 23%, respectively).

It might be argued in the present example that there is a temporal order in the events described which would explain why one answer is selected more than the other. To counteract this objection, the authors examined another, similar, problem (the IQ problem, pp. 2-4) in which there is no temporal relation and yet, the same general pattern of results was found.

The misconception highlighted by the problem above may be related to the illusion of validity (see below), as making predictions using the more redundant information is associated with greater over-confidence. Alternatively, this could be a case of representativeness as it may be easier to think about a decathlon winner than to think about an athlete who won the first event.

### *Correlation is not Causality*

There seems to be a tendency to over-interpret co-occurrences by searching for a causal explanation. This does resemble the preference for deterministic reasoning over statistical reasoning that we examined in Section II. However, we did not find any article on this subject related to statistical reasoning. An example (**z23**) says (erroneously) that in the United States, eating more ice cream increases the number of Shark attacks. While the correlation is almost perfect when comparing monthly statistics, the causation is wrong. A more probable explanation is that warmer temperatures invite more people to eat ice cream and to bathe in the ocean, a prerequisite to be attacked by a shark.

### *The Conjunction Fallacy*

The conjunction fallacy happens when one believes that two events are more likely to occur together than separately, despite the impossibility of this being true. The "Linda problem" is an example illustrating this bias (Tversky & Kahneman, 1982a):

*Linda is 31, single, very bright and not afraid to express her ideas. Her studies in philosophy led her to become interested in issues of discrimination and social justice. She has also taken part in protests against fossil fuels. Based on this description, we ask which of the following situations is the most likely:*
*a) Linda is a banker*
*b) Linda is a banker and a feminist*

The results indicate that 85% of participants thought answer b is more likely, even though, mathematically speaking, a has a higher probability of being true, since to obtain b, a must be true. Thus, 85% of participants committed a conjunction fallacy. Kahneman and Tversky (1973) attributed this result to the representativeness heuristic because instances with a more elaborate description are easier to recall.

Fiedler (1998), following Gigerenzer (1991), notes that when this problem is asked using frequencies (see below), the conjunction fallacy effect tends to disappear (from 85% of participants choosing b to 22%). These two authors argue that this decrease shows that the conjunction fallacy cannot be attributed to the representativeness heuristic but may have to do with a distinction between estimated confidence and estimated frequency (also see the overconfidence bias).

*The total sample included 100 people. To how many persons out of this sample can the following statements be applied?*

### Base Rate Neglect

This bias occurs when an individual ignores or minimizes information regarding prevalence in favor of individual information (Kahneman & Tversky, 1973). Turpin et al. (2020) propose the following problem (p. 385), inspired from Kahneman and Tversky (1973, p. 241):

> *In a study 1000 people were tested. Among the participants there were 5 engineers and 995 lawyers. Jack is a randomly chosen participant of this study.*
> *Jack is 36 years old. He is not married and is somewhat introverted. He likes to spend his free time reading science fiction and writing computer programs.*
> *What is the likelihood that Jack is a Lawyer?*

The results show a strong base rate neglect, where participants judged Jack more likely to be an engineer, despite the weak chance of choosing one randomly.

Turpin et al. (2020) argue that this problem has two components, some baseline prevalence (995 out of a 1000), but also a "causal" explanation (a mental model providing a strong belief that personality is related to career choice). When participants have a mental model providing an explanation, they would see no use to examine baseline information.

Tversky and Kahneman (1982b, p. 9) presented a similar problem in which the mental model becomes explicit:

> *A cab was involved in a hit and run accident at night. Two cab companies, the Green and the Blue, operate in the city. You are given the following data:*
> *a) 85% of the cabs in the city are Green and 15*
> *b) a witness identified the cab as Blue. The court tested the reliability of the witness under the same circumstances that existed on the night of the accident and concluded that the witness correctly identified each one of the two colors 80% of the time and failed 20% of the time.*
> *What is the probability that the cab involved in the accident was Blue rather than Green?*

The correct answer is 41%. However, the median answer given by participants is close to 80%, a value which coincides with the credibility of the witness (the mental model being explicit) and is less affected by the relative frequency of blue and green cabs (the baseline information).

Gigerenzer (1991) argues that the base rate neglect in the *Jack* problem can be nullified by controlling the selection issue (seen in Section I). He then considered the following problem (from Tversky & Kahneman, 1982b, p. 154):

> *If a test to detect a disease whose prevalence is 1/1000 has a false positive rate of 5%, what is the chance that a person found to have a positive result actually has the disease, assuming you know nothing about the person's symptoms or signs?*

The correct Bayesian answer is 2% but half of the participants responded 95%; the average across participants is 56%. Cosmides and Tooby (1996, p. 24) rephrased the problem using a frequentist formulation:

*One out of l000 Americans has disease X. A test has been developed to detect when a person has disease X. Every time the test is given to a person who has the disease, the test comes out positive. But sometimes the test also comes out positive when it is given to a person who is completely healthy. Specifically, out of every l000 people who are perfectly healthy, 50 of them test positive for the disease.*
*Imagine that we have assembled a random sample of 1000 Americans. They were selected by a lottery. Those who conducted the lottery had no information about the health status of any of these people. How many people who test positive for the disease will actually have the disease? ___ out of ___.*

Now more than half of the participants (56%) responded 2%; the average result is 5.8%. Gigerenzer used this result to further argue that judgments of frequencies are confused with judgments of confidence when asked to provide a probability. However, it could also be a case of confusion between test specificity and test sensibility (Section IIIb).

### Confirmation Bias

The confirmation bias asserts that individuals try to corroborate a prediction that confirms their conceptions rather than to look for counter-evidence that would invalidate the prediction (R. R. Nickerson, 1998).

Wason (1966) developed the card problem (for variations, see Wason & Shapiro, 1971; Pinker, 2021):

*The cards that follow have a letter on one side and a number on the other side. Please consider this rule: IF THERE IS A VOWEL ON ONE SIDE OF A CARD, THEN THERE IS AN EVEN NUMBER ON ITS OTHER SIDE. Which of the following four cards should you turn over to test that the rule is true?*
*E K 4 7*

Whereas the correct answer is "E" and "7" (chosen by 4% of participants), 46% choose "E" and "4" (failure to look for a counter-evidence) and 33% choose "E" only.

Griggs and Cox (1982, p. 415; also see Johnson-Laird & al., 1972) presented a more ecologically valid version of the problem:

*On this task, imagine that you are a police officer on duty. It is your job to ensure that people conform to certain rules. The cards in front of you have information about four people sitting at a table. On one side of a card is a person's age and on the other side of the card is what the person is drinking. Here is a rule: IF A PERSON IS DRINKING BEER, THEN THE PERSON MUST BE OVER 19 YEARS OF AGE. Select the card or cards that you definitely need to turn over to determine whether or not the people are violating the rule.*
*DRINKING A BEER, DRINKING A COKE, 16 YEARS OF AGE, 22 YEARS OF AGE*

In this version, 70% participants got the correct answer. One possible explanation for the low performance in the original version is that participants may have had a tendency to process implication relations ($A \implies B$) as if they were equivalence relations ($A \iff B$; Bronner, 2023). A more accepted explanation however is that people tend to confirm a rule rather than search to invalidate it (R. S. Nickerson, 2002).

### The Illusion of Validity

People may endorse a decision when there is a match between the information provided and the outcome considered. The illusion of validity is when the information matches the outcome of a mental model held by the person (e.g., a stereotype) without requiring them to consider base rate information or any other factor limiting predictability (Tversky & Kahneman, 1974). This illusion is strengthened when the information provided is homogeneous or redundant. As an example, Tversky and Kahneman, (1974, p. 1126) argued that:

> *People express more confidence in predicting the final grade-point average of a student whose first-year record consists entirely of B's than in predicting the grade-point average of a student whose first- year record includes many A's and C's.*

### *Overconfidence Bias*

Related to the above, another cognitive bias lies in how people rate their confidence in given that were given. The *overconfidence bias* refers to the tendency to have strong confidence in one's own judgments. Gigerenzer (1991, p. 87) considered the following example:

> *Which city has more inhabitants?*
> *a) Hyderabad,*
> *b) Islamabad*
> *How confident are you that your answer is correct?*
> *(check one of 50% 60% 70% 80% 90% 100%)*

In all the cases where subjects said, "I am 100% confident that my answer is correct," the relative frequency of correct answers was only about 80%, for an overconfidence effect of 20%. In Gigerenzer (1991, Exp. 1, $n = 80$), the mean confidence is 67% whereas the percent correct is only 52%. This difference between expressed confidence in correct responses and actual frequency of correct responses (here 15%) is the overconfidence effect.

Koriat et al. (1980) proposed that the overconfidence bias is caused by a "confirmation bias" where, after a response option is chosen, one searches for information that confirms the answer given without searching for information that could falsify it.

Gigerenzer (1991) examined the overconfidence bias by adding the following question at the end of 50 items similar to the one above:

> *How many of these 50 questions do you think you got right?*

In this version, participants must make a frequency judgment rather than a confidence judgment. The mean frequency judgment was 52% (i.e., the mean response was 26.0), matching well the actual rate of correct response (53%). Thus, in this version, the overconfidence bias seems to disappear.

### *Need for Structure*

According to Kruglanski and Ajzen (1983), the need for structure is the desire to have "any knowledge, as opposed to a state of ambiguity" (p. 16). The authors did not document this bias with a formal study. Nonetheless, it is added to the present paper as it provides some emphasis on the notion of mental model, that is, any belief that can be used to understand observations or make predictions about future observations.

An example of the need for structure can be found in Null Hypothesis Statistical Testing (NHST). In these procedures, it is well-known that a failure to reject $H_0$ is to be interpreted as an absence of evidence. Yet, it is typically, and falsely, interpreted as an acceptation of H0. Much research shows that this incorrect interpretation influences not only students, but researchers and statisticians as well. Thus, the overwhelming prevalence of this tendency may reveal a common cognitive bias. In Howell's words, "the problem of how to interpret a non-rejected null hypothesis has plagued students in statistics courses for over 75 years" (Howell, 2010, p. 93).

## General Discussion

The present paper examined various misconceptions, heuristics and biases that impede the use of formal reasoning when facing uncertainty. All these limitations were judged to be incorrect reasoning by Tversky and colleagues. To determine what is correct and what is incorrect, a set of formal reasoning rules must be used. Tversky always assumed that the correct reasoning rules were given by Bayesian procedures. Yet, nothing guarrantee that Bayesian rules are applicable to realy-life assessment of risk and odds. Gigerenzer (1991) argued that the human mind is poorly equipped for Bayesian estimation but that it is fully capable of performing frequentist estimation. With this assumption, he and others found that problems expressed with absolute frequencies (e.g., "Out of every 1000 people who are perfectly healthy, 50 of them...") are more accurately responded than problems expressed with relative frequencies (e.g., "5% of people..."). More recently, Liu (2019) argued that Bayesian procedures may have an immanent bias: there has to be a formal model given to all alternatives. The general convention that "uninformative priors" can be equated to a uniform distribution may actually return very wrong results. In particular, it fails to see that in situations where nothing is known, the worst-case scenario may actually be underlying the computations of frequencies (worst-case inference was also explored in Cousineau, 2020, using a different approach).

A two-way post-hoc classification is proposed. The first dimension concerns the locus of influence (sampling, variability, estimation, or inferences), the second being about the nature of the cause:

- Misconceptions happen when knowledge is incorrect or incomplete. For example, the incorrect belief that randomness is self-correcting is a misconception.
- Heuristics are observed when simple reasoning processes are used instead of correct processes. Basing a decision on how easy it is to retrieve instances (representativeness heuristic) is a heuristic.
- Cognitive biases happen when correct processes are not performed correctly. For example, the last number seen may bias approximation judgments.

Table 1, which illustrate the two-way classification, is most probably an illusory conjunction of ideas triggered by a need for structure but it is a first attempt at categorizing what affects (or prevents) formal reasoning.

Shaughnessy (1977) summarizes the studies on judgments under uncertainty by arguing that the representativeness heuristic can account for: I) insensitivity to prior probabilities and population proportions (the base rate neglect; Section IV); II) insensitivity to the effects of sample size (Section II); III) unwarranted confidence in a prediction that is based upon invalid input data (Section IV); IV) misconceptions of chance, such as the Gamblers' Fallacy and the fairness of randomness (Section I); and V) misconceptions about the tendency of data to regress to the mean (Section IIIb). This heuristic was among Tversky and Kahneman's three most important ones, along with the availability heuristic and the anchoring bias (also in Section IIIb).

The present compendium also suggests the necessity of a mental model; the effects described above may not exist if no mental model is present. Its role is highlighted in I) the expectancy of local representativeness; II) the base rate neglect; III) the illusion of validity; and IV) the need for structure. However, it may be underlying other limitations as well.

Finally, Table 1 indicates that more than one limitation can be activated simultaneously. Thus, cumulative effects and possible interactions should be examined. For example, it is suggested that the negative recency effect may be composed of the law of small numbers and the gambler's fallacy. By having an exhaustive view of all the known limitations, some useful integrative synthesis might be found that would simplify what currently is a plethora of diverse phenomenon. By correcting these misconceptions, biases and heuristics as well as their impact on decision making, individuals may strive to adopt a better approach that will promote better decisions.

## Authors' note

The authors would like to thank Jesika Walker for comments on an earlier version of this text.

## References

Alberoni, F. (1962). Contribution to the study of subjective probability, part 1. *Journal of General Psychology*, *66*(2), 241–264. doi: 10.1080/00221309.1962.9711840.

Allen, K. (2006). *The statistics concept inventory: Development and analysis of a cognitive assessment instrument in statistics [doctoral dissertation, university of oklahoma]*. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.2130143

Bar-Hillel, M., & Falk, R. (1982). Some teasers concerning conditional probabilities. *Cognition*, *11*(2), 109–122. doi: 10.1016/0010-0277(82)90021-x.

Béland, S. (2020). Introducing the teaching quantitative methods vignettes (tqmv). *The Quantitative Methods* *for Psychology*, *16*(2), v1–v3. doi: 10.20982/tqmp.16.2.v001.

Bhardwaj, P. (2019). Types of sampling in research. *Journal of the Practice of Cardiovascular Sciences; 5*, 157–63. doi: 10.4103/jpcs.jpcs_62_19.

Bronner, G. (2023). *L'empire de l'erreur: Biais cognitifs et société* (Vol. 260).

Chapman, L. J. (1967). Illusory correlation in observational report. *Journal of Verbal Learning and Verbal Behavior*, *6*(1), 151–155. doi: 10.1016/S0022-5371(67)80066-5.

Chavalarias, D. (2022). *Toxic data: Comment les réseaux manipulent nos opinions*. Flammarion.

Cosmides, L., & Tooby, J. (1996). Are humans good intuitive statisticians after all? rethinking some conclusions from the literature on judgment under uncertainty. *Cognition*, *58*(1), 1–73. doi: 10.1016/0010-0277(95)00664-8.

Cousineau, D. (2020). How many decimals? rounding descriptive and inferential statistics based on measurement precision. *Journal of Mathematical Psychology*, *97*(102362), 1–99. doi: 10.1016/j.jmp.2020.102362.

Cousineau, D., & Harding, B. (2017). Pourquoi les statistiques sont-elles difficiles à enseigner et à comprendre? *Quelques réflexions [Why statistics is difficult to teach: A few considerations]. Revue de Psychoéducation*, *46*(2), 397–419. doi: 10.7202/1042257ar.

DelMas, R. C., Garfield, J., Ooms, A., & Chance, B. (2007). Assessing students' conceptual understanding after a first course in statistics. *Statistics Education Research Journal*, *6*(2), 28–58. http://www.stat.auckland.ac.nz/serj

Fiedler, K. (1998). The dependence of the conjunction fallacy on subtle linguistic factors. *Psychological research*, *50*(2), 123–129. doi: 10.1007/BF00309212.

Garfield, J., & Ben-Zvi, D. (2004). Research on statistical literacy, reasoning, and thinking: Issues, challenges, and implications. In Ben-Zvi & J. Garfield (Eds.), *D* (pp. 397–409). Kluwer Academic Publishers.

Garfield, J., delMas, R., & Chance, B. (2002). *The assessment resource tools for improving statistical thinking (artist) project*. Retrieved March 19, 2024, from https://app.gen.umn.edu/artist/]

Garfield, J. B. (2003). Assessing statistical reasoning. *Statistics Education Research Journal*, *2*(1), 22–38. doi: 10.52041/serj.v2i1.557.

Gibeau, R.-M., & Cousineau, D. (2024). *The key statistical concepts from an investigation using four standardized tests* (tech. rep.).

Gibeau, R.-M., Maloney, E. A., Béland, S., Lalande, D., Cantinotti, M., Williot, A., & Cousineau, D. (2023). The correlates of statistics anxiety: Relationships with spatial anxiety, mathematics anxiety and gender. *Journal of Numerical Cognition*, *9*(1), 16–43. doi: 10.5964/jnc.8199.

Gigerenzer, G. (1991). How to make cognitive illusions disappear: Beyond "heuristics and biases". *European Review of Social Psychology*, *2*(1), 83–115. doi: 10.1080/14792779143000033.

Gigerenzer, G., & Gaissmaier, W. (2011). Heuristic decision making. *Annual Review of Psychology*, *62*, 451–82. doi: 10.1146/annurev-psych-120709-145346.

Gilovich, T., Griffin, D., & Kahneman, D. (2002). Cambridge University Press. doi: 10.1017/CBO9780511808098.

Griggs, R., & Cox, J. (1982). The elusive thematic-materials effect in wason's selection task. *British Journal of Psychology*, *73*, 407–420. doi: 10.1111/j.2044-8295.1982.tb01823.x.

Heath, C., & Tversky, A. (1991). Preference and belief: Ambiguity and competence in choice under uncertainty. *Journal of Risk and Uncertainty*, *4*, 5–28. doi: 10.1007/BF00057884.

Howell, D. C. (2010). *Statistical methods for psychology (7th edition)*. Wadsworth.

Husereau, T. (2021). *La compréhension des probabilités chez les pharmaciens et les étudiants universitaires* [Doctoral dissertation, Unpublished undergraduate thesis report].

Jacobs, V. R. (1997). Children's understanding of sampling in surveys.

Jarvik, M. E. (1951). Probability learning and a negative recency effect in the serial anticipation of alternative symbols. *Journal of Experimental Psychology*, *41*(4), 291–291. doi: 10.1037/h0056878.

Johnson-Laird, P. N., Legrenzi, P., & Legrenzi, M. S. (1972). Reasoning and a sense of reality. *British Journal of Psychology*, *63*, 395–400. doi: 10.1111/j.2044-8295.1972.tb01287.x.

Kahneman, D. (2011). *Thinking, fast and slow*. Farrar. doi: 10.1007/s00362-013-0533-y.

Kahneman, D., & Tversky, A. (1972). Subjective probability: A judgment of representativeness. *Cognitive Psychology*, *3*(3), 430–454. doi: 10.1016/0010-0285(72)90016-.

Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. *Psychological review*, *80*(4), 237–251. doi: 10.1037/h0034747.

Kaplan, J. J., Gabrosek, J. G., Curtiss, P., & Malone, C. (2014). Investigating student understanding of histograms. *Journal of Statistics Education*, *22*, 2–30. doi: 10.1080/10691898.2014.11889701.

Kish, L. (1965). *Survey sampling*. Wiley.

Konold, C. (1989). Informal conceptions of probability. *Cognition and Instruction*, *6*(1), 59–98. doi: 10.1207/s1532690xci0601_3.

Konold, C. (1995). Issues in assessing conceptual understanding in probability and statistics. *Journal of Statistics Education*, *3*(1), 1–9. doi: 10.1080/10691898.1995.11910479.

Konold, C., Pollatsek, A., Well, A., Lohmeier, J., & Lipson, A. (1993). Inconsistencies in students' reasoning about probability. *Journal for Research in Mathematics education*, *24*(5), 392–414. doi: 10.5951/jresematheduc.24.5.0392.

Koriat, A., Lichtenstein, S., & Fischhoff, B. (1980). Reasons for confidence. *Journal of Experimental Psychology: Human learning and memory*, *6*(2), 107–107. doi: 10.1037/0278-7393.6.2.107.

Kruglanski, A. W., & Ajzen, I. (1983). Bias and error in human judgment. *European Journal of Social Psychology*, *13*(1), 1–44. doi: 10.1002/ejsp.2420130102.

Liu, B. (2019). Uncertain urn problems and ellsberg experiment. *Soft Computing*, *23*, 6579–6584. doi: 10.1007/s00500-018-3306-7.

Meletiou-Mavrotheris, M., & Lee, C. (2005). *Exploring introductory statistics students' understanding of variation in histograms*. In Proceedings of the 4th Congress of the European Society for Research in Mathematics Education.

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

Nickerson, R. S. (2002). The production and perception of randomness. *Psychological Review*, *109*, 330–357. doi: 10.1037/0033-295X.109.2.330.

Pfannkuch, M., & Brown, C. (1996). Building on and challenging students' intuitions about probability: Can we improve undergraduate learning? *Journal of Statistics Education*, *4*(1), 1–19. doi: 10.1080/10691898.1996.11910502.

Pinker, S. (2021). *Rationality: What it is, why it seems scarce, why it matters*. Penguin books.

Pohl, R. F. (2017). *Cognitive illusions: Intriguing phenomena in thinking, judgment and memory*. Routledge. doi: 10.4324/9781315696935.

Reyna, V. F., & Brainerd, C. J. (2008). Numeracy, ratio bias, and denominator neglect in judgments of risk and probability. *Learning and Individual Differences*, *18*, 89–107. doi: 10.1016/j.lindif.2007.03.011.

Shaughnessy, J. M. (1977). Misconceptions of probability: An experiment with a small-group, activity-based, model building approach to introductory probability at the college level. *Educational Studies in Mathematics*, *8*(3), 295–316. doi: 10.1007/BF00385927.

Shaughnessy, J. M. (1992). Research in probability and statistics: Reflections and directions. In D. A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning: A project of the national council of teachers of mathematics* (pp. 465–494). Macmillan Publishing Co.

Shaughnessy, J. M., Ciancetta, M., & Canada, D. (2004). Types of student reasoning on sampling tasks. *International Group for the Psychology of Mathematics Education*, *4*, 177–184.

Snee, R. D. (1990). Statistical thinking and its contribution to total quality. *The American Statistician*, *44*(2), 116–121. doi: 10.2307/2684144.

Stuart, M. (1995). Changing the teaching of statistics. journal of the royal statistical society. *Series D (The Statistician)*, *44*(1), 45–54. doi: 10.2307/2348615.

Thompson, S. K. (2012). *Sampling (3rd edition)*. Wiley.

Tune, G. S. (1964). Response preferences: A review of some relevant literature. *Psychological Bulletin*, *61*(4), 286–302. doi: 10.1037/h0048618.

Turpin, M. H., Meyers, E. A., Walker, A. C., et al. (2020). The environmental malleability of base-rate neglect. *Psychonomic Bulletin and Review*, *27*, 385–391. doi: 10.3758/s13423-020-01710-1.

Tversky, A., & Kahneman, D. (1971). Belief in the law of small numbers. *Psychological Bulletin*, *76*(2), 105–110. doi: 10.1037/h0031322.

Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology*, *5*(2), 207–232. doi: 10.1016/0010-0285(73)90033-9.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, *185*(4157), 1124–1131. doi: 10.1126/science.185.4157.1124.

Tversky, A., & Kahneman, D. (1977). Causal schemas in judgments under uncertainty. In *Progress in social psychology* (pp. 49–72). Lawrence Erlbaum; associates.

Tversky, A., & Kahneman, D. (1982a). Evidential impact of base rates. In P. S. Kahneman & A. Tversky (Eds.), *D* (pp. 153–160). Cambridge University Press. doi: 10.21236/ADA099501.

Tversky, A., & Kahneman, D. (1982b). Judgments of and by representativeness. In P. S. Kahneman & A. Tversky (Eds.), *D* (pp. 84–98). Heuristics; Biases, . Cambridge University Press.

Tversky, A., & Kahneman, D. (1983). Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological Review*, *90*(4), 293–315. doi: 10.1037/0033-295X.90.4.293.

Wagenaar, W. A. (1970). Subjective randomness and the capacity to generate information. *Acta Psychologica*, *33*, 233–242. doi: 10.1016/0001-6918(70)90135-6.

Wason, P. C. (1966). Reasoning. In *New horizons in psychology* (pp. 59–99). Penguin Books.

Wason, P. C., & Shapiro, D. (1971). Natural and contrived experience in a reasoning problem. *Quarterly Journal of Experimental Psychology*, *23*, 63–71. doi: 10.1080/00335557143000068.

Watson, J. M., Kelly, B. A., Callingham, R. A., & Shaughnessy, J. M. (2003). The measurement of school students' understanding of statistical variation. *International Journal of Mathematical Education in Science and Technology*, *34*, 1–29. doi: 10.1080/0020739021000018791.

Wild, C. J., & Pfannkuch, M. (1999). Statistical thinking in empirical enquiry. *International Statistical Review*, *67*(3), 223–248. doi: 10.1111/j.1751-5823.1999.tb00442.x.

Wilks, S. S. (1951). Undergraduate statistical education. *Journal of the American Statistical Association*, *46*(253), 1–18. doi: 10.1080/01621459.1951.10500763.

Zeidner, M. (1991). Statistics and mathematics anxiety in social science students: Some interesting parallels. *British Journal of Educational Psychology*, *61*, 319–328.

### Citation

Husereau, T., Cousineau, D., Zang, S., & Gibeau, R.-M. (2024). A compendium of common heuristics, misconceptions, and biased reasoning used in statistical thinking. *The Quantitative Methods for Psychology*, *20*(1), 57–75. doi: 10.20982/tqmp.20.1.p057.



Received: 19/03/2024

Table 1 follows.

**Table 1** ■ Some classification of the limitations to statistics reasoning

| | Nature of the limitation | | |
|---|---|---|---|
| Locus of the limitation | Heuristics | Misconceptions | Cognitive biases |
| I. Sampling | | Law of small number<br>Expectancy of local representativeness<br>The selection issue<br>Fairness of randomness | Sampling process preference<br><br><br>The gambler's fallacy |
| II. Variability | | Population variability<br>Sampling variability<br>Insensitivity to sample size | Deterministic reasoning |
| III. Estimation | The outcome approach<br>Representativeness heuristic<br>Availability heuristic<br>Anchoring bias | The equiprobability bias | Failure to notice regression to the mean<br>Illusory correlation<br>Confusion precision/exactness |
| IV. Inference | | Non-symmetrical inferences<br>Correlation is not causality | Conjunction fallacy<br>Base rate neglect<br>Confirmation bias<br>The illusion of validity<br>The overconfidence bias<br>Need for structure |