# 5 OPERA SINGERS AND INFORMATION THEORY

## 5.1 THE GAME

| | |
|---|---|
| Number of players: | 5–20, plus 1 game show host with complete knowledge of what is going on. |
| You will need: | a stage divided into 2 separate areas, left and right, and 1 off-stage area (backstage or frontstage) |

The players represent a group of opera singers. Among them are two stars. These stars are either friends or enemies. Only the host knows who the stars are and the nature of their relationship (friends or enemies). So in this game, not even the two stars know each other: The stars do not know that they are stars themselves, they do not know who the other star is, and they do not know their relationship. Neither do the other singers know. The singers' common goal is to identify who the stars are and whether they are friends or enemies. To do so, they send two groups of singers to the different areas on the stage. The group on the left represents the singers from the first half of the opera, the group on the right from the second half. The remaining players in the backstage area do not take part in that opera performance. The host has complete knowledge of who's who and now reveals some information on the players' identities. He does so by providing feedback on each of these "performances" by classifying each performance as a "big success," "disaster," or "neutral," according to the following two rules:

1. When no stars or only one star appears on stage, the critics will not even write about the opera, and the host will classify the performance as "neutral."
2. If both stars appear on stage, the critics will be excited, and the opera performance will not be considered neutral. If the stars are friends and appear in the same half or if the stars are enemies and appear in different halves, the critics will declare the opera to be a "big success." Otherwise (that is, if the friends sing in different halves or the enemies sing together), the opera will be deemed a "disaster."

Let us consider two examples: In the first example, there are two enemies (called E1 and E2), but only one of them is on stage, as shown in the figure below on the left. Here, the outcome is neutral. In the second example, there are two friends (F1 and F2) singing in the same half, as shown on the right. This opera is a big success.


Outcome: neutral Outcome: big success

After receiving the critique, the group decides whom to send on stage next. The group of players is asked to develop a strategy that allows them to find out as quickly as possible who the stars are and whether they are friends or enemies. Fortunately, so many operas with two halves have been composed that there are no constraints on the number of singers in each group: The numbers of singers can differ between the two halves (as in the example above) and between performances. One or two of the groups of singers may even be empty.

*What is the optimal strategy for 5 singers? More precisely, which strategy requires the lowest number of performances, even in the worst case? Is there an easy strategy that comes close to optimal? What about the optimal strategy for 100 singers?*

For simplicity, let us first look at the example with 5 singers. It does not even seem obvious how to find a strategy that can be described easily. We can, for example, select a pair of singers, schedule them for the first half of the opera, and leave the second half empty. We can then repeat this strategy for all pairs of singers. In most cases, at least one of the stars will be left outside, and the outcome of the performance will be neutral. If, however, the outcome is not neutral, we immediately know that this pair of singers must be the pair of stars we are looking for, and we also know whether they are friends (the opera is a success) or enemies (it is a disaster). In the worst case, we have a nonneutral outcome only for the last pair of singers. Given that there are 5 singers in total, how many pairs of singers are there? There are 5 possibilities to choose the first singer of the pair and 4 possibilities to choose the second singer from the remaining singers, which gives $5 \cdot 4 = 20$ pairs. This, however, counts every pair exactly twice, since for forming a pair, it does not matter which singer was chosen first and which one second. In total, there are thus $5 \cdot 4/2 = 10$ pairs. This number is often denoted by $\binom{5}{2} = 10$ (see appendix B.4: What Is … a Binomial Coefficient?).

Thus, with this strategy, we obtain the correct answer after at most 10 performances.

This strategy, however, is not optimal. One of the explanations for this is that each performance carries only a little "information" on average (we will make this term precise later), since most of the performances will be considered neutral. One possibility to improve on the above strategy is to look at 2 pairs at the same time; that is, to have one pair in the first half and another in the second half. After 5 performances, we have seen all singers on stage and have, even in the worst case, identified a group of 4 singers that must contain the 2 stars. Using 2 more performances, it is possible to identify the 2 stars and their relation. This strategy thus works in at most $5+2$ attempts. We will see that this strategy is not optimal either.

We encourage the reader to pause here. Using the argument above, 7 is an upper bound on the number of performances required in the worst case. Is there a way to develop a lower bound?

## 5.2 HOW WELL CAN A STRATEGY WORK?

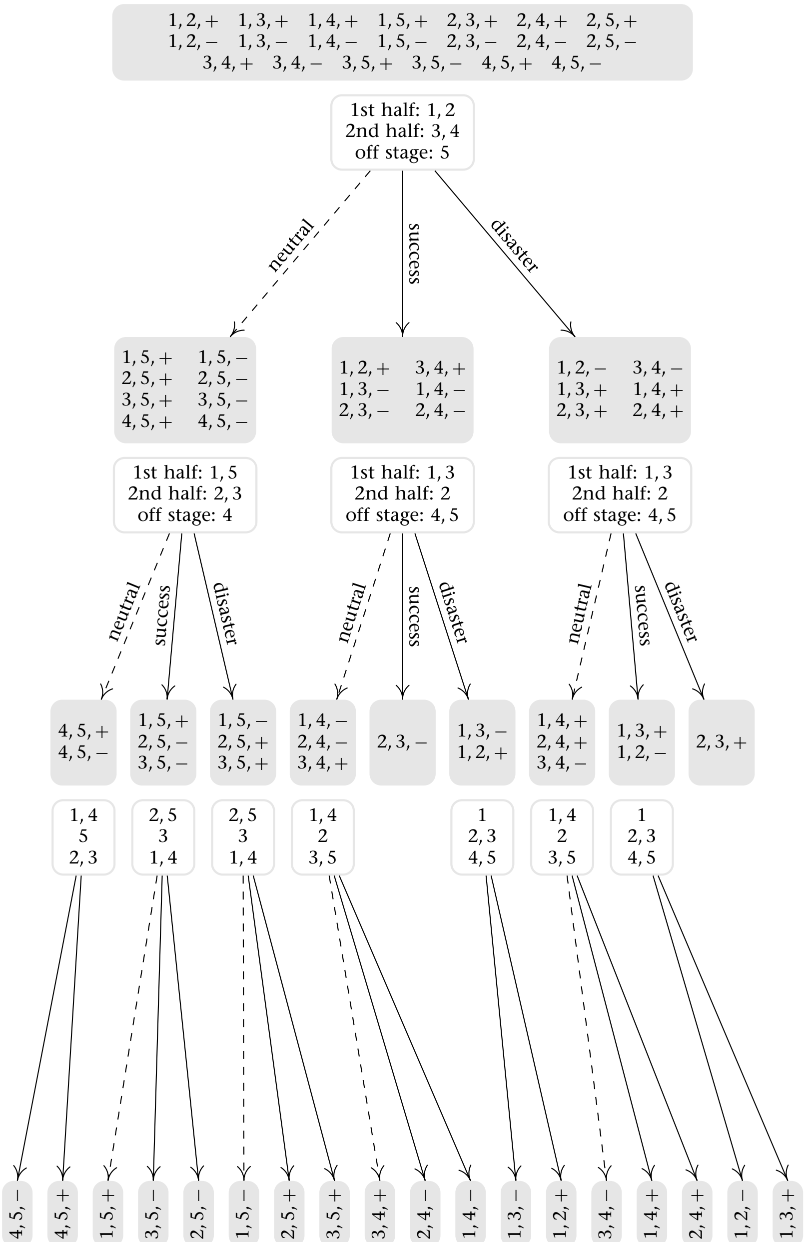
We first compute a lower bound on what we can hope for: What is the minimum number of performances required for us to be certain about who the stars are and what their relationship is? Let us assume that there are 5 singers. Then we can count the number of possible solutions. There are $5 \cdot 4/2 = 10$ different pairs that could be the pair of stars. Each pair of stars can either consist of 2 friends or 2 enemies, so there are 20 possible solutions in total.

Now, each opera performance can be seen as an experiment with 3 possible outcomes (neutral, disaster, or success). Suppose there is a deterministic strategy that will always enable us to know the true solution after 2 performances. Then, each of the 20 possible solutions will yield exactly 1 pair of outcomes, such as (success, neutral), (success, disaster), or (neutral, neutral). There are, however, only $3 \cdot 3 = 9$ possible outcomes with 2 performances. Several solutions must yield the same pair of outcomes, and we have no chance of distinguishing these solutions. As a consequence, there is no strategy that always succeeds with just 2 performances.[1] What about 3 performances? We then have $3 \cdot 3 \cdot 3 = 27$ possible outcomes, which could be enough to discover the truth among the 20 possibilities. We will see that there is indeed a strategy that will always find the truth after 3 performances. Section 5.3 presents the solution in the case of 5 singers and 3 performances. In section 5.4, we introduce some ideas from information theory that will help us understand why these strategies work and how to extend them to more general settings.

## 5.3 SOLUTION FOR 5 SINGERS

First, let us write down all possible pairs of opera singers together with an indicator (+ or −) as to whether they are friends (+) or enemies (−). We stated already that there are 20 hypotheses in total: $\Omega := \{(1,2,+),\ (1,2,-),\ (1,3,+),\ (1,3,-),\ (1,4,+),\ (1,4,-),\ (1,5,+),\ (1,5,-),\ (2,3,+),\ (2,3,-),\ (2,4,+),\ (2,4,-),\ (2,5,+),\ (2,5,-),\ (3,4,+),\ (3,4,-),\ (3,5,+),\ (3,5,-),\ (4,5,+),\ (4,5,-)\}$. A strategy that lets us identify the true solution in any of these cases is summarized in the following figure.

1. The same argument holds for randomized strategies. Each solution then yields a probability distribution over the 9 pairs of outcomes. Therefore, there will be at least 1 pair of outcomes that has a positive probability of occurring for at least 2 different solutions. For simplicity, however, we are going to restrict ourselves to deterministic strategies in the remainder of this chapter.


1, 2, + 1, 3, + 1, 4, + 1, 5, + 2, 3, + 2, 4, + 2, 5, +
1, 2, − 1, 3, − 1, 4, − 1, 5, − 2, 3, − 2, 4, − 2, 5, −
3, 4, + 3, 4, − 3, 5, + 3, 5, − 4, 5, + 4, 5, −
1st half: 1, 2
2nd half: 3, 4
off stage: 5
neutral
success
disaster
1, 5, + 1, 5, −
2, 5, + 2, 5, −
3, 5, + 3, 5, −
4, 5, + 4, 5, −
1, 2, + 3, 4, +
1, 3, − 1, 4, −
2, 3, − 2, 4, −
1, 2, − 3, 4, −
1, 3, + 1, 4, +
2, 3, + 2, 4, +
1st half: 1, 5
2nd half: 2, 3
off stage: 4
1st half: 1, 3
2nd half: 2
off stage: 4, 5
1st half: 1, 3
2nd half: 2
off stage: 4, 5
neutral
success
disaster
neutral
success
disaster
neutral
success
disaster
4, 5, +
4, 5, −
1, 5, +
2, 5, −
3, 5, −
1, 5, −
2, 5, +
3, 5, +
1, 4, −
2, 4, −
3, 4, +
2, 3, −
1, 3, −
1, 2, +
1, 4, +
2, 4, +
3, 4, −
1, 3, +
1, 2, −
2, 3, +
1, 4
5
2, 3
2, 5
3
1, 4
2, 5
3
1, 4
1, 4
2
3, 5
1
2, 3
4, 5
1, 4
2
3, 5
1
2, 3
4, 5
4, 5, −
4, 5, +
1, 5, +
3, 5, −
2, 5, −
1, 5, −
2, 5, +
3, 5, +
3, 4, +
2, 4, −
1, 4, −
1, 3, −
1, 2, +
3, 4, −
1, 4, +
2, 4, +
1, 2, −
1, 3, +

We start by putting singers 1 and 2 in the first half, and singers 3 and 4 in the second half (see the white box near the top of the figure). In the scenario where the opera is declared a success, for example, we are left with 6 possible hypotheses (see the gray box connected by “success”), and we then put singers 1 and 3 in the first half and singer 2 in the second half (see the white box below). No matter what the underlying truth is, after no more than 3 performances, there is only one hypothesis left, and so the group will know exactly who the stars are and whether they are friends or enemies.

## 5.4 SOME MATHEMATICS: INFORMATION THEORY

The key to solving this game is to ask the right questions. Similarly, in the game ‘Who am I?’, players receive hidden identities (e.g., based on famous people) and try to reveal their identities by asking questions that will provide as much information as possible. But how do we measure informativeness? This is the starting point of *information theory*.

### Measuring Information Content

Suppose you are playing a number-guessing game with your friend. Your friend is thinking of a number between 0 and 15, and your task is to find which number your friend is thinking of by asking yes/no questions. The goal is to ask as few questions as possible. Let us assume that your friend chooses her number at random and that she does not prefer any of the numbers: None of the 16 numbers is chosen with larger probability than the others.

To help us understand information content, it will be convenient to write your friend’s number as a *binary word*. If she thinks about the number 5, we write this as 0101. Why 0101? Because 5 can be written as

$$5 = \mathbf{0} \cdot 2^3 + \mathbf{1} \cdot 2^2 + \mathbf{0} \cdot 2^1 + \mathbf{1} \cdot 2^0.$$

Similarly, 14 is written as 1110, since

$$14 = \mathbf{1} \cdot 2^3 + \mathbf{1} \cdot 2^2 + \mathbf{1} \cdot 2^1 + \mathbf{0} \cdot 2^0.$$

The four digits in 0101 and 1110 are often called “bits,” which is short for “binary digits.” (You can find more details about binary expansions in “What Is … a Binary Number?” in appendix B.1.)

If you ask the question “Are you thinking of the number 7, that is, 0111?” we can quantify the information that is contained in your friend’s answer. If she answers “Yes, it is,” you received 4 bits of information: You know all four bits (0111) of your friend’s answer. However, if your friend answers “No, it is not,” there are 15 out of the 16 possibilities left, which means that you received very little information. How little, exactly?

In general, we define the *information content* of an event as

$$\log_2 \frac{1}{p},$$

where $p$ is the probability of that event. Here, $\log_2$ is the logarithm to base 2 (appendix B.3 contains more details on the logarithm function). In the example above, this yields

$$\log_2 16 = 4 \quad \text{for the “yes” answer,}$$
$$\log_2 \frac{16}{15} \approx 0.093 \quad \text{for the “no” answer.}$$

The unit of information content is often also called a “bit.” In our number guessing game, that name fits well.

As we may have expected, it is not a good idea to ask questions whose answers we already know: If the question is such that the answer has probability 1, the information content will be $\log_2(1) = 0$ bits. It is not helpful to ask questions such as “Is your number smaller than 1401?”, for example, because the answer does not contain any helpful information—the information content equals zero bits.

In the opera game, there are not only 2 (yes/no) but also 3 possible answers (neutral/success/disaster). And if we consider two performances in a row, there are even $3 \times 3 = 9$ possible outcomes. But the key idea of information content remains the same. If there are $m$ possible answers (with probabilities $p_1, p_2, \ldots, p_m$, say), then the information content of the $m$ answers are, respectively,

$$\log_2 \frac{1}{p_1}, \log_2 \frac{1}{p_2}, \ldots, \log_2 \frac{1}{p_m}.$$

**Information Content and Selecting Hypotheses**

Information content is closely related to the ability to identify the true hypothesis among a set of possible hypotheses. This set is often written as $\Omega$. Let us start again with the number guessing game described near the start of this section. There, $\Omega = \{0, 1, 2, \ldots, 15\}$.

We now describe any strategy as a map to help us navigate from the set of possible hypotheses to a set of possible answers. Let $\omega^* \in \Omega$ denote the secret number that your friend thinks about but that is unknown to us. Asking a yes/no question now corresponds to constructing a function ("$O$" for outcome),

$$O : \Omega \to \{0, 1\} \qquad \text{(one no/yes question)}$$

and asking your friend for the value of $O(\omega^*)$; that is, $O(\omega^*) = ?$. Here, we simply identify "yes" with 1 and "no" with 0. The question, "Are you thinking of the number 7?", for example, translates into a function $O$ that maps 7 to 1 and all other values to 0. So, $O(\omega^*) = 1$ if $\omega^*$ happens to be 7, and $O(\omega^*) = 0$ otherwise. In this case, we have $P(O = 1) = 1/16$, because only one element in $\Omega$ maps to 1.

Asking two yes/no questions in a row corresponds to constructing a function $O : \Omega \to \{0, 1\}^2$. The set $\{0, 1\}^2$ of answers contains all possible combinations of yes and no answers: $(0, 0), (0, 1), (1, 0)$, and $(1, 1)$. Or, in words, (no, no), (no, yes),

(yes, no), and (yes, yes). Finally, asking $k$ questions corresponds to constructing the function

$$O:\Omega\to\{0,1\}^k \qquad \text{(asking } k \text{ no/yes questions)}$$

and, again, asking your friend for the value of $O(\omega^*)$.

We thus think about the truth (that is, the secret number) as an element $\omega^*\in\Omega$ that is unknown to us. Asking questions or combinations of questions corresponds to constructing functions $O:\Omega\to\{0,1\}^k$. The answer, given by our friend, equals $O(\omega^*)$, the evaluation of the truth $\omega^*$. This turns out to be a convenient description of the problem. The set $\Omega$ tells us the number of hypotheses, and the set $\{0,1\}^k$ tells us the number of possible answers. Now, for $k=3$, for example, the set $\{0,1\}^k$ contains 8 elements. But since $\Omega$ contains 16 elements, $\#\Omega=16$, there are at least 2 elements in $\Omega$ that map to the same answer. That is, no matter what strategy we choose, we can never be absolutely certain that we will identify the correct number after $k=3$ questions. If, however, we are lucky, and the information content of a given answer $O=i$, say, is $\geq 4$ bits, then we have identified the secret number $\omega^*$:

$$\log_2\left(\frac{1}{P(O=i)}\right)\geq 4 \quad\Rightarrow\quad \frac{1}{P(O=i)}\geq 16 \quad\Rightarrow\quad P(O=i)\leq\frac{1}{16}.$$

Thus at most one element $\omega^*$ maps to the answer $x$ (otherwise, the probability of obtaining that answer would be at least 2/16)—so this must be the true secret number $\omega^*$! These two observations hold true in general, and we reiterate them here:

$$\#\Omega > \#\text{possible answers} \quad \overset{\text{Rule 1}}{\Longrightarrow} \quad \omega^* \text{ cannot always be identified,}$$

$$2^{\text{information content}}\geq\#\Omega \quad \overset{\text{Rule 2}}{\Longrightarrow} \quad \omega^* \text{ is identified.}$$

Rule 1 gives us a lower bound on the number of questions that we need to ask to be able to identify the truth even in the worst case. Rule 2 tells us that an answer with high information content allows us to identify the truth.

What about the opera game described in section 5.1? There, the set $\Omega$ contains 20 elements and is listed at the start of section 5.3. We will again describe strategies by using maps. But unlike for the number guessing game, we have 3 possible outcomes, so this time, any strategy can be described by

$O: \Omega \to \{0, 1, 2\}^k$ ($k$ performances with neutral/success/disaster).

We can now exploit the two rules that we learned about above. On one hand, $k = 2$ performances do not suffice to always identify the true solution, because $\#\Omega = 20 > 9 = \#$possible aswers. On the other hand, receiving more than $\log_2(20) \approx 4.322$ bits of information content is in principle enough to identify the two stars and whether they are friends or enemies. This is because $2^{\log_2(20)} = 20 \geq \#\Omega$.

Because of the rules of the game (and unlike in the number-guessing game), we cannot implement any arbitrary function $O$ as a strategy. For example, we cannot simply ask "Is this singer a star?" But the other direction works fine: Any strategy can be described as a function $O$.

In summary, receiving a lot of information content is good for identifying the correct hypothesis. In practice, we can control the information content by designing questions or experiments with certain outcome probabilities. We will now be looking at how to design the best questions (i.e., the ones that will prompt answers with a high information content). These answers will then help us identify the correct hypothesis.

### Principle 1: Maximizing the Minimal Information Content

Let us assume that in the first opera performance, we put 3 of the 5 singers in the first ensemble and the remaining 2 in the second ensemble—all players are on stage. The probabilities of success and disaster will then both be 0.5 (each pair can yield either outcome, depending on whether they are friends or enemies). The neutral option will never come up. This means, we obtain the following bits of information:

| Outcome | neutral | success | disaster |
|---|---|---|---|
| Information content | ✗ | $\log_2 2 = 1$ | $\log_2 2 = 1$ |

Is this a good question to start with? It turns out that we can do much better in terms of information content. If instead, we leave 1 singer off-stage, there will be a neutral reaction in 8 out of 20 cases (namely, in the cases where the left-out singer is one of the stars). We would therefore obtain:

| Outcome | neutral | success | disaster |
|---|---|---|---|
| Information content | $\log_2(20/8) \approx 1.32$ | $\log_2(20/6) \approx 1.74$ | $\approx 1.74$ |

bits for the neutral, success, and disaster outcomes, respectively. Every outcome of the second performance will therefore contain more information than any outcome of the first performance.

The theory of information provides us with a powerful tool for designing questions. If we want to minimize the maximal number of questions required to find out the truth, we need to ensure that even in the worst case, the answer's information content is as large as possible. Let us assume that there are $m$ possible outcomes, and the probability $P(O = i)$ of receiving outcome $i$ equals $p_i$. The outcome with the smallest information content is the one corresponding to

$$\min_{i \in \{1,\ldots,m\}} \log_2 \frac{1}{p_i},$$

that is, the one corresponding to the largest probability $p_i$. This yields the following rule:

> *We maximize the minimal amount of information by making the most probable answer as unlikely as possible (that is, by making the largest* $p_i$ *as small as possible).*

**Principle 2: Maximizing the Average Information Content**

Another idea is to find a question whose answer contains a lot of information *on average*. Consider a question that is answered "yes" with probability $p$ and therefore "no" with probability $1-p$. The expected information is then defined as

$$H(p) = p\log_2(p) + (1-p)\log_2(1-p),$$

and is usually referred to as the *entropy*. We adapt the widely used convention of defining $0 \cdot \log_2 0 := 0$.[2]

In the example of the number-guessing game, the entropy of our question "Are you thinking of the number 7, that is, 0111?" equals

$$\frac{1}{16}\log_2 16 + \frac{15}{16}\log_2 \frac{16}{15} \approx 0.337,$$

that is, we receive 0.337 bits of information on average. Most people would not start with the above question, but would rather ask: "Is your number greater than or equal to 8?" for example. The entropy clarifies in which sense this question is more informative. On average,[3] the latter question yields

$$\frac{1}{2}\log_2 2 + \frac{1}{2}\log_2 2 = 1$$

bit of information. This makes sense intuitively: The answer to "Is your number greater than or equal to 8?" tells us the first bit in the binary expansion of the secret number.

In the general case of $m$ possible answers, we define the *entropy* as the expected information content:

$$H := H(O) := H(p_1, \ldots, p_m) := p_1 \log_2 \frac{1}{p_1} + \cdots + p_m \log_2 \frac{1}{p_m}.$$

2. This convention makes sense, as $p \cdot \log_2 p$ is converging to 0 if $p > 0$ is converging to 0.

3. In fact, the minimal information content is also 1 bit in this case.

Appendix B.6, “What Is … an Expectation?” contains a few more details on why this is the *expected* information content. In many games, we are not interested in worst-case behavior, but we would like to receive a lot of information on average. To do so, we should design the question so that its entropy is as large as possible. But how do we achieve that? In short, the entropy is maximized if all $m$ answers are equally likely, that is, $p_1 = p_2 = \cdots = p_m = 1/m$. A proof of that statement can be found in appendix C.3. This observation yields the following rule:

> *The average amount of information is maximized if all* $p_i$*s take the same value.*

What does that mean, intuitively? If all answers to a certain question are equally likely, the average information content of the answer is as large as possible. The other extreme is if one of the $m$ answers has probability 1 and all others have probability 0. Then, the average information content is 0.

If it is possible to find a question such that all answers have the same probability, both the average information content and the minimal information content are maximized and equal. This is what we should aim for! Often, however, finding questions with equal answer probabilities is impossible, and the two concepts differ from each other.

### Finding Optimal Strategies

With these principles in mind, we can now try to construct useful strategies. In the opera game, there are three outcomes, so we should design a sequence of performances, represented by the map

$$O : \Omega \to \{0, 1, 2\}^k$$

such that all the outcomes have (roughly) equal probability. This, however is a difficult task. In practice, we can often design the performances one-by-one. Starting with $O_1 : \Omega \to \{0, 1, 2\}$, we try to design the performance in such a way that each outcome has probability 1/3. In our solution on page 78, the probabilities equal 6/20, 6/20, and 8/20 and are as close as possible to 1/3, 1/3, and 1/3. But what happens to the information if we add a second performance? The amount of information content that the second answer adds to the first answer depends on how related the two questions are. Assume, for example, that in the number-guessing game, we ask two yes/no questions: “Is your number even?” and “Does the last digit in the binary expansion equal one?”. Both of these questions individually receive on average 1 bit of information ($1 = 1/2 \log_2 2 + 1/2 \log_2 2$). But both of the answers together are still worth 1 bit of information. The reason is that the two answers depend on each other: If the first question is answered with “yes,” then the second answer must be “no” and vice versa. If, however, the questions are constructed in such a way that the answers are independent (e.g., “Does the first digit equal 1?” and “Does the second digit equal 1?”), the average amount of information of the combined questions is maximized and equals the sum of both entropies: 2 bits. This is not too hard to prove—see appendix B.5 for details on independence.

No matter what the outcome of the first performance, we thus need to construct a second performance whose outcome is as independent as possible from that of the first performance. In practice, this yields the following rule:

*At each point in time, ask a question the answers to which, given current knowledge, have roughly equal probability.*

In fact, the solution shown on page 78 was constructed by such a step-wise procedure. At each point in time, we constructed the performances in such a way that any outcome had roughly equal probability.

This procedure is simple and can be applied to many games, where one needs to identify a hypothesis (the section 5.5 discusses another game of this type). However, there is a drawback to the stepwise approach: It does not always yield an optimal strategy (we provide an example in appendix C.3). In our case, however, we are lucky. We know from before (page 83) that there is no strategy based on only two performances or fewer that can always identify the true hypothesis. Since our strategy involves three performances, it must be optimal.

The idea of maximizing information content not only helps with mathematical problems. We usually consider the conversations with high information content as particularly interesting. Questions whose answers we can guess with high probability are (on average) boring. But in a few cases, they yield high information content. For example, asking a stranger about a friend you may have in common usually results in a negative answer; but if it does not, the surprise is large—and so is the information content.

## 5.5 VARIATION: BALL WEIGHING

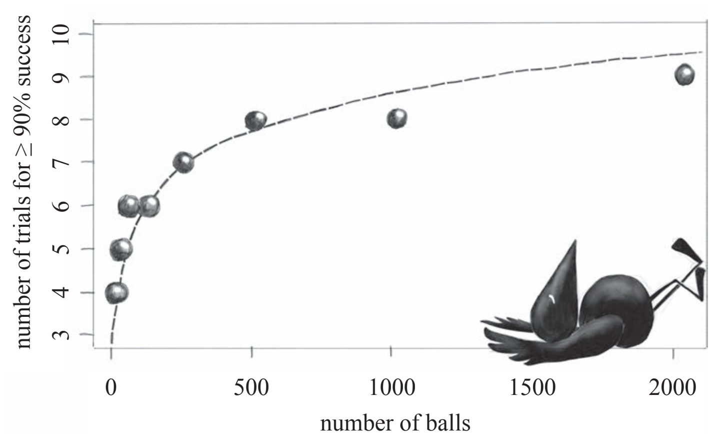
The opera singers problem is a variation on a famous ball-weighing problem. That problem is formulated as follows. There are 12 balls, 11 of which are of equal weight. The twelfth ball, however, is different; it weighs a different amount, but that difference is unknown: It is either lighter or heavier than the other balls. Since the odd ball out is the same size and color as the other balls, it is visually indistinguishable from the others. The task is now to use a balance scale to identify the "oddball" with as few weighings as possible (see the figure on the next page). Here, the balance only indicates whether both sides are equally heavy and, if they are not, which side is heavier.

The ideas we discussed in this chapter can also be used to solve the weighing problem. Which weighing shall we start with? Do we expect to gain more information from weighing 6 against 6 balls, 4 against 4, or 1 against 1? Section 5.7 tells you where to find the solution.

## 5.6 RANDOM STRATEGIES

For most problems, designing strategies that optimize information content is difficult. We have seen that the stepwise approach can be helpful in practice but does not necessarily provide the best strategy (see appendix C.3). However, there is a simple and surprisingly useful alternative: We can simply design the experiments randomly. In the ball-weighing problem, for example, we can put 4 randomly chosen balls on the left-hand side of the balance and 4 other randomly chosen balls on the right-hand side. After 3 such random weighings, we can then check how many hypotheses could be used to explain the results. We repeated this experiment 10, 000 times (fortunately, we were able to use a computer) and found that in about 47% of the cases, only a single hypothesis was left. If more than one hypothesis is left (e.g., ball 3 could be lighter or ball 9 heavier than the other balls), we can also randomly choose one such hypothesis (e.g., we guess "Ball 9 is heavier than the other balls"). This allows us to correctly identify the special ball in about 72% of cases. For 4 weighings, these numbers increase even further to 81% and 90%, respectively.

This random strategy is extremely simple and fast. Each weighing is randomly constructed without the need to adapt to the result of the previous weighing. When considering a large number of balls, for example, finding the optimal strategy is very difficult—even the stepwise procedure is a bit cumbersome to compute. The random strategy, however, remains simple. Motivated by our deliberations on information content and entropy, we always choose $\approx p/3$ randomly selected balls on each side of the balance, where $p$ is the number of balls. The following graph shows how many weighings are needed to obtain a success probability of greater than or equal to 90% (again allowing for random guesses if there are ties). For $p = 2{,}048$ balls, we only need 9 weighings!




It seems surprising that the number of necessary weighings grows so slowly with the number of balls. But there is mathematical theory that explains this behavior. It tells us that for a growing number $p$ of balls, we only require (roughly) $C \cdot \log(p)$ weighings. The figure on the previous page shows a graph of that function for a specific choice of $C$: It fits the black dots quite nicely. The underlying theory is sometimes called "compressed sensing" or "compressive sampling." To apply it to our problem, we represent the truth by a long vector $\beta$ of length $p$ that contains only 0s, apart from at the position of the secret ball, where the vector has either a $+1$ or a $-1$, indicating whether the ball is heavier or lighter than the others. A weighing can now be described by a vector $x$ of length $p$, for which the entries $-1$, 1, and 0 indicate, respectively, that the corresponding ball is on the left-hand side of the balance, on its right-hand side, or is not included. The result of a weighing then corresponds to the measurement $\sum_{i=1}^{p} x_i \beta_i$. The values $-1$, 1, and 0 indicate that the left-hand side is heavier, the right-hand side is heavier, or that they are equally heavy, respectively. The theory of compressed sensing now tells us that the number of measurements we need to reconstruct the vector $\beta$, grows only like the logarithm of $p$. One of the reasons the random strategy works so well is that out of all these balls, only one ball differs from the others. Random strategies still work if there are only a few balls that are different from the majority, but if there were lots of balls like this, we would have no chance of recovering them with only a few random weighings. The theory of compressed sensing can not only be used for our weighing problem. It can also be used in many other areas, such as seismology, photography, and medical imaging.

## 5.7 SHORT HISTORY

To the best of our knowledge, the earliest formulation of the weighing problem appeared in a 1945 article in the *American Mathematical Monthly* [Schell, 1945], with a more general game published the following year [Eves, 1946]. You can read more about the ball-weighing problem and its solution (there is a strategy that uses at most 3 weighings) in the book *Information Theory, Inference, & Learning Algorithms* [MacKay, 2002] on page 68, for example. Our figure on page 78 is inspired by David MacKay's figure on page 69 of his book. The concept of entropy [e.g., Cover and Thomas, 2006] and its connection to information content has been studied by Claude Shannon, who was one of the pioneers in that area [Shannon, 1948]. His original motivation was to study communication, and today the theory is widely applied not only in communication and signal processing but also in data compression and data science, as well as in other areas. The theory of compressed sensing [Candès et al., 2006; Donoho, 2006] was developed by Emmanuel Candès, David Donoho, Justin Romberg, and Terence Tao.

## 5.8 PRACTICAL ADVICE

In a first instance, the game can be played competitively. The host divides the players into three groups in each round of the game. She can do the division in any way she likes and declares the performance again either a success, disaster, or neutral, according to the same rules. The players have to guess who the stars are and what their relationship is. The first player to get it right wins the game (one can allow for one false answer per person). Once the players have played this competitive round, they can move to the cooperative version of the game, where they have to agree on the division of the group and guess the identity of the stars and their relationship together. The players should aim to be faster (i.e., to use fewer rounds) than in the competitive round of the game.