Chapter 3
The Helmholtz Principle

The Helmholtz principle can be formulated two ways. The first way is commonsensical. It simply states that we do not perceive any structure in a uniform random image. In this form, the principle was first stated by Attneave [Att54]. This gestaltist was to the best of our knowledge the first scientist to publish a random noise digital image. This image was actually drawn by hand by U.S. Army privates using a random number table. In its stronger form, of which we will make great use, the Helmholtz principle states that whenever some large deviation from randomness occurs, a structure is perceived. As a commonsense statement, it states that “we immediately perceive whatever could not happen by chance”. Our aim in this chapter is to discuss several intuitive and sometimes classical examples of exceptional events and their perception. We will see how hard it can be to calculate some rather simple events. This difficulty is solved by introducing a universal variable adaptable to many detection problems, the Number of False Alarms (NFA). The NFA of an event is the *expectation of the number of occurrences* of this event. Expectations are much easier to compute than probabilities because they add. After we have treated three toy examples in Section 3.1, we will define in Section 3.2 what we call ε*-meaningful events*, namely events whose NFA is less than ε. This notion is then applied to a first realistic problem: the dot alignment detection in an image.

3.1 Introducing the Helmholtz Principle: Three Elementary Examples

3.1.1 A Black Square on a White Background

Assume two scholars are looking at a picture of, say, 100×100 size, namely 10,000 pixels. Assume the figure contains somewhere a 10×10 black square; all other pixels are white (see Figure 3.1). Common sense tells us that such a figure could not arise just by chance: We are “sure” that this organization corresponds to an intention; somebody drew a square there and this is why we see it. Now, the obvious intuition

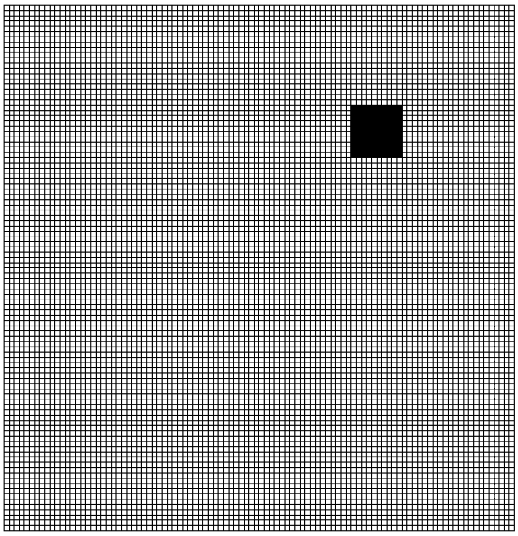

Fig. 3.1 A grid of size 100×100 containing a 10×10 square of blackened pixels; all other pixels are white. The probability of this particular configuration is $2^{-10,000}$, assuming that the pixels are independent and black or white with probability $1/2$.

that, for example, white noise cannot generate the black square must be quantified. It is actually possible that a white noise generates a black square, particularly if we are allowed to repeat the same experiment many times. As the following dialogue will show, there are many difficulties to overcome before we can "prove" the existence of the square as a meaningful event. Our dialogue takes place between a sceptic and an enthusiast. The enthusiast is sure that he sees a square and that its existence can be proven by probabilistic arguments. The sceptic will try to find, and will succeed in finding, many objections.

— Sceptic: "You think you see a square; but, all I see is a set of white or black pixels. They just fell together by chance and built this square just by chance."

— Enthusiast: "Anybody looking at a picture and knowing what a square is will claim "This is a square". Now, why are they so sure? Since you talk about chances, let us interpret their decision in a probabilistic way: Assume indeed that the pixels are white or black just by chance. Assume black and white have a probability of $1/2$. Then the probability of the black square appearing is just $(1/2)^{10,000}$ that is, about 10^{-3000}. Thus, the event is very, very, very unlikely."

— Sceptic: "Your calculation is wrong: I never said that the probabilities for white and black are equal. Any Bernoulli distribution is possible."

— Enthusiast: "Well, I am pretty sure you remember enough of your Feller reading to acknowledge that when 10,000 samples have been observed, 100 of which were black and 9900 white, then the probability of a pixel being black is likely to be close to $1/100$, is not true? So, the probability of the square happening just by chance is just about $(1/100)^{100}(99/100)^{9900}$ and it's still very small."

— Sceptic: "I concede that whatever probability p for the black you assume, the probability of the observed square is $p^{100}(1-p)^{9900}$ and therefore very small. But you've made a beginner's mistake. You know very well that probabilities must be calculated *a priori*. So you are not allowed to compute *a posteriori* the probability that particular square happened. You indeed ignored *a priori* where and what it was."

— Enthusiast: "I knew you would raise this objection! But it's easily fixed. I won't compute the probability of *that particular square happening*, but the probability of *any* square happening! Let us call k the square's side. Let's call (x,y) the position of its top-left corner: There are $(100-k+1)^2$ possible positions for this corner. Since all square events are disjoint, I can compute the probability that "some square appears somewhere in the image" and this number will be very small anyway. You will agree that the events "a square of side-length k appears at position (x,y)" are all disjoint when k varies. So I can just sum up their probabilities and I get

$$\mathbb{P}(\text{any square happening}) = \sum_{k=1}^{k=100} (100-k+1)^2 p^{k^2} (1-p)^{10{,}000-k^2}.$$

With $p = 1/100$, you'll agree it's a very small number anyway."

— Sceptic: "Ha! The more complicated you make it, the more objections you'll have. First of all, p depends on k. Are you forgetting that p is just an *a posteriori* estimate of the chances of a pixel being black, drawn from actual observation?"

— Enthusiast: "Well, all right then. Let's take the "unbiased estimate" of p. It's $p_k = k^2/10{,}000$. We can just replace p by p_k in my formula. And the sum is still very, very small."

— Sceptic: "No, it's not! Let's see... the largest term must be the first one. It's $10^4 \times 10^{-4}(1-10^{-4})^{9999} \approx e^{-1} \geq 1/3$. Do you call $1/3$ very small? Forget about your complicated formulas and use your own common sense if you have any left: If you are observing a single black pixel, are you allowed to call it a square? A dot has no shape."

— Enthusiast: "I went too far, I confess. But there must be a minimal size above which we are sure we *see* a square. So, I propose finding the *minimal size k above which we are sure we see a square*. My feeling is that if k exceeds, say 10, we are already sure we see a square. This is common sense as you call it! So I claim that the following number is very small for $k_{min} \geq 10$:

$$\mathbb{P}(\text{any square happening of size larger than } k_{min}) = \sum_{k=k_{min}}^{k=100} (100-k+1)^2 p_{k_{min}}^{k^2} (1-p_{k_{min}})^{10{,}000-k^2}."$$

— Sceptic: "I'm sure you agree this computation isn't quite convincing: to start with, p is assumed to depend upon k, as $p(k) = k^2/10{,}000$. So you cannot fix a single k_{min}. Clearly, k_{min} depends on k. Now I'll be fair: There might be something out there. This probability you propose is indeed very small. What do you say to my next objection: The square is nothing special in your computations. You could do the same computations about *any* configuration of black pixels. So whatever random image might be presented to me, I can, by your very same computations, claim that its probability was *a priori* very low and deduce that I see something exceptional. Let me be more specific: Any realization of white noise has an equally low probability! In the case $p = 1/2$, all configurations have probability $2^{-10{,}000}$. So any one of them is "exceptional" along your line of thought since it has this low probability. All the same, one of them will occur. Since there are $2^{10{,}000}$ possible configurations, the sum of their probabilities is 1. So the event of "one of these exceptional configurations occurring" has probability $2^{10{,}000} \times 2^{-10{,}000} = 1$. No surprise there!"

— Enthusiast: "Hm, You know what? This objection poses a real problem and all the other ones were mere child's play. So I feel forced to enlarge and simplify my model. You'll agree that we do not usually recognize shapes in a white noise image. What I claim is this: The number of shapes known to humans is limited. Let us say there are as many objects as words in a good dictionary, namely 10^5. Let's assume 10^{10} aspects of the same object due to different pauses and ways it was built, angles of view, and so forth. Let's allow also for 10^3 different ways light can be shed on the same object. This means that the number of all possible black and white silhouettes of all world objects is about 10^{18}. All the same, this number is very small with respect to $2^{10{,}000}$. So, if we see the silhouette of a known object inside our 100×100 image, we'll immediately recognize it. The probability of each one of the familiar silhouettes occurring is $2^{-10{,}000}$, so the probability of any one of the silhouettes occurring is less than $2^{-10{,}000}.10^{18}$ which is again a very small number. So, you see, I stand my ground, since a 10×10 black square simply is one of those familiar silhouettes."

— Sceptic: "As the French say, *vous vous échappez par les branches*. We were talking about a square, and all of a sudden you start talking about *all* shapes in the world and making fantastic estimates about their number. I really don't think we're on the same page!"

3.1.2 Birthdays in a Class and the Role of Expectation

Black squares on a white background are a tough and abstract subject. So let us return to a more familiar problem: the classical problem of shared birthdays in a class. Is it surprising that two alumni have the same birthday in a class of 30? And if not, would it be surprising to observe three alumni having the same birthday? Even such a simple situation can be formalized in different ways, depending on the various answers we may put forth.

We have looked at a class of 30 students. Let us assume that their birthdays are independent and uniformly distributed variables over the 365 days of the year. We call, for $1 \le n \le 30$, C_n the number of n-tuples of students in the class having the same birthday (this number is computed exhaustively by considering all possible n-tuples. If, for example, students 1, 2, and 3 have the same birthday, then we count three pairs $(1,2)$, $(2,3)$, and $(3,1)$). We also consider $\mathbb{P}_n$, the probability that there is at least one n-tuple with the same birthday and p_n, the probability that there is at least one n-tuple and no $(n+1)$-tuple. In other terms, $\mathbb{P}_n$ is the probability of the event "$C_n \ge 1$", that is, "there is at least one group of n alumni having the same birthday" and p_n the probability of the event "the largest group of alumni having the same birthday has cardinality n". We are primarily interested in the evaluation of $\mathbb{P}_n$ and of the expectation $\mathbb{E}C_n$ as good indicators for the exceptionality of the event.

Proposition 1 *The probability that no two alumni have the same birthday in a class of 30 is* $(365 \times 364 \times \ldots \times 336)/(365^{30}) \approx 0.294$. *The probability that at least two alumni were born on the same day therefore is*

$$\mathbb{P}_2 \approx 0.706.$$

Proof — Number the alumni from 1 to 30. Given any date among the 365 possible, the probability that alumnus 1 has this birthday is $1/365$. So the probability that alumnus 2 has the same birthday as alumnus 1 is $1/365$. The probability that their birthdays differ therefore is $1 - 1/365 = 364/365$. In the same way the probability that alumnus 3 has a birthday different from alumni 1 and 2 is $1-2/365 = 363/365$. Since the birthdays are supposed independent (no twins in the class), we arrive at the expected result. □

At this point, we notice that without a computer, we would have been in some pain to compute a good approximation of this probability. There is, however, another way to demonstrate the likeliness of two alumni having the same birthday. As usual, when a probability is difficult to compute, we may compute an expectation. By the Markov inequality, expectations give hints on probabilities.

Proposition 2 *The expectation of the number of pairs of alumni having the same birthday in a class of 30 is* $\mathbb{E}C_2 = \frac{30\times 29}{2\times 365} \approx 1.192$. *The expectation of the number of n-tuples is* $\mathbb{E}C_n = \frac{1}{365^{n-1}}\binom{30}{n}$. *By an easy calculation,* $\mathbb{E}C_3 \approx 0.03047$ *and* $\mathbb{E}C_4 \approx 5.6 \times 10^{-4}$.

Proof — Enumerate the students from $i = 1$ to 30 and call E_{ij} the event "students i and j have the same birthday". Also, call $\chi_{ij} = \mathbb{1}_{E_{ij}}$. Clearly, $\mathbb{P}(E_{ij}) = \mathbb{E}\chi_{ij} = 1/365$. Thus, the expectation of the number of pairs of students having the same birthday is

$$\mathbb{E}C_2 = \mathbb{E}\left(\sum_{1\le i<j\le 30} \chi_{ij}\right) = \sum_{1\le i<j\le 30} \mathbb{E}\chi_{ij} = \frac{30\times 29}{2}\frac{1}{365} \approx 1.192.$$

The general formula follows by analogous reasoning. □

"On the average", we can expect to see 1.192 pairs of alumni with the same birthday in each class. Unfortunately, this information is a bit inaccurate, since the large number of pairs on the average could be due to exceptional cases where one observes a lot of pairs. We only know by Markov inequality that $\mathbb{P}_2 \le \mathbb{E}C_2$. The situation would be quite different if $\mathbb{E}C_n$ were small. In that case, an estimate on $\mathbb{E}C_n$ will give us a "cheap" estimate on $\mathbb{P}_n$. This is what we get from the estimates of $\mathbb{E}C_3$ and $\mathbb{E}C_4$. Both tell us immediately that triplets or quadruplets are not likely. They also yield an estimate of $\mathbb{P}_3$ and $\mathbb{P}_4$ from above. How good that estimate is can be derived from the following results (see Exercise 3.4.1 at the end of the chapter).

$$p_2 = \frac{1}{365^{30}} \sum_{i=1}^{15} \left[\frac{\prod_{j=0}^{i-1} \binom{30-2j}{2}}{i!} \prod_{k=0}^{29-i} (365-k) \right] \approx 0.678$$

and, after a brave computation, $\mathbb{P}_3 \approx 0.0285$. In the same way,

$$p_3 = \frac{1}{365^{30}} \sum_{i=1}^{10} \frac{\prod_{j=0}^{i-1} \binom{30-3j}{3}}{i!} \times \left[\prod_{k=0}^{29-2i} (365-k) + \sum_{l=1}^{\left[\frac{30-3i}{2}\right]} \frac{\prod_{m=1}^{l} \binom{30-3i+2-2m}{2}}{l!} \prod_{n=0}^{29-2i-l} (365-n) \right]$$

so that $p_3 \approx 0.027998$ and $\mathbb{P}_4 \approx 5.4 \times 10^{-4}$. (We denote by $[r]$ the integer part of a real number r.) To summarize, the value of $\mathbb{P}_2$ told us that it is likely to have two alumni with the same birthday. The value of $\mathbb{P}_3$ tells us that it is rare to observe triplets and the value of $\mathbb{P}_4$ tells us that quadruplets are very unlikely. It is, however, noticeable how complicated the computation of $\mathbb{P}_3$ or $\mathbb{P}_4$ has been. Of course, the formulas of $\mathbb{P}_n$ are worse and rather counterintuitive. At this point, it is noticeable how simple and intuitive the computation of $\mathbb{E}C_n$ is. For $n \ge 3$, this computation gives us exactly the same information as the computation of $\mathbb{P}_n$, namely the unlikeliness of n-uplets. More striking is that the values of $\mathbb{P}_n$ and $\mathbb{E}C_n$ differ by a very small amount. They actually give exactly the same orders of magnitude! (See the table in Exercise 3.4.1.)

3.1.3 Visible and Invisible Alignments

We return now to more visual examples. Our aim is to evaluate how well vision, put in a random environment, perceives meaningful deviations from randomness. We will try to see where the threshold stands between visible and masked alignments. On the left of Figure 3.1.3, we display roughly 400 segments. This image has size $N_1 \times N_2 = 1000 \times 600$ pixels, and the mean length of the segments is $l \simeq 30$ pixels. Thus, their directional accuracy (computed as the width-length ratio) is $\pm 2/l$, which corresponds to about ± 4 degrees. Assuming that the directions and the positions of the segments are independent and uniformly distributed, we can compute a rough estimate for the expectation of the number of alignments of four segments or more (we say that segments are aligned if they belong to the same line, up to a given

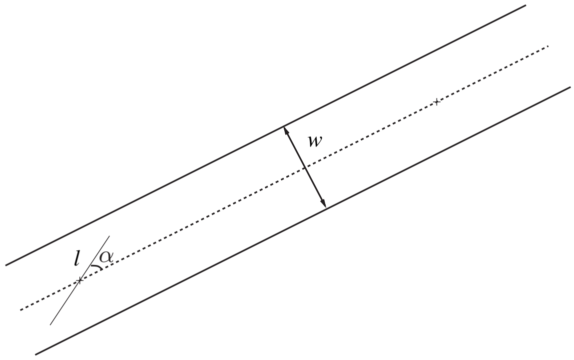

Fig. 3.2 The length l segment is said to be aligned with the dashed line with an accuracy w if the angle α is such that $\frac{l}{2}|\sin\alpha| \leq \frac{w}{2}$. Since l is much larger than w and since α is assumed to be uniformly distributed in $[-\frac{\pi}{2}, \frac{\pi}{2}]$, the probability of this event is roughly equal to $\frac{2w}{l\pi}$.

accuracy). Let M denote the number of segments, and let w denote the accuracy of the alignments (see Figure 3.2 for an illustration of this). In the following computations, we will take $w = 6$ pixels. If we consider a set of four segments denoted by S_1, S_2, S_3, and S_4, then the probability that they are aligned is roughly given by the probability that the centers of S_3 and S_4 fall at a distance less than $w/2$ from the line defined by the centers of S_1 and S_2. The relative area of the strip thus defined is approximatively $w/\max(N_1,N_2)$. Thus, this probability is $(w/\max(N_1,N_2))^2$, times the probability that the directions of the four segments are aligned with the direction of the strip. Since the segments are independent and the directions uniform, this last probability is $(2w/(l\pi))^4$. Thus, a rough estimate of the expectation of the number of alignments of four segments or more is

$$\binom{M}{4} \times \left(\frac{w}{\max(N_1,N_2)}\right)^2 \times \left(\frac{2w}{l\pi}\right)^4.$$

For the left image in Figure 3.1.3, the number of segments is $M = 400$. Using the previous formula, the expectation of the number of aligned 4-tuples of segments is about 10. It shows that we can expect some such alignments of four segments in this image. They are easily found by a computer program. Do you see them? On the right image, we performed the same experiment with about $M = 30$ segments, with the same accuracy. The expectation of the number of groups of four aligned segments is about $1/4000$. Most observers detect them immediately.

3.2 The Helmholtz Principle and ε-Meaningful Events

The three preceding examples have illustrated the promises of a general perception principle that we call the Helmholtz principle. We refer to Figure 3.2.2 for another illustration. The Helmholtz principle can be stated in the following generic way.

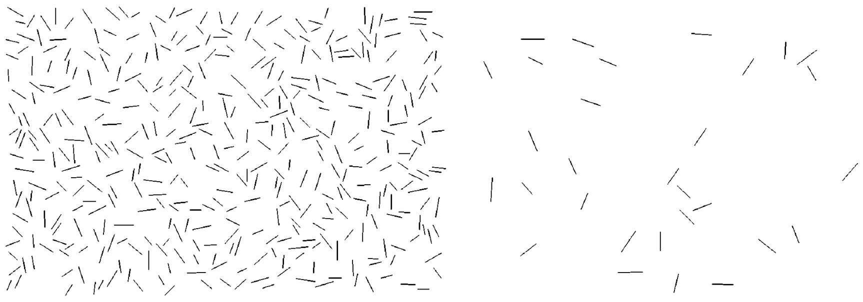

Fig. 3.3 The Helmholtz principle in human perception:

A group of four aligned segments exists in both images, but it can hardly be seen on the left-hand side image. Indeed, such a configuration is not exceptional in view of the total number of segments. In fact, the expectation of the number of aligned segments 4*-tuples is about* 10*. In the right-hand image, we immediately perceive the alignment as a large deviation from randomness that could hardly happen by chance. In this image, the expectation of the number of groups of four aligned segments is about* $1/4000$*.*

Assume that atomic objects O_1, $O_2, \ldots, O_n$ are present in an image. Assume that k of them, say $O_1, \ldots, O_k$, have a common feature (same color, same orientation, position, etc.). We then face a dilemma: Is this common feature happening by chance or is it significant and enough to group $O_1, \ldots, O_k$? To answer this question, let us make the following mental experiment: Assume *a priori* that the considered quality had been randomly and uniformly distributed on all objects $O_1, \ldots, O_n$. In the mental experiment, the observed position of objects in the image is a random realization of this uniform process. We finally ask the question: Is the observed repartition probable or not? If not, this proves *a contrario* that a grouping process (a gestalt) is at play. The Helmholtz principle states roughly that in such mental experiments, the numerical qualities of the objects are assumed to be uniformly distributed and independent.

Definition 1 (ε-meaningful event [DMM00]). *We say that an event that is ε-meaningful if the expectation of the number of occurrences of this event is less than ε under the a-contrario random assumption. When $\varepsilon \le 1$, we simply say that the event is meaningful.*

This definition is very generic. It must be completed by a discussion of perceptually relevant events. Adequate a-contrario models must also be given. In many cases, the a-contrario random assumption is that numerical qualities of objects are independent and uniformly distributed, but the a-contrario model can be more general.

If the Helmholtz principle is true, we perceive events if and only if they are meaningful in the sense of the preceding definition. The alignment in Figure 3.1.3 (right) is meaningful, whereas the left-hand figure contains no meaningful alignment of 4 segments.

The example of birthdays has explained why we prefer to detect unlikely events by estimating the expectation of their number instead of their probability. As an example of generic computation that we can do with the ε-meaningfulness definition, let us assume that the probability that a given object O_i has the considered quality is equal to p. In the case of birthdays, we had $p = \frac{1}{365}$ and in the black square example, $p = \frac{1}{2}$.

Under the independence assumption, the probability that at least k objects out of the observed n have this quality is

$$\mathcal{B}(n,k,p) = \sum_{i=k}^{n} \binom{n}{i} p^i (1-p)^{n-i}$$

that is, the tail of the binomial distribution. To get an upper bound for the number of false alarms (i.e. the expectation of the number of geometric events happening by pure chance), one simply multiplies the above probability by the number of tests performed on the image. This number of tests N_{conf} corresponds to the number of different possible configurations one could have for the searched gestalt. Thus, in most cases that we will consider in the next sections, a considered event will be defined as ε-meaningful if

$$\text{NFA} = N_{conf} \cdot \mathcal{B}(n,k,p) \leq \varepsilon.$$

We call NFA the left-hand member of this inequality. It stands for "number of false alarms". The NFA of an event measures the "meaningfulness" of this event. The smaller it is, the more meaningful the event is. (Good things come in small packages.)

The definition of meaningful events is, of course, related to the statistical framework of hypothesis testing and of multiple tests. We will discuss this link and also explain the differences in Chapter 15.

3.2.1 A First Illustration: Playing Roulette with Dostoievski

Dostoievski's *The Player* is all about the links of chance and destiny. The hero of the novel believes in some regularities in chance and also believes that he can detect them and win a long series. Twice in the novel, he comments on the *exceptional* event that on some day red came in 22 times in a row, which was unheard of. We quote from [Dos69]. We translate it as follows:

That time, as if on purpose, a circumstance arose which, incidentally, recurs rather frequently in gambling. Luck sticks, for example, with red and does not leave it for ten or even fifteen turns. Only two days before, I had heard that red had come out twenty two times in a row in the previous week. One could never recall a similar case at roulette and it was spoken of with astonishment.

And earlier in the novel he writes:

In the succession of fortuitous events, there is, if not a system, at least some kind of order. (...) It's very odd. On some afternoon or morning, black alternates with red, almost without any order and all the time. Each color only appears two or three times in a row. The next day or evening, red alone turns, for example, up to twenty times in a row.

Why 22? The probability that red appears 22 times in a row is $\left(\frac{18}{37}\right)^{22}$, namely about 10^{-7}. The computation of the probability that this happens in a series of n trials may be a bit intricate. We can, instead, directly compute the expected number of occurrences of the event as $\mathrm{NFA}(n) = (n-21) \times \left(\frac{18}{37}\right)^{22}$. The event is likely to happen if its NFA is larger than 1, which yields roughly $n \geq 10^7$. Thus, we are led to compute how many trials a passionate gambler may have done in his life. Considering that a professional gambler would play roulette at 100 evenings of 5 hours a year for 20 years, estimating in addition that a roulette trial may take about 30 seconds, we deduce that an experienced gambler would observe at the most, in his gambling life span, about $n = 20 \times 100 \times 5 \times 120 \simeq 10^6$ trials. We deduce that 1 out of 10 professional gamblers can have observed such a series of 22. Actually, Dostoievski's information about the possibility of 22 series is clearly based on conversations with specialists. The hero says:

I own a good part of these observations to Mr. Astley, who spends all of his mornings by the gambling tables but never gambles himself.

If this professional observer spent his time by several tables, maybe 10 simultaneously, he is, according to our computations, likely to have observed a series of 22. As we computed, 22 is somewhat a limit for an observable series. On the other hand, the hero mentions this occurrence as having happened just a few days before he was playing. There is no contradiction here, since, according to Aristotle, it is a rule of poetry, epics, and tragedy to put their heroes in exceptional situations. As he notices in his *Poetics*, exceptional situations do happen. Dostoievski twice puts his hero in an unlikely, but not impossible, situation. First, as we mentioned, is when a series of 22 occurs just a few days before the hero gets interested in roulette, second, a few days later, is when the hero observes a series of 14 reds and takes advantage of it to win a fortune. A series of 14 is unlikely to be observed by a beginner. The NFA of this happening to the hero during the three evenings he plays at the Roulettenbourg casino is, by the same kind of calculations as above, about $\mathrm{NFA} = 3 \times 5 \times 120 \times \left(\frac{18}{37}\right)^{14} \simeq 4.10^{-3}$. Thus, this event is unlikely, yet, again, not impossible and therefore fits Aristotle's criterion.

Our comments would be incomplete if we did not also notice that the gambler's perception obeys gestalt laws. According to Dostoievski, most of their observations of roulette focus on a very small number of specific kinds of series that are clearly the only ones likely to be perceived as exceptional. These specific series are, according to Dostoievski's comments, the following:

– long monochromatic series (reds or blacks);
– periodic or quasi-periodic series, namely two or three reds alternating with two or three blacks all the time.

Thus, we can rule out the main objection raised by the sceptic of NFA calculations. He argued that all possible long sequences are equally exceptional since they all have a very low probability. There would therefore be no surprise in an exceptional one happening, since one of the sequences must happen. In fact, the observers have a very small list of gestalts and perceive all other sequences as usual and not to be noticed. Our preceding estimates should, however, take into account the number of possible gestalts, not just monochromatic series. Following Dostoievski, we can estimate to 10 the various gambler's gestalts, namely:

– long enough series of red;
– long enough series of black;
– long enough series of alternate black and red;
– long enough series of alternate pairs black-black-red-red;
– long series of alternate triples;
– long enough series alternating one red and two blacks;
– long enough series alternating one black and two reds.

There may be a few more, but little more. Let us call N_g the number of such gestalts. Then we can calculate again the NFA of the event that "any of those gestalts is observed". This NFA simply is the former NFA multiplied by N_g and our conclusions remain valid.

3.2.2 *A First Application: Dot Alignments*

Dots in a dot image will be called aligned if they all fall into a strip thin enough and in sufficient number (see Figure 3.2.2.) Of course, the Helmholtz a-contrario

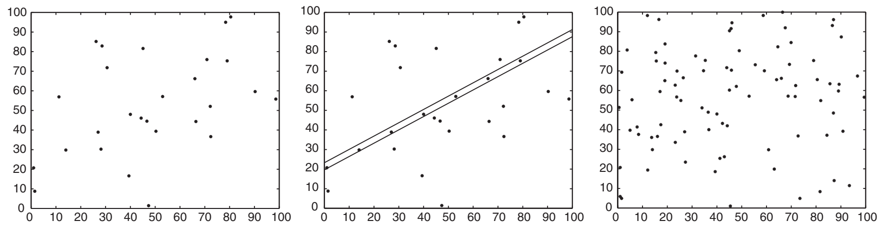

Fig. 3.4 The Helmholtz principle:

Noncasual alignments are automatically detected by the Helmholtz principle as a large deviation from randomness. Left: 20 uniformly randomly distributed dots and 7 aligned added. Middle: This meaningful and visible alignment is detected as a large deviation from randomness. Right: same alignment added to 80 random dots. The alignment is no more meaningful (and no longer visible). In order to be meaningful, it would need to contain at least 12 points.

assumption is that the dot positions are uniform, independent random variables, namely a uniform (Poisson) spatial distribution.

Let M be the number of dots in the image. The precision of the alignment is measured by the width of the strip. Let S be a strip of width a. Let $p(S)$ denote the prior probability for a point to fall in S, and let $k(S)$ denote the number of points (among the M) that are in S.

Definition 2. *A strip S is ε-meaningful if*

$$\mathrm{NFA}(S) = N_s \cdot \mathcal{B}(M, k(S), p(S)) \leq \varepsilon,$$

where N_s is the number of considered strips.

3.2.3 The Number of Tests

We now have to discuss what the considered strips will be, since we have to evaluate their number. A simple tiling argument shows that if the strip width is small with respect to the size of the image, then $N_s \simeq 2\pi(R/a)^2$, where R is the diameter of the image domain Ω and a is the minimal width of a strip. There is indeed about that number of strips to be tested if we want to ensure that any rectangle of the image with width less than $a/2$ is contained in at least one of the strips with width a. To be more generic, we should not, however, fix an arbitrary a. So one can sample all considered strip widths a in a finite logarithmic scale up to the smallest possible width. Thus, one obtains N_s as the total number of strips of all possible (quantized) widths. Then the final number of strips N_s only depends on the size of the image and this yields an unsupervised detection method. This is the first way to compute and test the possible strips.

Second testing method. Another way to define the actual tests that speeds up detection considerably and makes it perceptually realistic is to only consider strips whose endpoints are observed dots. In such a case, we obtain

$$N_s = \alpha \frac{M(M-1)}{2},$$

where α denotes the number of considered widths (about 10) and $\frac{M(M-1)}{2}$ is the number of pairs of points. Both methods for computing N_s are valid, but they do not give the same result! Clearly, the first method would be preferable in the case of a very dense set of points, assimilable to a texture, and the second method when the set of points is sparse. Notice, however, the slight obvious change in the computation of $k(S)$. It denotes the number of dots that fell into the strip, with the exception, of course, of the two endpoints defining the strip.

At this point, we must address an objection: are we not cheating and choosing the theory that gives the better result? We have two possible values for N_s and the smallest N_s will give the largest number of detections. When two testing methods are available, perception must obviously choose the one giving the smaller test number. Indeed, there is perceptual evidence that grouping processes may depend on density and that different methods could be relevant for dense and for sparse patterns. Hence, the second testing method should be preferred for sparse distributions of points, whereas the initial model based on density would give a smaller number of tests when the number of points is large. This economy principle in the number of tests has been developed in recent works by Donald Geman and his collaborators [FG01][BG05].

Let us compare both definitions of object alignments in the examples of Figure 3.2.2. When we use the larger N_s corresponding to the all strips with all widths (from 2 to 12 pixels), *we simply do not detect any alignment*. Indeed, for this image (size 100×100), we have $R = 100\sqrt{2}$ and thus $N_s = \sum_{a=2}^{12} 2\pi(R/a)^2 \simeq 10^5$. On the other hand the alignment of 7 points is included in a strip with width $a = 3$ and thus has a probability $\mathcal{B}(M, 7, a/100)$, which has value $\simeq 10^{-5}$ when $M = 27$ and has value $\simeq 10^{-2}$ when $M = 87$. Thus, in both cases, the alignment is not meaningful. This is due to the testing overdose: We have tested many times the same alignments and have also tested many strips that contained no dots at all. The second definition of N_s happens to give a perceptually correct result. One has $N_s \simeq 3 \times 10^3$ for the image with $M = 27$ points and thus the alignment becomes meaningful. For the image with $M = 87$ points $N_s \simeq 4 \times 10^5$ and the alignment is not meaningful since its NFA is larger than 1. This result is displayed in Figure 3.2.2 in the middle, where we see the only detected strip. This same alignment is no more detectable on the right. The tested widths range from 2 to 12; strips thinner than 2 pixels are nonrealistic in natural (nonsynthetic) images and strips larger than 12 no longer give the appearance of alignments in a 100×100 image.

3.3 Bibliographic Notes

The program stated here has been proposed several times in Computer Vision. We know of at least two instances: David Lowe [Low85] and Witkin-Tenenbaum [WT83]. Here we quote extensively David Lowe's program, whose mathematical consequences are developed in this book.

We need to determine the probability that each relation in the image could have arisen by accident, $P(a)$. Naturally, the smaller that this value is, the more likely the relation is to have a causal interpretation. If we had completely accurate image measurements, the probability of accidental occurrence could become vanishingly small. For example, the probability of two image lines being exactly parallel by accident of viewpoint and position is zero. However, in real images there are many factors contributing to limit the accuracy of measurements. Even more important is the fact that we do not want to limit ourselves to perfect instances of each relation in the scene – we want to be able to use the information available from even approximate instances of a relation. Given an image relation that holds within some degree *of accuracy, we wish to calculate the probability that it could have arisen by accident to within that level of accuracy. This can only be done in the context of some assumption regarding the surrounding distribution of objects, which serves as the null hypothesis against which we judge significance. One of the most general and obvious assumptions we can make is to assume that a background of independently positioned objects in three-space, which in turn implies independently positioned projections of the objects in the image. This null hypothesis has much to recommend it. (...) Given the assumption of independence in three-space position and orientation, it is easy to calculate the probability that a relation would have arisen to within a given degree of accuracy by accident. For example if two straight lines are parallel to within 5 degrees, we can calculate that the chance is only* $5/180 = 1/36$ *that the relation would have arisen by accident from two independent objects.*

Some main points of the program that we will mathematically develop are contained in the preceding quotation, particularly, the idea that significant geometric objects are the ones with small probability and the idea that this probability is anyway never zero because of the inherent lack of accuracy of a digital image. However, the preceding program is not accurate enough to give the right principles for computing gestalt. The above-quoted example is not complete. Indeed we simply cannot fix *a priori* an event such as "these two lines are parallel" without merging it into the set of all events of the same kind – that is, all possible groups of parallel lines in the considered image. If the image has many lines, it simply likely that two of them will be quite parallel. So we have to take into account the number of possible pairs of parallel lines. If this number is large, then we will, in fact, detect many non-significant pairs of parallel lines. Only if the expected number of such pairs is much below 1, can one decide that the observed parallelism makes sense. Although, in accordance with the former quotation, the general principle proposed in this chapter should be attributed to Lowe, it is also stated by Zhu in [Zhu99] and attributed to Helmholtz [vH99]: *Besides Gestalt Psychology, there are two other theories for perceptual organization. One is the likelihood principle [vH99] which assigns a high probability for grouping two elements such as line segments, if the placement of the two elements has a low likelihood of resulting from* accidental arrangement. Viewed that way, the Helmholtz principle is exactly opposite to the so-called *Prägnanz* principle in gestalt psychology : *"...of several geometrically possible organizations that one will actually occur which possesses the best, simplest and most stable shape"*, quoted in [Zhu99] from Koffka's book [Kof35].

3.4 Exercise

3.4.1 Birthdays in a Class

Consider a class of 30 students and assume that their birthdays are independent and uniformly distributed variables over the 365 days of the year. We call, for $1 \leq n \leq 30$, C_n the number of n-tuples of students of the class having the same birthday.

(This number is computed exhaustively by considering all possible n-tuples. If (for example) students 1, 2, and 3 have the same birthday, then we count three pairs, $(1,2)$, $(2,3)$, $(3,1)$.) We also consider $\mathbb{P}_n = \mathbb{P}(C_n \geq 1)$, the probability that there is at least one n-tuple with the same birthday and p_n, the probability that there is at least one n-tuple and no $(n+1)$-tuple.

1) Prove that $\mathbb{P}_n = 1 - \sum_{i=1}^{n-1} p_i$ and $\mathbb{P}_n = \mathbb{P}_{n-1} - p_{n-1}$.

2) Prove that $\mathbb{E}C_n = \frac{1}{365^{n-1}}\binom{30}{n}$. Check that $\mathbb{E}C_2 \approx 1.192$, $\mathbb{E}C_3 \approx 0.03047$, *and* $\mathbb{E}C_4 \approx 5.6 \times 10^{-4}$.

3) Prove that $\mathbb{P}(C_2 = 0) = \frac{365\times\cdots\times 336}{365^{30}} \approx 0.294$. Deduce that $P_2 \approx 0.706$.

4) Prove that

$$p_2 = \frac{1}{365^{30}} \sum_{i=1}^{15} \frac{\prod_{j=1}^{i} \binom{32-2j}{2}}{i!} \prod_{k=0}^{29-i} (365-k).$$

5) Compute by a small computer program (in Matlab for example): $p_2 \approx 0.678$.

6) Deduce that $\mathbb{P}_3 \approx 0.0285$.

7) We denote by $[r]$ the integer part of a real number. Prove that

$$\begin{aligned} p_3 = {} & \frac{1}{365^{30}} \sum_{i=1}^{10} \frac{\prod_{j=1}^{i} \binom{33-3j}{3}}{i!} \\ & \times \left[\prod_{k=0}^{29-2i} (365-k) + \sum_{l=1}^{\left[\frac{30-3i}{2}\right]} \frac{\prod_{m=1}^{l} \binom{30-3i+2-2m}{2}}{l!} \prod_{n=0}^{29-2i-l} (365-n) \right]. \end{aligned}$$

8) Deduce by a computer program that $p_3 \approx 0.027998$ and $\mathbb{P}_4 \approx 5.4 \times 10^{-4}$.

9) Be courageous and give a general formula for p_n.

10) Prove that $\mathbb{E}C_{30} = \mathbb{P}_{30} = \frac{1}{365^{29}}$, $\mathbb{E}C_{29} = \frac{30}{365^{28}}$, and $\mathbb{P}_{29} = \frac{30\times 364+1}{365^{29}}$.

11) The following table summarizes the comparative results for $\mathbb{E}C_n$ and P_n as well as the relative differences. Check it.

n	$\mathbb{E}C_n$	$\mathbb{P}_n$	$\frac{\mathbb{E}C_n - \mathbb{P}_n}{\mathbb{P}_n}$
2	1.192	0.706	68.84%
3	0.0347	0.0285	21.75%
4	5.6×10^{-4}	5.3×10^{-4}	5.66%
...	...	...	...
29	$\frac{30}{365^{28}}$	$\frac{30\times 364+1}{365^{29}}$	0.27%
30	$\frac{1}{365^{29}}$	$\frac{1}{365^{29}}$	0%

12) Explain why $\mathbb{P}_n$ and $\mathbb{E}C_n$ are so close for $n \geq 3$.

Springer

http://www.springer.com/978-0-387-72635-9

From Gestalt Theory to Image Analysis
A Probabilistic Approach
Desolneux, A.; Moisan, L.; Morel, J.-M.
2008, Hardcover
ISBN: 978-0-387-72635-9