École Normale Supérieure de Cachan

On salience and non-accidentalness: comparing human vision to *a contrario* algorithms

A dissertation presented by

Samy Blusseau

in fulfillment of the requirements
for the degree of Doctor of Philosophy
in the subject of Applied Mathematics

Committee in charge

Referees	Andrés Almansa	-	Télécom ParisTech, FR
	Laurent Perrinet	-	Aix-Marseille Université, FR
	Johan Wagemans	-	KU Leuven, BE
Advisors	Jean-Michel Morel	-	ENS de Cachan, FR
	Rafael Grompone von Gioi	-	ENS de Cachan, FR
Examiners	Yann Gousseau	-	Télécom ParisTech, FR
	Alessandro Sarti	-	EHESS, FR

September 2015

Abstract

The present dissertation compares the human visual perception to computer vision algorithms based on a mathematical model called *a contrario* theory. To this aim, it focuses on two visual tasks that are at the same time easy to model and convenient to test in psychophysical experiments. Both tasks consist in the perceptual grouping of oriented elements, namely Gabor patches. The first one is the detection of alignments and the second one extends to curves, that is to say to more general arrangements of elements in good continuation. In both cases, alignments and curves, psychophysical experiments were set up to collect data on the human visual perception in a masking context.

The non-accidentalness principle states that spatial relations are perceptually relevant when their accidental occurrence is unlikely. The *a contrario* theory is a formalization of this principle, and is used in computer vision to set detection thresholds accordingly. In this thesis, the *a contrario* framework is applied in two practical algorithms designed to detect non-accidental alignments and curves respectively. These algorithms play the part of artificial subjects for our experiments.

The experimental data of human subjects is then compared to the detection algorithms on the very same tasks, yielding two main results. First, this procedure shows that the Number of False Alarms (NFA), which is the scalar measure of non-accidentalness in the *a contrario* theory, strongly correlates with the detection rates achieved by human subjects on a large variety of stimuli. Secondly, the algorithms' responses match very well the average behavior of human observers.

The contribution of this thesis is therefore two-sided. On the one hand, it provides a rigorous validation of the *a contrario* theory's relevance to estimate visual thresholds and implement visual tasks in computer vision. On the other hand, it reinforces the importance of the non-accidentalness principle in human vision.

Aiming at reproducible research, all the methods are submitted to IPOL journal, including detailed descriptions of the algorithms, commented reference source codes, and online demonstrations for each one.

Résumé

Dans cette thèse nous comparons la vision humaine à des algorithmes de vision par ordinateur basés sur un modèle mathématique appelé théorie *a contrario*. Pour cela, nous nous concentrons sur deux tâches visuelles dont la modélisation d'une part, et l'expérimentation psychophysique d'autre part, sont simples. Celles-ci relèvent du groupement perceptuel d'éléments orientés appelés patchs de Gabor. Dans la première tâche il s'agit de détecter des alignements, et dans la seconde des courbes, soit des configurations plus générales dans lesquelles les éléments sont en bonne continuation. Dans les deux cas, des expériences psychophysiques ont été menées afin de collecter des données sur la perception visuelle humaine dans un contexte de masquage.

Le principe de non-accidentalité désigne le fait que les relations spatiales entre des éléments prennent un sens pour la perception lorsqu'il semble invraisemblable qu'elles soient le fruit du hasard. Ce principe trouve une formalisation dans la théorie *a contrario*, qui est utilisée en vision par ordinateur pour déterminer des seuils de détection en accord avec la non-accidentalité. Les méthodes *a contrario* sont appliquées ici dans l'implémentation de deux algorithmes, conçus pour détecter respectivement des alignements et des courbes non-accidentels. Ces algorithmes jouent le rôle de sujets artificiels dans nos expériences.

Les données expérimentales des sujets humains ont donc été comparées aux algorithmes sur les mêmes taches, conduisant à deux principaux résultats. Premièrement, le Nombre de Fausses Alarmes (NFA), qui est la mesure de non-accidentalité dans la théorie *a contrario*, est en forte corrélation avec les taux de détection obtenus par les sujets humains sur un large éventail de stimuli. Deuxièmement, les réponses des algorithmes sont très similaires à celles de la moyenne des sujets humains.

La contribution de cette thèse est donc double. D'une part, elle valide de façon rigoureuse la pertinence des méthodes *a contrario* dans l'estimation de seuils perceptuels, et leur application en vision par ordinateur. D'autre part, elle souligne l'importance du principe de non-accidentalité dans la vision humaine.

Dans le but de rendre ces recherches reproductibles, les méthodes décrites dans la thèse sont soumises pour publication dans le journal IPOL. Ces publications fournissent le détail des algorithmes, leur code source commenté, ainsi qu'une démonstration en ligne pour chacun d'eux.

Remerciements

Les années de thèse ont été pour moi une période privilégiée, durant laquelle il m'a été offert d'apprendre en me consacrant pleinement à un projet unique, et ce dans les meilleures conditions. Je suis sincèrement reconnaissant envers tous ceux qui ont participé à me donner cette chance.

J'ai trouvé au CMLA un cadre idéal pour travailler dans la sérénité et la bonne humeur. Je le dois en grande partie à Carine, Christophe, Micheline, Nicolas P., Sandra, Véronique et Virginie, dont la bienveillance et la compétence ont assuré le bon déroulement de mon séjour au laboratoire, tant sur les plans logistique et administratif qu'humain. Je remercie également tous les professeurs du CMLA pour leur accessibilité et l'image positive qu'ils m'ont donnée du métier d'enseignant-chercheur.

Mes remerciements vont tout particulièrement à mes directeurs Jean-Michel Morel et Rafael Grompone von Gioi, qui m'ont laissé la plus grande liberté dans mon travail tout en me prévenant d'éventuels égarements par leurs idées et leurs conseils. Ils ont fait preuve d'une disponibilité exceptionnelle que je souhaite à tous les doctorants. Par ailleurs, je garderai un souvenir enthousiaste du groupe "images" du CMLA, où la qualité scientifique va de paire avec un esprit de partage et d'entraide. Merci également aux chercheurs du CibPSI, en particulier Alejandra Carboni et Alejandro Maiche, et à Gregory Randall qui a encouragé cette collaboration avec l'Universidad de la República de Montevideo.

Les journées au laboratoire n'auraient pas été aussi plaisantes sans mes camarades du CMLA: entre *risotti*, débats et plaisanteries, la cuisine a été le lieu des meilleures pauses grâce à vous.

Enfin, merci à toutes celles et tous ceux que je devrais remercier chaque jour - ils se reconnaîtront.

Contents

1 Introduction **11**
1.1 Motivation . . . 11
1.2 Salience of alignments . . . 14
1.3 Salience of curves . . . 17
1.4 Summary of the contributions . . . 21

2 Psychophysics, Gestalts and Games **23**
2.1 Introduction . . . 23
2.2 The non-accidentalness principle in vision . . . 25
2.3 Introduction to the *a contrario* theory . . . 28
2.4 Detection Theory versus Gestaltism . . . 34
2.5 Detection Theory versus Psychophysics . . . 39
2.6 Conclusion . . . 46

3 Measuring the salience of alignments by their non-accidentalness **47**
3.1 Introduction . . . 47
3.2 Methods . . . 49
3.3 Experiment I ($N = 200$) . . . 51
3.4 Experiment II ($N = 100$) . . . 55
3.5 Experiment III ($N = 600$) . . . 55
3.6 Model and algorithm . . . 56
3.7 Subjects compared to the algorithm . . . 61
3.8 Discussion . . . 64
3.9 Conclusion . . . 67

4 Measuring the salience of curves by their non-accidentalness **69**
4.1 Introduction . . . 69
4.2 Methods . . . 71
4.3 Experiment I . . . 74
4.4 Experiment II . . . 79
4.5 Model and algorithm . . . 83
4.6 Subjects compared to the algorithm . . . 88
4.7 Discussion . . . 96
4.8 Conclusion . . . 97

5 Conclusion and perspectives **99**

1 Introduction

1.1 Motivation

Since the emergence of the field of Computer Vision [57] about fifty years ago – initially as a branch of the Artificial Intelligence working with robots and its artificial senses – there have been many attempts at formalizing vision theories and especially the one formulated by the Gestalt school of psychology. Its members, Wertheimer, Köhler, Koffka, Kanizsa among others [90, 41, 12, 58, 37], developed from the twenties to the eighties an original *modus operandi*, leading to the conclusion that the first steps of visual perception are based on a reduced set of geometrical grouping laws [88, 89]. These laws, also called *Gestalts*, describe the configurations in which most human observers can't help interpreting different elements as one shape or group. Figure 1.1 illustrates the laws of proximity, symmetry and good continuation.

Desolneux *et al.* [11] designed the *a contrario* theory as one of the attempts to provide mathematical foundations to these laws. Several properties of the *a contrario* framework argue in favor of its suitability for visual perception modeling.

First, it applies Attneave's principle, stating that we do not perceive any structure in white noise [2]. Computer vision algorithms based on the *a contrario* theory, like for example the Line Segment Detector [27, 28], typically avoid detections in images like Figure 1.2. This idea is also consistent with the more general *non-accidentalness* principle, introduced by Witkin, Tenenbaum and Lowe [48, 96, 95], and of great importance in the study of perception. This principle states that spatial relations are perceptually relevant when their accidental occurrence is unlikely. Therefore, in accordance with Attenave's statement, these meaningful spatial relations should not appear by chance in a random image like Figure 1.2, in which all pixels are independent.

Another strength of the *a contrario* theory is that it allows the design of parameterless detection algorithms. These make as few assumptions as needed to adapt *automatically* to a large variety of inputs. Similarly, our visual system is able to perform equally well in very different conditions, without the need to re-learn, each time, priors on the world.

Promising results have already been obtained by comparing *a contrario* algorithms to perception [9, 11], and may be confirmed and extended by further research. This is precisely one of the lines we intend to follow in the present dissertation.

To link the Gestaltist qualitative description of perceptual grouping laws to a mathematical model, a quantitative gap must be bridged. Human perceptual behavior has been the subject of quantitative experimentation since the times of Fechner, the founder of Psychophysics. This relatively new science investigates the relationship between the stimulus intensity and the perceived sensation [78]. Concerning visual perception, psychophysical studies have added quantitative measurements to the qualitative observations made by the Gestaltists. Among them, one finds experiments investigating the contour integration phenomenon, and in particular the role of the Gestaltic

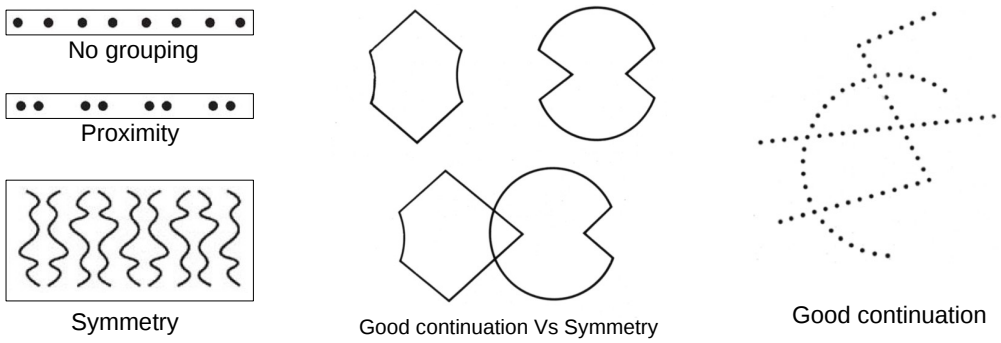


Figure 1.1: Illustration of some of the grouping laws defined by the Gestalt school of psychology. The left hand image, adapted from [65], shows how proximity and symmetry forces us to group dots and lines two by two. The right hand drawing is a figure from [36], which we interpret as three smooth lines because of the good continuation grouping law. The central image, adapted from [37], illustrates relationships between grouping laws when several of them occur simultaneously. In this case, when the two symmetrical shapes are joined, we can't help seeing two shapes with more regular contours.

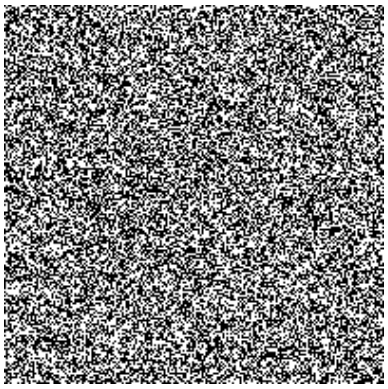


Figure 1.2: An example of white noise as in Attneave's experiment [2]. The state of each pixel was determined randomly and independently from the others, with the same odds for black and white. Such an image gives an impression of homogeneity, and no particular pattern emerges from it.

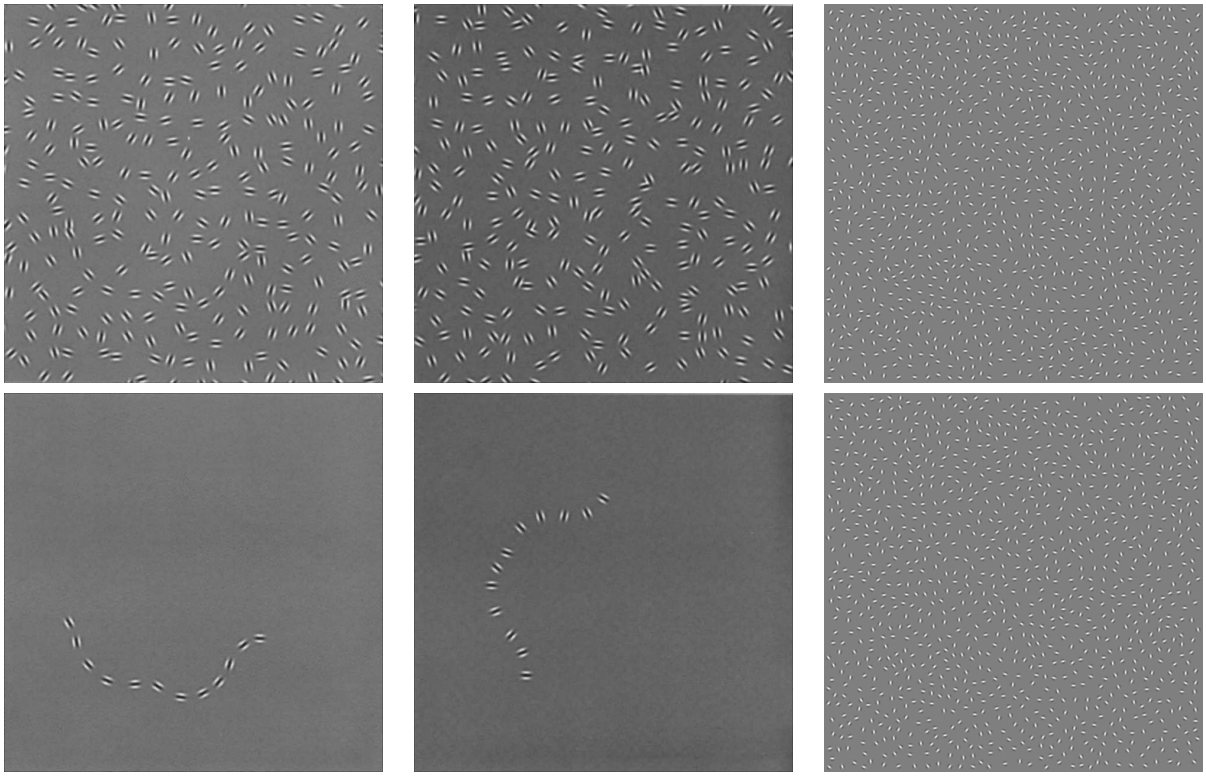


Figure 1.3: Examples of arrays of Gabor patches used to study good continuation in [21] (first two columns) and [73] (third column). Whereas in the upper image of the first column most observers perceive the smooth path reproduced just below, this task becomes much harder in the second column, although the path represented in the bottom image is present in the top one. Similarly, the shape of a bell, visible in the upper image of the right hand column, vanishes in the bottom image although the same elements are present, but with random orientations.

good continuation grouping law (fig. 1.1), which can be rephrased in Palmer's words: "All else being equal, elements that can be seen as smooth continuations of each other tend to be grouped together" [64, p. 259]. In this thesis, we will focus on that particular field of Psychophysics, hence good continuation will be the perceptual grouping law of interest.

More specifically, we will build on a method introduced in the early 90s [21, 42] and still used in the study of good continuation [44, 63, 55, 54, 84, 74]. It consists in embedding discrete contours in a cluttered background of Gabor patches, and measuring the ability of human observers to perceive such contours, as a function of their properties. As illustrated in Figure 1.3, a result that is common to the previously mentioned experiments is that the visual system is good at detecting smooth paths formed by elements that are roughly oriented like the local tangent of the contour. These studies also reported decreasing detection performance when contours were deviating from this ideal configuration. An important step towards the understanding of the mechanisms underlying these observations is the definition of a model that could *explain* and *predict* these results obtained experimentally.

With that purpose in mind, the second motivation of this thesis is the conjecture that the *a contrario* framework, as a mathematical formulation of the non-accidentalness principle, is particularly suited to provide such quantitative interpretations. Our proposal is precisely to design a model able to evaluate *a priori* the detectability, by human perception, of linear structures like

those of Figure 1.3, based on given parameters such as the total number of elements in the image, the number of elements composing the target path, and their orientations.

To that aim, the methods presented in the dissertation will stand on the concept of *Turing Test* [81]. Alan Turing proposed in 1950 to define the machine intelligence through what he called the *Imitation Game*: a machine that eventually could not be distinguished from humans should be considered intelligent. Similarly, a good model of visual perception should translate into an algorithm able to mimic human subjects in visual tasks. Therefore, we propose a research methodology to develop machine vision algorithms on the one hand, and quantitative psychophysical protocols on the other hand, so that computer and human vision match on certain classes of detection tasks. As we said earlier, we will use the *a contrario* framework to define artificial observers, and we will focus on the detection of good continuation in arrays of Gabor patches. We will first study the salience of alignments, and then apply our methods to more general contours.

1.2 Salience of alignments

Perhaps the simplest possible gestalt to start with is the perception of alignments. Building on methods illustrated in Figure 1.3, we will work with arrays of Gabor patches as well, but the embedded contours will be rectilinear ones, as shown in Figure 1.4, left hand column. In our stimuli, each array contains a total number N of patches, n of which are aligned and regularly spaced. Whereas the background elements are randomly oriented, the aligned elements are assigned the orientation of their line, affected by a random angular noise of variable intensity, called *angular jitter*.

In Figure 1.4, the displayed elements are centered on the same positions in all three images. Whereas an alignment is perceptible in the left hand image because the aligned elements are affected by very little angular jitter, the same alignment vanishes in the right hand image, where the angular jitter has maximal level. As illustrated by the central image of the same figure, the patches' coordinates are set to fill the image homogeneously and to make the alignment almost impossible to detect from their positions only. Consequently, the difficulty to detect the alignments in such stimuli is essentially ruled by the total number N of patches, the number n of aligned elements, and by the angular jitter affecting their orientations.

In Chapter 3 we present our first complete experimental set up, which involved 32 subjects and was based on the stimuli we just described. It consisted in three versions of a straight contours

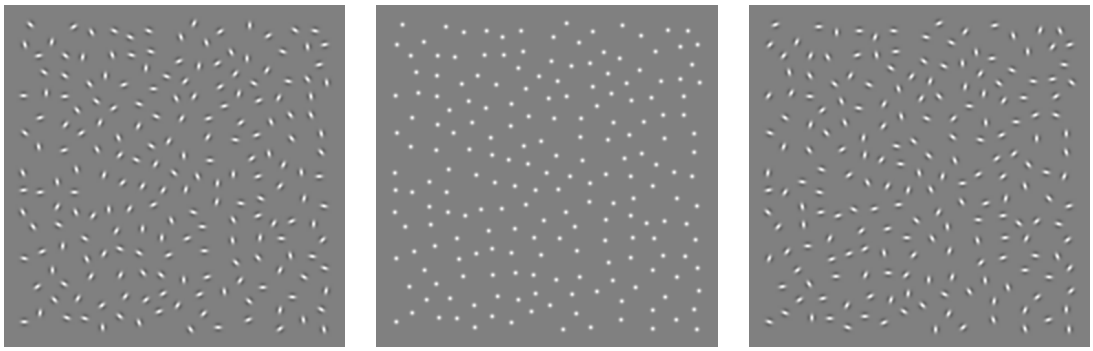
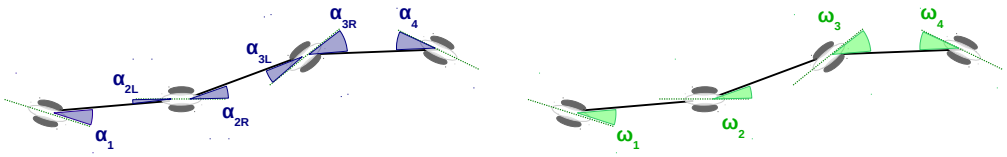


Figure 1.4: Illustration of the perception of alignments in arrays of Gabor patches. The displayed elements have the same positions in all three images, but an alignment is perceptible only in the left hand image, where the aligned elements have the same orientation as their line.



(a) (b)

Figure 1.5: In Chapter 3 we present an algorithm designed to detect non-accidental alignments in arrays of Gabor patches. Here we represent the variables the algorithm measures when analyzing a chain of patches. For a given chain, the angles ω, represented in (b) are deduced from the angles α, shown in (a), as follows: $\omega_1 \stackrel{\text{def}}{=} \alpha_1$, $\omega_n \stackrel{\text{def}}{=} \alpha_n$ and $\omega_i \stackrel{\text{def}}{=} \max(\alpha_{iL}, \alpha_{iR})$ otherwise.

detection task, corresponding to three tested values for N: $N = 100, 200$ and 600 Gabor patches per image[1].

In each trial of these experiments, the subjects were shown on a screen an array of Gabor patches like those of Figure 1.6, top row. They knew that every stimulus contained an alignment, but did not know where, how long and how jittered it was. They were asked to click on an element they perceived as part of the alignment. Each click made by a subject was associated to the nearest Gabor element in the image, and counted as a detection when that element belonged to the target alignment. The reaction time was measured as well.

As will be discussed, the results were mostly as expected, because they were consistent with those obtained in previous studies on the influence of orientation jitter on contour detection. The more jittered the alignments, the lower the average detection performance, and the longer the subjects took in looking for them. It was also harder to detect short alignments than long ones. However, as will be pointed out, we observed two unexpected effects. First, subjects performed above chance level in the cases of maximal jitter level, which indicated insufficient masking of the target alignments' coordinates in some cases. Second, the detection rate was not a decreasing function of the number N of patches per image, since the worst average performance was measured for $N = 200$. We will see that this could be related to the use, in the latter case, of a different spatial frequency in the definition of the Gabor functions.

Subsequently, we present an *a contrario* algorithm, which was designed to detect non-accidental alignments in our stimuli. The algorithm analyses, in each stimulus, a set of chains that might be significant alignments. For every chain, it considers the variables $\omega_1, \dots, \omega_n$ (Fig. 1.5(b)) defined from the angles represented in Figure 1.5(a), as $\omega_1 \stackrel{\text{def}}{=} \alpha_1$, $\omega_n \stackrel{\text{def}}{=} \alpha_n$ and $\omega_i \stackrel{\text{def}}{=} \max(\alpha_{iL}, \alpha_{iR})$ otherwise. Then the maximum of these angular deviations $\omega^* \stackrel{\text{def}}{=} \max\{\omega_1, \dots, \omega_n\}$ is computed and the algorithm calculates the Number of False Alarm of a chain c as

$$\text{NFA}(c) \quad \stackrel{\text{def}}{=} \quad N_T \times \mathbb{P}\left(\Omega^* \leq \omega^*\right), \tag{1.1}$$

where N_T is the total number of tested chains, and the second factor is the probability that a lower deviation than ω^* occur under a null hypothesis. The algorithm then returns the chain with smallest NFA. We will show that, in a random array of Gabor patches, the expected number of chains c such that $\text{NFA}(c) \leq \varepsilon$, is less than ε, for any $\varepsilon > 0$ (Theorem 5). This is what we will call the *non-accidentalness property* of the NFA. Its consequence is that a chain with a small NFA, typically $\text{NFA}(c) \leq 1$, is a rare event to be observed by chance less than once per image, and is therefore expected to be detected. In turn, such a small NFA corresponds to small changes in direction and

[1]These experiments can be tried online at `http://bit.ly/na_alignments`

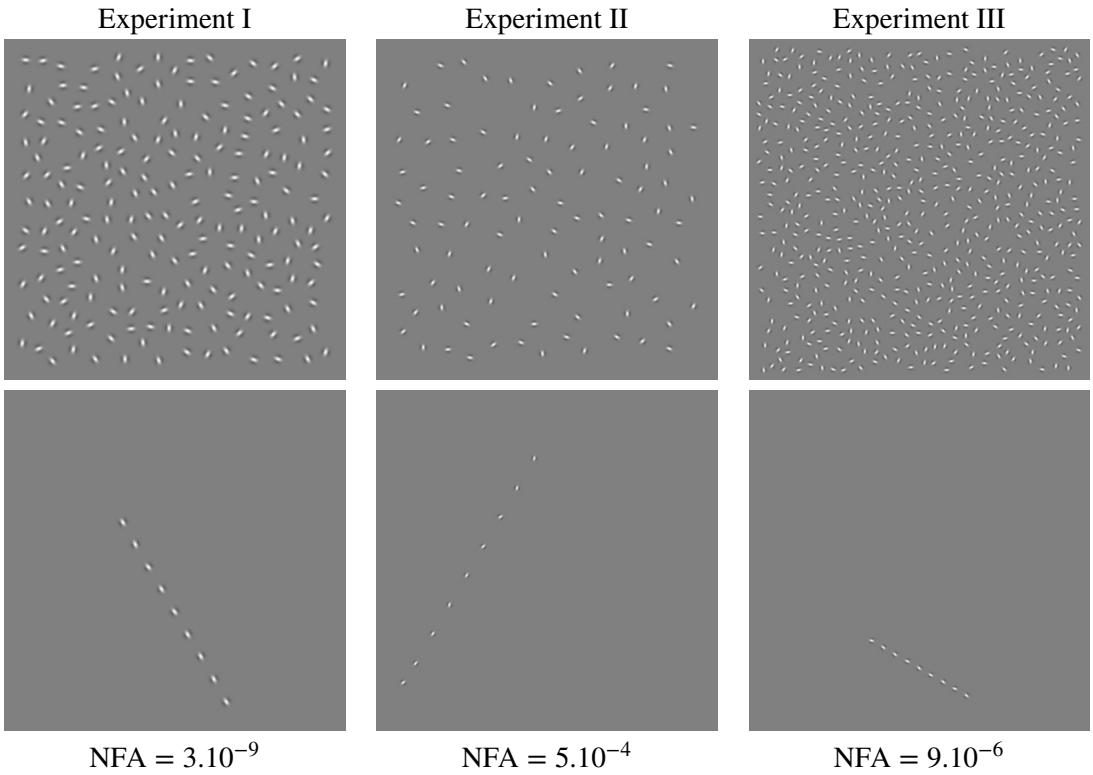


Figure 1.6: Top row: examples of stimuli used in the experiments of Chapter 3; bottom row: the corresponding output of our detection algorithm, along with the NFA of the detected chains.

Gabor patches roughly tangent to their chain. Figure 1.6 provides examples of detections by the algorithm.

Then, to compare the perceptual results obtained by human subjects to the algorithm, viewed as an artificial subject, the algorithm's "clicked point" is defined for each stimulus as the most central element among the returned chain. In Figure 1.7, the results of the subjects of all three experiments are averaged and compared to the algorithm. The first and second plots show the detection rate as a function of the jitter level and the number of aligned elements respectively, whereas the third and fourth columns display respectively the detection rate and the reaction time as functions of $\log_{10}(\text{NFA})$ of the target alignment. In the last sections of Chapter 3, the apparent fit between subjects and algorithm will be confirmed by statistical evidence, and we will also discuss the observed differences.

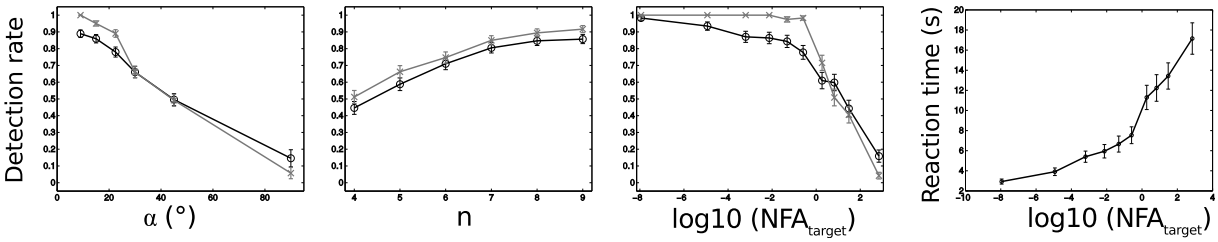


Figure 1.7: Comparison of the subjects of Chapter 3's experiments to the *a contrario* detection algorithm for alignments. The solid black lines represent the subjects, whereas the gray solid lines represent the algorithm's results. The data shown here is the average over all three experiments. The detection rates are plotted as functions of the jitter intensity α (first column), the number n of aligned elements (second column) and the $\log_{10}$ of the target's NFA (third column). The latter is also the x variable in the last column, the y axis representing the reaction time.

1.3 Salience of curves

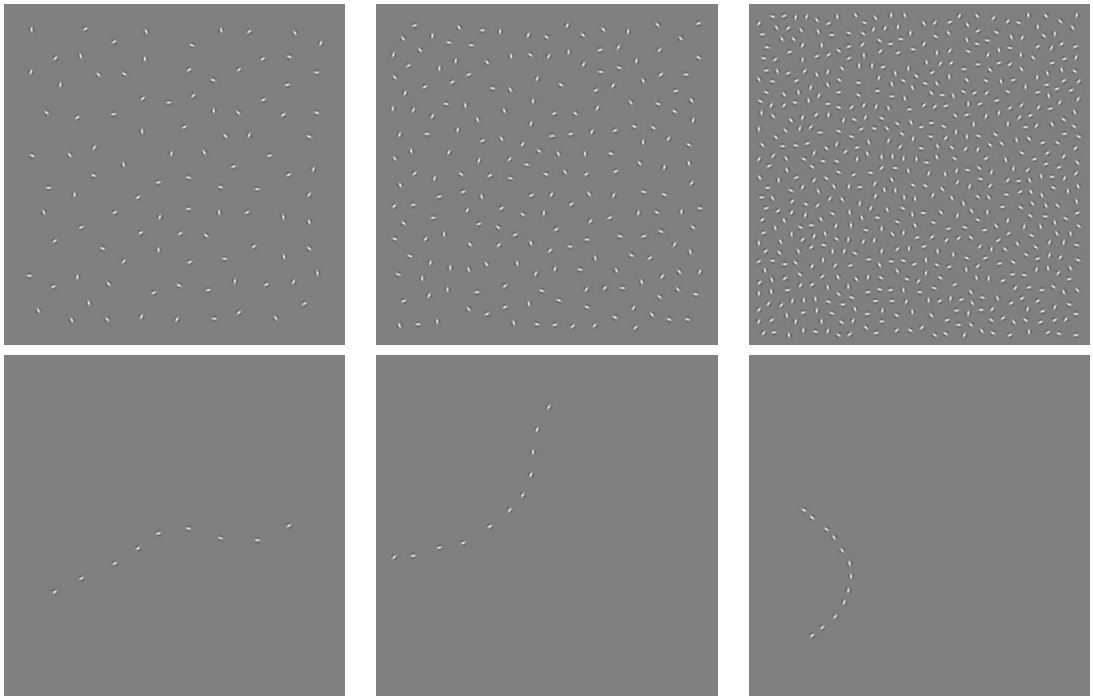
Encouraged by the results obtained on the salience of alignments, we will try to get closer to our initial target: a quantitative prediction of the perception of good continuation, intended in a more general sense than straight contours only, as illustrated by Figure 1.3. To that aim, the framework developed in Chapter 4 is very similar to the methods of Chapter 3, with an attempt to overcome certain flaws pointed out in the latter, and to strengthen its conclusions.

In particular, we define a new class of good continuation stimuli, for which there is no masking issue in the setting of the coordinates of target and background elements. For a given stimulus, all N points are set according to a unique process designed to ensure a homogeneous filling of the image. The target path is *then* chosen from this initial set, as a chain of n neighboring points forming a smooth path. Finally, stimuli such as those illustrated in Figure 1.8 are obtained by orienting the target elements like the local tangents of the path, and setting the background orientations randomly. Obviously, these ideal orientations for the target patches are affected by an angular jitter of variable intensity. Therefore, as for alignments, the detectability of the paths depends again on the total number N of patch per image, the quantity n of elements composing the target path, and the level of angular jitter. The same three values of N are tested (100, 200 and 600), and again the subjects are asked to click on an element they perceived as part of a smooth path.

In Chapter 4 we present two experiments. Both include the just mentioned three kinds of stimuli, and differ from each other in their procedures and apparatus. As will be explained, the first experiment was run on a web interface[2], and the subjects saw three sequences of images containing each time the same number N of patches; whereas the second experiment was programmed with PsychToolBox, run in a controlled environment, and its subjects saw three sequences of images with a different value of N in each sequence. Despite these variations, no significant difference in the detection rates is found. As expected and in both cases, the detection rate appears to increase with n, but to decrease with growing jitter levels and, this time, with N.

Let us get now to the *a contrario* algorithm that was compared to the perception of our subjects. In short, it performs a search among a set of smooth chains, in a similar fashion as for alignments. However, as illustrated by Figure 1.9, here it focuses on variables of two different natures: angles α_is on the one hand, and the distances ℓ_is between adjacent elements on the other hand. To handle the latter, we assumed that the ℓ_is are independent and approximated their cumulative distribution function by an empirical histogram $\tilde{F}$, estimated directly on the data. The algorithm then merges

[2]The experimental sessions can be tested online at `http://bit.ly/ac_curves`.



(a) NFA = 0.028 (b) NFA = 0.29 (c) NFA = 0.36

Figure 1.8: Examples of stimuli used in the experiments of Chapter 4, and the corresponding output of our detection algorithm.

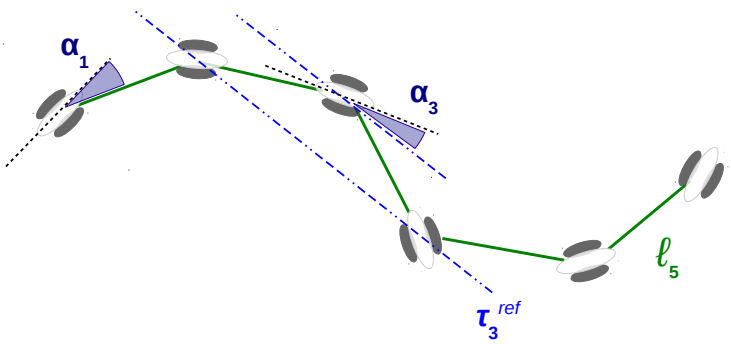


Figure 1.9: Illustration of the variables of interest in the detection algorithm.

the two sets of variables in a sum s

$$s = \sum_{i=1}^{n} \frac{2\alpha_i}{\pi} + \sum_{i=1}^{n-1} \tilde{F}_\ell(\ell_i). \tag{1.2}$$

This yields the following definition for the NFA of a chain:

$$\mathrm{NFA}(c) \stackrel{\text{def}}{=} N_T \cdot \mathbb{P}(S_{2n-1} \leq s), \tag{1.3}$$

where N_T is the number of tested chains and the variable S_{2n-1} follows the Irwin-Hall distribution of order $2n-1$.

As we will see, despite our approximations, simulations tend to show that the algorithm still complies with the non-accidentalness property, in the sense that the number of detections in noise is still upper-bounded.

More importantly, the comparison between humans and our artificial observer yields a very accurate match. Figure 1.10 sums up this comparison, following the same pattern as Figure 1.7. The data of the subjects and the algorithm is presented, averaged over all experimental sessions. The first two plots show the detection rate as a function of the jitter level and the number of target elements respectively, whereas the last two columns display respectively the detection rate and the reaction time as functions of $\log_{10}$(NFA) of the target path. These plots will be supported by further analysis, showing that the algorithm's performance is statistically similar to a human subject on this task. In addition, the information shown in Figure 1.11 is worth commenting. Indeed, beyond the correlation between detectability and NFA, it draws a perceptual equivalence among stimuli characterized by the same NFA.

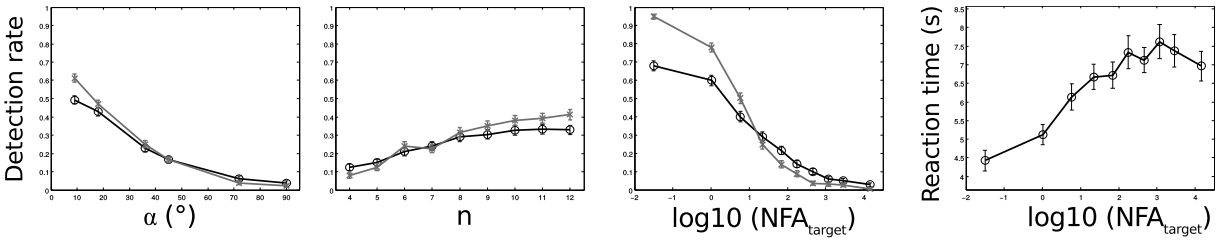


Figure 1.10: Comparison of the subjects of Chapter 4's experiments to the *a contrario* detection algorithm for curves. The solid black lines represent the subjects, whereas the gray solid lines represent the algorithm's results. The data shown here is the average over all subjects. The detection rates are plotted as functions of the jitter intensity α (first column), the number n of aligned elements (second column) and the $\log_{10}$ of the target's NFA (third column). The latter is also the x variable in the last column, the y axis representing the reaction time.

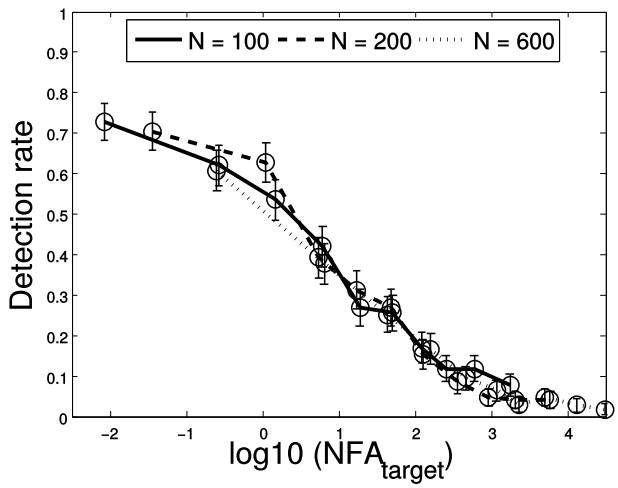


Figure 1.11: Detection rates achieved by the subjects, as functions of $\log_{10}$(NFA) of the target paths, for the three different values of N. The superimposition of the three curves indicates a perceptual equivalence among stimuli characterized by the same NFA: for a given NFA value, the detection rates are substantially the same, regardless of the value of N.

1.4 Summary of the contributions

The proposal of this thesis is driven by a main methodological concern: comparing systematically visual perception to an algorithm. Two benefits may be derived from this approach. First, it is a way to test the reliability and accuracy of computer vision methods, when these aim at reaching the performance of the visual system. Second, it is the ultimate proof for any model of vision. In the dissertation, we focus on the *a contrario* theory, which has always claimed to provide Gestalt theory with quantitative tools based on the non-accidentalness principle.

Our first contribution is to explore several possible implementations of the methodology. In Chapter 2, we exploit the interactivity offered by web interfaces in three original set-ups. One, that we called *gestaltic game*, invites the user to build stimuli and compare their own interpretation to the output of an *a contrario* algorithm. We also propose an informal short psychophysical experiment which can be taken on the web. This allows to collect easily a significant amount of data, to be averaged and compared to a second algorithm. In the same vein, the third set up is an online game in which the perception of the player is pushed to its limits. This exploration stands as a proof of concept. Indeed, the comparison of the visual perception of alignments to *a contrario* algorithms, yields promising preliminary results.

In Chapter 3, we deepen the psychophysical aspect of our program. We build a more rigorous experimental set up, consisting in three versions of a straight contours detection task, and we design an *a contrario* algorithm able to detect the same straight contours. By comparing the data of human subjects to the algorithm's detections, we further investigate the role of non-accidentalness in this perceptual task. We show that the humans' detection performance is strongly correlated to the NFA of the target alignments, and that the algorithm matches quite well this performance in a large variety of stimuli.

This approach is generalized in Chapter 4, where we focus on the perception of curvilinear contours. Like in the previous chapters, the first experiment on the detection of curves is run on the subjects' computer devices, through a web interface. We validate this protocol by showing that it gives similar results to a second experiment, carried out in one single controlled environment. Again, we design an *a contrario* algorithm specially tailored to perform the same detection task as human subjects. What we observe through the comparison of the perceptual data to our model, is in accordance with the previous results on the perception of alignments. Not only does the NFA correlate with the detection performance of human subjects, but we actually find an equivalence between the detectability of the target stimuli and their NFA.

The work in this thesis has led to the following publications:

- Lezama, J., Blusseau, S., Morel, J.M., Randall, G., and Grompone von Gioi, R. *Psychophysics, Gestalts and games*. In Sarti, A. and Citti, G., editors, Neuromathematics of Vision. Springer, 2014.

- Blusseau, S., Carboni, A., Maiche, A., Morel, J.-M. and Grompone von Gioi, R. *A psychophysical evaluation of the a-contrario detection theory*. In IEEE International Conference on Image Processing (ICIP) 2014.

- Blusseau, S., Carboni, A., Maiche, A., Morel, J.-M. and Grompone von Gioi, R. *Measuring the salience of alignments by their non-accidentalness*. Vision Research, accepted for publication.

- Blusseau, S. Morel, J.-M., and Grompone von Gioi, R. *Measuring the salience of curves by their non-accidentalness*. To be submitted.

- Submissions to IPOL journal are in preparation. Supplementary material concerning Chapters 3 and 4 is already accessible at `http://bit.ly/na_alignments` and `http://bit.ly/ac_curves`.

2 Psychophysics, Gestalts and Games

Many psychophysical studies are dedicated to the evaluation of the human gestalt detection on dot or Gabor patterns, and to model its dependence on the pattern and background parameters. Nevertheless, even for these constrained percepts, Psychophysics have not yet reached the challenging prediction stage, where human detection would be quantitatively predicted by a (generic) model. On the other hand, Computer Vision has attempted at defining automatic detection thresholds. This chapter sketches a procedure to confront these two methodologies inspired in gestaltism.

Using a computational quantitative version of the non-accidentalness principle, we raise the possibility that the psychophysical and the (older) gestaltist setups, both applicable on dot or Gabor patterns, find a useful complement in a Turing test. In our perceptual Turing test, human performance is compared by the scientist to the detection result given by a computer. This confrontation permits to revive the abandoned method of gestaltic games. We sketch the elaboration of such a game, where the subjects of the experiment are confronted to an alignment detection algorithm, and are invited to draw examples that will fool it. We show that the method leads to a more precise definition of the alignment gestalt and of its computational formulation.

Detection algorithms might also be relevant to more classic psychophysical setups, where they can again play the role of a Turing test. To a visual experiment where subjects were invited to detect alignments in Gabor patterns, we associated a single function measuring the alignment detectability in the form of a number of false alarms (NFA). The first results indicate that the values of the NFA, as a function of all simulation parameters, are highly correlated to the human detection. This fact, that we intend to support by further experiments in the next chapter, might end up confirming the relevance of non-accidentalness as a mechanism in alignment detection by humans.

2.1 Introduction

Alan Turing advanced a controversial proposal in 1950 that is now known as the *Turing Test* [81]. Turing's aim was to discuss the problem of machine intelligence and, instead of giving a premature definition of thinking, he framed the problem in what he called the *Imitation Game*: A human interrogator interacts with another human and a machine, but only in typewritten form; the task of the interrogator is to ask questions in order to determine which of its two interlocutors is the human. Turing proposed that a machine that eventually could not be distinguished from humans by its answers should be considered intelligent. This influential suggestion sparked a fruitful debate that continues to this day [69].

Our concern here is however slightly different. We are studying perception and Turing pre-

cluded in his test any machine interaction with the environment other than the communication through the teletype; he concentrated on the pure problem of thinking and to that aim avoided fancy computer interactions, that anyway did not exist at his time. Yet, machine perception is still a hard problem for which current solutions are far from the capacities of humans or animals[1]. Our purpose in this chapter, is to discuss a variety of *perceptual imitation games* as a research methodology to develop machine vision algorithms on the one hand, and quantitative psychophysical protocols on the other.

Human perceptual behavior has been the subject of quantitative experimentation since the times of Fechner, the founder of Psychophysics. This relatively new science investigates the relationship between the stimulus intensity and the perceived sensation [78]. But this approach does not provide a perceptual theory on which machine vision and an imitation game could be based.

The Gestalt school, Wertheimer, Köhler, Koffka, Kanizsa among others [90, 41, 12, 58, 37], developed from the twenties to the eighties an original *modus operandi*, based on the invention and display to subjects of clever geometric figures [88, 89]. A considerable mass of experimental evidence was gathered, leading to the conclusion that the first steps of visual perception are based on a reduced set of geometrical grouping laws. Unfortunately these Gestalt laws, relevant though they were, remained mainly qualitative and led to no direct machine perception approach.

Since the emergence of the field of Computer Vision [57] about fifty years ago – initially as a branch of the Artificial Intelligence working with robots and its artificial senses – there have been many attempts at formalizing vision theories and especially Gestalt theory [72]. Among them one finds models of neural mechanism [30], theories based on logical inference [15], on information theory [43], invoking minimum description principles [99], or grammars of visual elements [98, 31]. Nevertheless, only a small fraction of these proposals has been accompanied by systematic efforts to compare machine and human vision. An important exception is the Bayesian theory of perception [60] that has attracted considerable attention in cognitive sciences, leading to several experimental evaluations [97, 16, 56, 40, 13, 91]. A recent groundbreaking work by Fleuret et al. [22] compared human and machine performing visual categorization tasks. Humans are matched against learning algorithms in the task of distinguishing two classes of synthetic patterns. One class for example may contain four parallel identical shapes in arbitrary position, while the other class contains the same shapes but with arbitrary orientation and position. It was observed that humans learn the distinction of such classes with very few examples, while learning algorithms require considerably more examples, and nevertheless gain a much lower classification performance. The experimental design was more directed at pointing out a flaw of learning theory, though, than at contributing to Psychophysics.

Such experiments stress the relevance of computer vision as a research program in vision, in addition to a purely technological pursuit. Its role should be complementary to explanatory sciences of natural vision by providing, not only descriptive laws, but actual implementations of mechanisms of operation. With that aim, perceptual versions of the imitation game should be the Leitmotiv in the field, guiding the conception, evaluation and success of theories.

Here we will present comparisons of human perception to algorithms based on the *non-accidentalness principle* introduced by Witkin, Tenenbaum and Lowe [48, 96, 95] as a general grouping law. As we will explain in Section 2.2, this principle states that spatial relations are perceptually relevant only when their accidental occurrence is unlikely. We shall use the *a contrario* framework, a par-

[1]It is a common practice in Internet services to use the so-called CAPTCHAs to ensure that the interaction is made with a human and not an automatic program. A CAPTCHA, which stands for *Completely Automated Public Turing test to tell Computers and Humans Apart*, usually consists in a perceptual task, simple to perform for humans but hard for known algorithms. This suggests that visual and auditory perception currently provides the most effective Turing test.

ticular formalization of the principle due to Desolneux, Moisan and Morel [10, 11] as part of an attempt to provide a mathematical foundation to Gestalt Theory.

This chapter is intended to illustrate the main ideas of our research program, as well as its first results that led to the deeper studies presented in the next chapters. Here the settings were reduced to the bare minimum, concentrating in one simple geometric structure, namely alignments. The methodology however is general. By using such a simple structure we present different possible versions of the imitation game approach: a research procedure inspired in the methodology of the Gestalt school and the use of online games for psychophysical experimentation.

A review of the non-accidentalness principle in vision is given is Section 2.2, followed by an introduction to the *a contrario* methods in Section 2.3.

Gestaltism created clever figures in which humans fail to perceive the expected structures, generating illusions. In the *gestaltic game*, as we shall call our first proposed methodology, the experimenter tries to fool the algorithm by building a particular data set that produces unnatural results. This methodology is discussed in Sect. 2.4.

The last part, in Sect. 2.5, is dedicated to a first attempt at a psychophysical evaluation of the same theory. There is a difference with classic psychophysical experiments in which detection thresholds are measured; here each stimulus will be shown to human subjects but also to an algorithm, and both will answer yes or no to the visibility of a given structure. In a second variation, both humans and machine will also have to point to the position of the observed structure. This last variation is proposed as an online game, used as a methodology to facilitate experimentation and the attraction of volunteers. The experiments presented in the next chapters were built on these ideas.

Being the result of a very early work, no final conclusion will be drawn. Our overall goal here is to advocate for new methods in the quantitative study of visual perception, before developing them in further parts of the dissertation.

2.2 The non-accidentalness principle in vision

Our visual system is stimulated by photons coming directly from light-sources or indirectly after the reflection on objects' surfaces. The efficiency of our perception lies in its ability to group those independent stimuli into sets which actually correspond to real objects. For example, the set of photons reflected by the surface of a wooden table is perceived as one single object, that we easily distinguish from another set of stimuli, such as those coming from a wooden chair close to the table. Moreover, the Gestaltists proved that the same process can occur even when we do not recognize any familiar object in the observed scene [90, 37]. Quoting Witkin and Tenenbaum, *"we organize the data even when we have no idea what it is we are organizing."*[95, p. 482]. To illustrate this phenomenon, they give the example of a picture, reproduced in Figure 2.1, representing the surface of the planet Jupiter. They point out that *"Although difficult to capture in words, the entities we see might be variously described as textured bands, channels, swirling patterns, puffs, spots, and so forth. What we see is* structure - *orderly, regular, or coherent patterns and relationships."*[95, p. 495]. What properties characterize those elements that our visual perception groups into sets called *structures*? The Gestalt school observed that they are bound by some grouping laws, such as proximity, similarity, convexity, smoothness, symmetry [90, 37]. Witkin and Tenenbaum conjecture that, more generally, we perceive as a structure a set of stimuli that have been produced by a *common cause*. According to this hypothesis, the visual system has the ability to detect the spatiotemporal regularity that is *"unlikely to arise by the chance interaction of independent entities"*[95, p. 481]. Looking at Figure 2.1, it is hard to see one of the dark smooth curves crossing

the image, as a set of independent dark points. Even if we cannot explain such a structure, we can't help seeing it as a whole, and guessing a unique cause behind it. Similarly, the parallelism between several of these curves literally pops out and binds them together perceptually (*"the spurious appearance of parallelism is extremely* unlikely *to arise among causally unrelated curves"*[95, p. 505]).

Figure 2.1: A picture of the planet Jupiter, in which a non expert observer usually does not recognize any familiar object but does perceive structures. Figure taken from [95].

The *non-accidentalness principle* is this idea that perception relies on non-casual relationships to segment an observed scene into meaningful structures, no matter if their meaning is accessible to the observer, or not. Although it was particularly well addressed in [95], previous or simultaneous formulations of this principle exist. In the early 80s, several computer vision scientists dealt with the issue of recovering three dimensional scenes from their two dimensional pictures [76, 77, 35, 3, 53, 92]. To do so, they usually focused on what would be called later *non-accidental properties* [49, 50, 86, 87], that is to say relations that exist between some elements in an image, and that must have been created by similar relations in the physical world or, in other words, that are unlikely to have been created by independent physical phenomena.

At first, Stevens called these properties *constraints* [76]. Indeed, the assumption that such relations are very likely to hold in three-dimensional scenes when observed in two-dimensional representations, constrains the ill-posed problem of inferring 3D-scenes from images, and helps the visual system solve it. Among these image properties are collinearity, parallelism, proximity, smoothness, straightness, which are configurations that are unlikely to be caused only by the viewpoint peculiarity and without any meaningful 3-D arrangement. In [77], Stevens compares this principle to the *general position* assumption, which precisely states that, in general, pictures are not taken from a viewpoint that produce accidental illusions ("general position *[...] means the viewpoint is not misleading–that the image is taken from a representative viewpoint."*[77, p. 55]; *"Note that one may find special situations that violate one or the other form of general position—either non-parallel curves on some non-cylindrical surface which look parallel from a particular viewpoint, or parallel curves on some non-cylindrical surface which are critically placed on the*

surface–but these are arguably improbable occurrences in nature."[77, p. 59]). This is quite close to the concept of *structural stability* mentioned and illustrated in [75].

In [35], a method is proposed to recover 3D-shapes from pictures by *"mapping image regularities into shape constraints"*[35, p. 422]. More precisely, Kanade relies on *"nonaccidental regularities"*, which comply with the following heuristic: *"Regularities observable in the picture are not by accident, but are some projection of real regularities."* [35, p. 422]. Among those regularities, [35] focuses particularly on parallelism of lines and skewed symmetry. The exposed method *"quantitatively recovers the probable shapes by exploiting the assumptions which exclude accidental alignments and regularities in the picture."* [35, p. 410].

Another example is the set of assumptions Binford makes to infer surfaces from images [3]. The *Coincidence assumption* is one of them, and Binford phrases as follows: (*"In absence of other evidence, assume that if features coincide in an image, their inverse images coincide in space if sufficiently constrained."*[3, p. 215]). Even more illustrative is the example he draws on the end points of two random curves in space: if one curve appears to terminate exactly on one point of the other curve in the image of the scene, then it is unlikely, under the coincidence assumption, that this relationship does not hold in space. This concept was then rephrased in [53], as *coincidence assumptions "that various coincidental features or alignments in an image are unlikely to arise without reason."*[53, p. 613].

Similar to the latter idea is the *rejection-of-coincidence* principle [71] enunciated by Rock (*"Certain well-known phenomena can be interpreted in terms of the rejection-of-coincidence principle. Consider the example of the launching effect described by Michotte: one visual object, A, moves until it reaches a second stationary object, B, whereupon A stops and B moves away in the appropriate direction and speed. The impression that A has caused B to move is irresistible. Not to perceive causation here is to accept the relationship between the behavior of the two objects as purely coincidental; the spatial and temporal contiguity, the shared direction of motion, and the equivalent velocity."*[71, p. 137]).

Not only has non-accidentalness inspired computer vision since the 1980s, but it has also led to actual breakthroughs in this area. The concept of scale-space, introduced by Witkin [94] and that have become a fundamental tool in signal and image processing since then, is based on the idea that meaningful patterns in a signal are stable through a large range of scales. Conversely, the consistency of features across many scales is unlikely to occur without a relevant structure in the signal to explain them (*"natural scales occur where the description (i.e., locations of discontinuities) is stable over a range of scales. Such stability can, in fact, be viewed as just another instance of a structural relation (i.e., replication over scale) that would be unlikely to arise by accident."*[95, p. 518]). Later on, several remarkable contributions from Lowe were rooted in the non-accidentalness principle [49, 50], and his essential work on scale-invariant features in images [51, 52] is partly due to a deep understanding of scale-space filtering.

Whereas these works in computer vision relied on the intuition of the perceptual relevance of non-accidentalness, psychologists and psycophysicists wanted to provide it with a firm experimental basis. In [86, 87], the wish to validate this principle as a perceptual mechanism is explicitly formulated. The latter studies were focused on axial symmetry. The tested hypothesis was that, if symmetry was *used* by subjects as a non-accidental property, it should help them recover symmetrical shapes from transformed versions of them. Non-accidentalness was thus intended here as a cue to infer real world properties from the observation of a planar representation. It was shown that *"it was always easier to match a pattern with affine transformed versions of it, when the pattern contained symmetry"*[86, p. 244]. According to the author, these results only indicated that symmetry was *detected* by subjects and helped perform the task, but were not enough to validate the

initial hypothesis. What was missing was a proof that subjects *interpreted* a skewed-symmetrical pattern as a mirror-symmetrical one *"oriented in depth"*. But the author observed no evidence that subjects inferred such a slant of the shape's plan. Instead, they seemed to rely on local patterns, such as alignments or clusters (present in some random patterns but favoured or enhanced by symmetry), to match shapes by inferring a transformation *within the plane*. However, as the author points out eventually, by these "tricks" the subjects were actually using the non-accidental repetition of patterns, in the sense that some patterns were too similar to arise independently in two different random shapes.

More recently, the contribution of [18] went actually beyond the experimental measure of the particular status of non-accidental properties in perception. It drew a parallel with their special status in the *minimal model* theory [15, 17]. The latter theory was then featured a Bayesian structure [19], binding together non-accidentalness, the likelihood principle and the simplicity principle.

The non-accidentalness principle can be interpreted as imposing minimal reliability conditions to percepts. A configuration that could have arisen by chance does not provide reliable information and should be rejected. Natural selection evolved reliable perception mechanisms which may obey a similar sound requirement, that no detection should occur in noise [2]. Undoubtedly, this minimal reliability condition is not the only factor involved in the salience of a particular structure. Nevertheless, it is interesting to evaluate whether this factor alone can predict the visibility of a specific Gestalt to some extent. Our purpose is precisely to study to what extent the *a contrario* formulation of non-accidentalness can account for visual salience.

2.3 Introduction to the *a contrario* theory

We propose to introduce the intuitive ideas and formal concepts behind the *a contrario* approach. We shall restrict the description to the minimal content required for this dissertation, but we refer to [11, 29] for a more general presentation.

2.3.1 Ideas and intuitions

Desolneux *et al.* [11] designed the *a contrario* theory as a mathematical formulation of non-accidentalness, although the authors called *Helmholtz principle* what is more commonly called non-accidentalness. The expression *a contrario* (Latin for "by or from contraries") refers to the fact that the theory focuses the *probabilistic* modeling on *noise*, understood as the absence of signal. The model of the signal itself is a deterministic description of an ideal structure. When an *a contrario* detection algorithm analyzes a candidate feature, it evaluates the expected number of features with lower or equal error *relative to the ideal structure*, that could happen by chance in noise. In a way, the *a contrario* theory may remind of an aspect of the signal detection theory, as *"a computational framework that describes how to extract a signal from noise"* [25, p. 1]. In signal detection theory, a detection in absence of signal is called a *false alarm*. Similarly, the expected number of accidental detections in noise, associated to a given feature by an *a contrario* algorithm, is called *Number of False Alarms* (NFA). The lower this number, the less likely it is that the detection of this feature be a false alarm. Thus, if the NFA is low enough, the feature is termed non-accidental and therefore significant.

As a probabilistic formalization rooted in information theory, the *a contrario* framework may be compared to Bayesian approaches (see for example [14, 16, 20, 91, 13], for illustrations of the Bayesian theory applied to the good continuation problem). A Bayesian formulation uses prior probabilities and likelihood terms. This requires hypothesis about the distributions and parameters estimations. In some problems, these are natural and lead to optimal predictions; in others, the op-

timality is undermined by the absence of sufficient knowledge. The *a contrario* methodology does not require a *probabilistic* model of the sought Gestalt; it only requires a deterministic description of the ideal structure and a background (or *a contrario*) hypothesis H_0, which assumes the maximal entropy for the distribution of the data, and models the absence of the relevant pattern. The Bayesian approach is preferable when all the distributions and their parameters are well known, as it is able to take full advantage of this information. In the lack of such knowledge, the *a contrario* methodology offers a good alternative, as it requires fewer assumptions.

A whole family of unsupervised algorithms in computer vision are based on the *a contrario* theory. The framework has been used to detect line segment and curves [11, 28, 68] as well as to perform shape matching and identification [62, 5], image segmentation [6], clustering [79], and mirror symmetry detection [67]. These algorithms apply the *a contrario* theory to adjust the detection thresholds so that the *expected* number of accidental detections is provably bounded by a small constant ε.

In order to identify *accidental* detections, an *a contrario* method requires two models: a deterministic model and a probabilistic one.

In the scope of the present work, as well as in many applications of the *a contrario* methods, the deterministic model is a geometric description of an ideal structure to be detected, that is why we will also call it *geometric model*. It is characterized by a target value x^* of a given measurement x, that is made on any structure candidate to detection. For example, imagine we are interested in detecting, among a set of points, triplets forming equilateral triangles. Then, for each triplet $\{a, b, c\}$, a possible measurement could be the difference of the greatest angle and 60°, that is $x = \max\{\hat{a}, \hat{b}, \hat{c}\} - 60°$. In this case, the target value would be $x^* = 0$, corresponding to a perfect equilateral triangle. Obviously, in real data, like for example a set of points extracted from a natural image, this target value might never be observed. But what we are actually looking for are triplets for which the observed value x is *suspiciously* close to x^*. On the contrary, what we don't want to detect are configurations that could be observed by chance in a random set of points. Therefore, to decide when the observed measurement is close enough to the target value to yield a detection, we need the probabilistic model.

The probabilistic model, also called *a contrario* or *background* model, is the statistical hypothesis H_0 of the absence of relevant structure - the so called *"null hypothesis"*. It represents the most general assumption on the data. Consistently with Attneave's principle stating that we do not perceive any structure in white noise [2], an *a contrario* model generally assumes the maximal entropy distribution for the building blocks of the percept. For example, if these building blocks are oriented, their orientations must be random, independent, and uniformly distributed over all possible angles; if they are non-oriented points, they are generally assumed to be uniformly and independently distributed over the region of interest. By Attneave's principle the emergence of a percept in a realization of such a model should be unlikely, and in any case purely accidental. Therefore this background model is used to test the significance of an observation, as follows. Let X be a random variable distributed according to H_0, and suppose that we have a function $d_{x^*}(\cdot)$ measuring the deviation of an observation from the target value. Given an observed feature x with a deviation $d_{x^*}(x)$, its relevance is measured by the probability $\mathbb{P}_{H_0}(d_{x^*}(X) \leq d_{x^*}(x))$ to be at least as close to x^* under H_0. Small deviations yield small probabilities, and are thus rare events in the background model.

The purpose of the *a contrario* framework is precisely to help decide when an observed deviation is small enough to be considered non-accidental. The occurrence of an event does not only depend on its probability. It also depends on the number of observations. For example, at the roulette game, the odds of a given number n between 0 and 36 is $1/37 \approx 0.027$. With a single bet

on that given n, it would be quite lucky to win. On the other hand, if a player keeps betting 1000 times on that same number, it would be surprising not to win at least once. This intuition finds a formalization in the Bonferroni correction method to test statistical significance of p-values, as it is well explained in [61, pp. 114-118].

The *a contrario* methodology uses the same kind of correction term: to evaluate the non-accidentalness of an event, one should take into account the whole set of observations that were necessary to come across that particular event. For example, when looking for smooth paths in an array of Gabor patches, we define *a priori* a family of sets of patches, that will be tested as candidates to be perceptible curves. This is what is usually called the *family of tests*, noted T, and N_T denotes the number of tests in T. Then, given a tested candidate x that yielded the deviation $d_{x^*}(x)$, denote by $X_1, X_2, \dots, X_{N_T}$, the random issues to the N_T tests, under the background model H_0. The question is now how common or, on the contrary, how surprising it would be to observe a deviation smaller than $d_{x^*}(x)$ among these N_T tests. An answer is to compute the expected number of variables X_i that verify $d_{x^*}(X_i) \leq d_{x^*}(x)$. This quantity is called *Number of False Alarms* associated to x, noted NFA(x). Indeed, since we are looking for a non-accidental small deviation, accidental ones are considered as "false alarms".

By definition, if NFA(x) ≥ 1, we expect to observe at least one deviation smaller than, or equal to $d_{x^*}(x)$, in random data consistent with H_0. In other words, it would not be surprising that such an event happen by accident, and the observation of x should not lead to reject H_0. Inversely, NFA(x) < 1 means that a deviation smaller than $d_{x^*}(x)$ is rather unexpected under H_0, and may indicate the presence of a relevant pattern. Thus, small NFAs characterize non-accidentalness and large ones accidental features, with a transition zone around NFA(x) $= 1$.

2.3.2 Formal results

Let's now present formally the framework and the fundamental results of the *a contrario* theory, to which we will refer in the remaining of the dissertation. The developments here are very similar to existing ones in the *a contrario* literature[11, 26, 29]. Their presentation is however adapted to the content of the thesis.

In this section, for simplicity and because it is consistent with most of the models defined later on, we only consider positive measurements and assume 0 is the target value. Better said, although the following developments hold in general, they only make sense as a detection framework if the meaningful values are those which are suspiciously close to 0. Note that, in many cases, detection problems can be framed that way.

Identically distributed measurements Let X a positive random variable, N_T an integer and $X_1, X_2, \dots, X_{N_T}$ random variables distributed like X (but not necessarily independent). We want to characterize how surprising it is to observe a value as low as a number x among N_T observed measurements $x_1, \dots, x_{N_T}$, assuming that they are sampled from $X_1, \dots, X_{N_T}$. Here we do it by the following definition of the NFA.

Definition 1. *The* NFA *associated to an observation $x \in \mathbb{R}^+$ is the expected number of observations with values lower than x, among the random measurements $X_1, \dots, X_{N_T}$:*

$$\text{NFA} : \begin{cases} \mathbb{R}^+ & \rightarrow & \mathbb{R}^+ \\ x & \mapsto & \mathbb{E}\left[\sum_{i=1}^{N_T} \mathbb{1}_{[0,x]}(X_i)\right] \end{cases} \tag{2.1}$$

It is important to remark that this NFA can be written in a simpler form. Indeed, for $x \in \mathbb{R}^+$,

$$\begin{aligned} \mathbb{E}\left[\sum_{i=1}^{N_T} \mathbb{1}_{[0,x]}(X_i)\right] &= \sum_{i=1}^{N_T} \mathbb{E}[\mathbb{1}_{[0,x]}(X_i)] \\ &= \sum_{i=1}^{N_T} \mathbb{P}[X_i \leq x]. \end{aligned}$$

Since each X_i is distributed like X, we get the expression

$$\mathrm{NFA}(x) = N_T \cdot \mathbb{P}[X \leq x]. \tag{2.2}$$

As we will see, the NFAs defined in the thesis will be written in a similar fashion. Let's now get to the fundamental result linking NFA and non-accidentalness.

Theorem 1 (Non-accidentalness property of the NFA)**.** *Let $\varepsilon \in [0, N_T]$. Then the expected number of tests such that $\mathrm{NFA}(X_i) \leq \varepsilon$, is less than ε:*

$$\mathbb{E}\left[\sum_{i=1}^{N_T} \mathbb{1}_{\mathrm{NFA}(X_i)\leq\varepsilon}\right] \leq \varepsilon. \tag{2.3}$$

If, in addition, the law of X is absolutely continuous[2]*, this expected number is equal to ε.*

This result confirms theoretically the intuition formulated at the end of the previous section. Setting $\varepsilon = 1$, it tells us that an observation x with $\mathrm{NFA}(x) < 1$ is expected to happen less than once in data distributed like X. If this distribution follows some null hypothesis H_0 modeling the absence of relevant structure, then the observed x should lead to reject H_0 and should be interpreted as meaningful.

Proof. To prove Theorem 1, it is sufficient to show that $\mathbb{P}[\mathrm{NFA}(X) \leq \varepsilon] \leq \frac{\varepsilon}{N_T}$ for any $\varepsilon \in [0, N_T]$, and that we have an equality if the law of X is absolutely continuous. Indeed, we would then have

$$\begin{aligned} \mathbb{E}\left[\sum_{i=1}^{N_T} \mathbb{1}_{\mathrm{NFA}(X_i)\leq\varepsilon}\right] &= \sum_{i=1}^{N_T} \mathbb{P}[\mathrm{NFA}(X_i) \leq \varepsilon] \\ &= N_T \cdot \mathbb{P}[\mathrm{NFA}(X) \leq \varepsilon] \\ &\leq N_T \cdot \frac{\varepsilon}{N_T} = \varepsilon. \end{aligned} \tag{2.4}$$

the inequality becoming an equality if $\mathbb{P}[\mathrm{NFA}(X) \leq \varepsilon] = \frac{\varepsilon}{N_T}$. This preliminary result is trivial for $\varepsilon = N_T$, and we now prove it for $\varepsilon \in [0, N_T)$.

Let F_X the cumulative distribution function of X. For any $y \in [0, 1)$ we note

$$\mathcal{A}_X(y) \stackrel{\text{def}}{=} \{x \in \mathbb{R}, F_X(x) \leq y\} \tag{2.5}$$

the set of values x such that the probability $\mathbb{P}[X \leq x]$ is not greater than y. This set is not empty since it contains any $x < 0$, and it is upper bounded because F_X is an increasing function and $\lim_{x\to+\infty} F_X(x) = 1$. Thus we can define $\hat{x}(y) \stackrel{\text{def}}{=} \sup \mathcal{A}_X(y)$, and it is a real number.

[2]Note that any variable X such that $\mathbb{P}[X = x] \neq 0$ for at least one value x, is not absolutely continuous with respect to Lebesgue measure. As an example, we can consider a discrete variable, such as the number of "tails" after tossing a fair coin n times: $\mathbb{P}[X = x] = \binom{n}{x}\frac{1}{2^n}$.

Now, let $\varepsilon \in [0, N_T)$. Two cases arise: $\hat{x}(\frac{\varepsilon}{N_T}) \in \mathcal{A}_X(\frac{\varepsilon}{N_T})$, and $\hat{x}(\frac{\varepsilon}{N_T}) \notin \mathcal{A}_X(\frac{\varepsilon}{N_T})$.
Case 1: $\hat{x}(\frac{\varepsilon}{N_T}) \in \mathcal{A}_X(\frac{\varepsilon}{N_T})$.
Then $F_X(\hat{x}(\frac{\varepsilon}{N_T})) = \frac{\varepsilon}{N_T}$, and for any $x \in \mathbb{R}$, $\mathrm{NFA}(x) \leq \varepsilon \Leftrightarrow x \leq \hat{x}(\frac{\varepsilon}{N_T})$. Thus,

$$\begin{aligned}
\mathbb{P}[\mathrm{NFA}(X) \leq \varepsilon] &= \mathbb{P}[X \leq \hat{x}(\tfrac{\varepsilon}{N_T})] \\
&= F_X(\hat{x}(\tfrac{\varepsilon}{N_T})) \\
&= \tfrac{\varepsilon}{N_T}.
\end{aligned}$$

Case 2: $\hat{x}(\frac{\varepsilon}{N_T}) \notin \mathcal{A}_X(\frac{\varepsilon}{N_T})$.
Then for any $x \in \mathbb{R}$, $\mathrm{NFA}(x) \leq \varepsilon \Leftrightarrow x < \hat{x}(\frac{\varepsilon}{N_T})$. Thus,

$$\begin{aligned}
\mathbb{P}[\mathrm{NFA}(X) \leq \varepsilon] &= \mathbb{P}[X < \hat{x}(\tfrac{\varepsilon}{N_T})] \\
&= \lim_{x \uparrow \hat{x}(\frac{\varepsilon}{N_T})} F_X(x) \\
&\leq \tfrac{\varepsilon}{N_T}
\end{aligned}$$

Therefore, in both cases we have $\mathbb{P}[\mathrm{NFA}(X) \leq \varepsilon] \leq \frac{\varepsilon}{N_T}$, with equality in Case 1. If the law of X is absolutely continuous, then F_X is continuous and Case 1 holds, which completes the proof. □

The assumption of identical distribution of the measurements in all tests, is convenient to introduce a first formal definition of the NFA. In particular, it shows how the NFA permits to dominate the average number of false detections. However, in practice, we won't build on this assumption in the different detection problems addressed in the dissertation.

Non-identically distributed measurements Now we assume that each test is characterized by two variables, a measurement x and what we will call a *size variable*, noted s. In addition, we assume that the significance of x depends on the value of s. For example, in Section 2.4.3, we deal with the detection of dots alignments. Each tested feature is a rectangle r, and we focus on the number $k(r)$ of dots which fell in that rectangle. Naturally, the surprise produced by this measurement depends on the area of r: the larger the rectangle, the more likely it is that random dots fell in it; thus two tested rectangles showing the same number of dots, represent very different events if the rectangles have very different areas.

Similarly here, although we are still interested in detecting suspiciously low values of x, we shall take into account that a measurement x_1, given the size value s_1, might be more suspicious than the event (x_2, s_2), even if x_2 is smaller than x_1.

Therefore, it does not make sense anymore to define the NFA as in Definition 1, since lower values than an observed x are not necessarily more significant. Recall that the purpose of the NFA is to impose an upper bound on the average number of false detections in a background distribution. In the previous paragraph, this property arose from the expression 2.2 of the NFA, in which the number of tests is multiplied by the odds of a more relevant value than the observed one. Similarly, we shall define here $\mathrm{NFA}(x, s) = N_T \cdot \mathbb{P}[X \leq x | S = s]$.

Let X and S two real valued random variables, with $X \geq 0$, and such that the function $F(x, s) = \mathbb{P}[X \leq x | S = s]$ is well defined for any $x \in \mathbb{R}$ and almost every $s \in \mathbb{R}$. Let N_T an integer and $(X_1, S_1), (X_2, S_2), \ldots, (X_{N_T}, S_{N_T})$ random couples distributed like (X, S). Despite this last

hypothesis, note that the variable $X_i|(S_i = s_i)$, which represents the measurement of interest for test i, is not distributed like $X_j|(S_j = s_j)$ as soon as $s_i \neq s_j$.

Definition 2. *The* NFA *associated to the measurement x given the size value s is*

$$\mathrm{NFA}(x, s) = N_T \cdot \mathbb{P}\Big[X \leq x | S = s\Big]. \tag{2.6}$$

We will see that in the thesis, all the defined NFAs will follow this definition, although we might sometimes omit the condition on the size variable. As expected, this NFA still shows the non-accidentalness property.

Theorem 2. *Let $\varepsilon \in [0, N_T]$. Then the expected number of tests such that* $\mathrm{NFA}(X_i, S_i) \leq \varepsilon$*, is less than ε:*

$$\mathbb{E}\left[\sum_{i=1}^{N_T} \mathbb{1}_{\mathrm{NFA}(X_i,S_i)\leq\varepsilon}\right] \leq \varepsilon. \tag{2.7}$$

If, in addition, for any $s \in \mathbb{R}$, $X|(S = s)$ is continuous, this expected number is equal to ε.

In particular, given (x, s), the expected number of tests (X_i, S_i) with NFA smaller than $\mathrm{NFA}(x, s)$, is less than $\mathrm{NFA}(x, s)$. In other words, $\mathrm{NFA}(x, s) = N_T \cdot \mathbb{P}[X \leq x|S = s]$ *dominates* the expected number of tests that are at least as good as (x, s), whereas $\mathrm{NFA}(x)$ in Definition 1 is *equal to* the expected number of tests at least as good as x. We shall now prove Theorem 2.

Proof. Naturally, the proof is analogue to the one of Theorem 1. With the same methods, we show that for any $s \in \mathbb{R}$ and any $\varepsilon \in [0, N_T]$,

$$\mathbb{P}[\mathrm{NFA}(X, s) \leq \varepsilon] \leq \frac{\varepsilon}{N_T} \tag{2.8}$$

with equality if $X|(S = s)$ is continuous. As a consequence, we can integrate over s and get

$$\begin{aligned} \mathbb{P}[\mathrm{NFA}(X, S) \leq \varepsilon] &= \int \mathbb{P}[\mathrm{NFA}(X, s) \leq \varepsilon]\, dP_S(s) \\ &\leq \int \tfrac{\varepsilon}{N_T} dP_S(s) \\ &= \tfrac{\varepsilon}{N_T}. \end{aligned} \tag{2.9}$$

With the same development as in Eq. 2.4, this suffices to complete the proof. □

We shall make one last remark on the interest of the NFA as defined here, with respect to Definition 1. As we said, for a given test, the variable at stake here was not the measurement x alone, but this value *given* the size s. The function $F(x, s) = \mathbb{P}[X \leq x|S = s]$, precisely, is a formulation of this criterion, and is therefore able to sort two observed couples (x_i, s_i) and (x_j, s_j) according to their relevance: (x_i, s_i) is more suspicious than (x_j, s_j) if and only if $F(x_i, s_i) \leq F(x_j, s_j)$. As a consequence, we could consider the variable $z_i = F(x_i, s_i)$ as measurement for each test. Then we would be in the case presented in the previous paragraph, where the measurements are identically distributed. We would define $\mathrm{NFA}(x, s) = N_T \cdot \mathbb{P}[F(X, S) \leq F(x, s)]$, which is equivalent to Definition 1 since all (X_i, S_i) are identically distributed, and Theorem 1 would hold. The reason for the alternative NFA we introduced in Definition 2, is that in practice the law of S, and thus the joint law of (X, S), are often unknown, making the the exact computation of $\mathbb{P}[F(X, S) \leq F(x, s)]$ impossible. We will see that in every detection algorithm we designed, we applied the framework induced by Definition 2.

This concludes our introduction to the *a contrario* theory. In the last two sections of this chapter, we explore several ways to compare the principles we just presented to visual perception.

2.4 Detection Theory versus Gestaltism

Here and in most of the text we shall call "gestalt" any geometric structure emerging perceptually against the background in an image. We stick to this technical term because it is somewhat untranslatable, meaning something between "form" and "structure". According to Gestalt theory, the gestalts emerge by a grouping process in which the properties of similarity (by color, shape, texture, etc.), proximity, good continuation, convexity, parallelism, alignment can individually or collaboratively stir up the grouping of the building elements sharing one or more properties.

2.4.1 Online Gestaltic Game

One of the procedures used by Gestalt psychology practitioners was to create clever geometric figures that would reveal a particular aspect of perception when used in controlled experiments with human subjects. They pointed out the grouping mechanisms, but also the striking fact that geometric structures objectively present in the figure are not necessarily part of the final gestalt interpretation. These figures are in fact counterexamples against simplistic perception mechanisms. Each one represents a challenge to a theory of vision that should be able to cope with all of them.

The methodology we propose in order to design and improve automatic geometric gestalt detectors is in a way similar to that of the gestaltist. One starts with a primitive method that works correctly in very simple examples. The task is then to produce data sets where *humans* clearly see a particular gestalt while the rudimentary method produces a different interpretation. Analyzing the errors of the first method gives hints to improve the procedure in order to create a second one that produces better results with the whole data set produced until that point. The same procedure is applied to the second method to produce a third one, and successive iterations refine the methods step by step. The methodology used by the Gestalt psychologist to study human perception is used here to push algorithms to be similar to their natural counterpart. Finding counterexamples is less and less trivial after some iterations and the counter-examples become, like gestaltic figures, more and more clever.

We decided to render this process interactive by drawing figures in a computer interface that delivers a detection result immediately. To facilitate the search we created an online interface where everyone can easily play the gestaltic game inventing new counterexamples[3]. The exploration of counterexamples is in that way transformed into an active search where previous examples are gradually modified in an attempt to fool the detection algorithm. The figures are all collected to be later used at the analysis stage. The *gestaltic game* is at the same time a method to produce interesting data sets, a methodology to develop new detection algorithms and a collaborative tool for research in the computational gestalt community. Each detection game will only stop when it eventually passes the Turing test, the algorithm's detection capability becoming indistinguishable from that of a human.

2.4.2 Dot Alignments Detection

For its simplicity, dot patterns are often used in the study of visual perception. Several psychophysical studies led by Uttal have investigated the effect of direction, quantity and spacing in dot alignment perception [82, 83]. The detection of collinear dots in noise was the target of a study attempting to assess quantitatively the masking effect of the background noise [80]. A recent work by Preiss analyzes various perceptual tasks on dot patterns from a psychophysical and computational perspective [70]. An interesting computational approach to detect gestalts in dot patterns

[3] `http://dev.ipol.im/~jirafa/ipol_demo/point_alignment_detection`

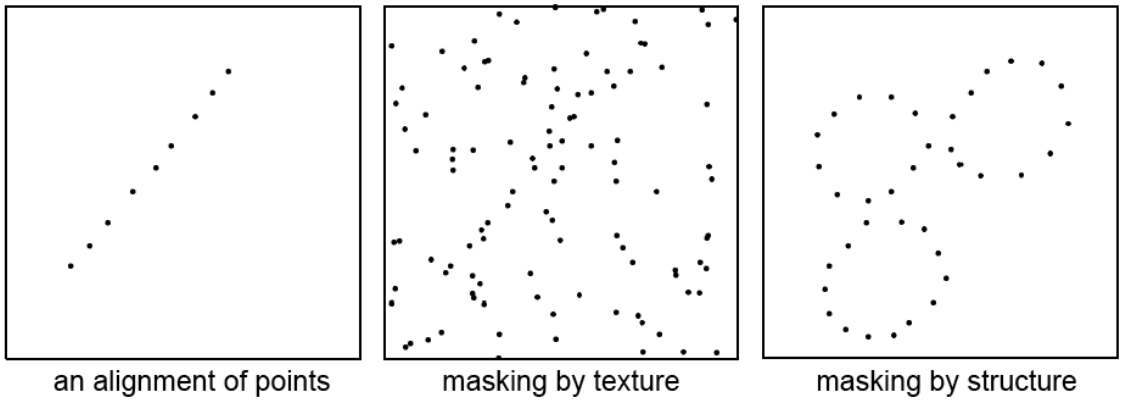


Figure 2.2: Exactly the same set of aligned dots is present in the three images, but it is only perceived as such in the first one. The second one is a classic masking by texture case and the third a masking by structure one, often called "Gestalt conflict".

is presented in [1], although the study is limited to very regularly sampled patterns. A practical application of alignment detection is presented in [85].

In a gestaltic sense, an alignment of points is a group of points sharing the property of being aligned in one direction. While it may seem a simple gestalt, Fig. 2.2 shows that a perceptual definition of an alignment is not straightforward. From a purely factual point of view, the same alignment is present in the three figures. However, it is only perceived as such by most viewers in the first one. The second and the third figures illustrate two occurrences of the *masking phenomenon* discovered by gestaltists [38]: the *masking by texture*, which occurs when a gestalt is surrounded by a clutter of randomly distributed similar objects or *distractors*, and the *masking by structure*, which happens when the alignment is masked by other perceptually more relevant gestalts, a phenomenon also called *perceptual conflict* by gestaltists [58, 59, 37]. The magic disappearance of the alignment in the second and third figures can be accounted for in two very different ways. As for the first one, we shall see that a probabilistic *a contrario* model [11] is relevant and can lead to a quantitative prediction. As for the second disappearance, it requires the intervention of the more general grouping law, namely *good continuation* [36].

These examples show that a mathematical definition of dot alignments is required before even starting to discuss how to detect them. A purely geometric-physical description is clearly not sufficient to account for the masking phenomenon. Indeed, an objective observer making use of a ruler would be able to state the existence of the very same alignment at the same precision on all three figures. But this statement would contradict our perception, as it would contradict any reasonable computational (definition and) theory of alignment detection.

The detection of an alignment is therefore a complex question, and must be decided by confronting mathematical definitions and detection algorithms to perception. As the patterns of Fig. 2.2 already suggest, simple computational definitions with increasing complexity will nevertheless find perceptual counterexamples. There is no better way to describe the ensuing "computational gestaltic game" than describing how the dialogue of more and more sophisticated alignment detection algorithms and counterexamples help build up a perception theory.

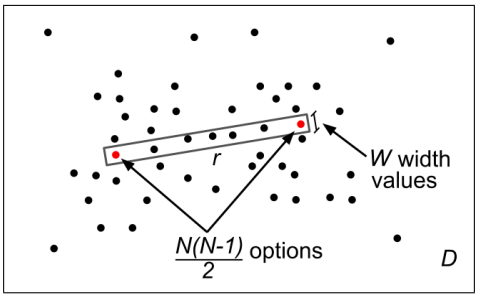


Figure 2.3: A schematic representation of the evaluated rectangle. In an image with N points, there are $(N(N-1) \times W)/2$ possible rectangles defined by two dots. In the case shown in this figure, $N = 49$ and the number of dots inside the rectangle is $k(r, \mathbf{x}) = 7$.

2.4.3 Basic Dot Alignment Detector

A very basic idea that could provide a quantitative context-dependent definition of dot alignments is to think of them as thin, rectangular shaped point clusters. In that case, the key measurements would be the relative dot densities inside and outside the rectangle. The algorithm described in this section is a simple illustration of the *a contrario* methodology introduced in Section 2.3.

Consider a dot pattern defined on a domain D with total area S_D and containing N dots, see Fig. 2.3. We are interested in detecting groups of dots that are well aligned. A first reasonable *a contrario* hypothesis H_0 for this problem is to suppose that the N dots are the result of a random process where points are independent and uniformly distributed in the domain. The question is then to evaluate whether the presence of aligned points contradicts the *a contrario* model or not.

Given an observed set of N points $\mathbf{x} = \{x_i\}_{i=1\ldots N}$ and a rectangle r (the candidate to contain an alignment), we will denote by $k(r, \mathbf{x})$ the number of those points observed inside r. We decide whether to keep this candidate or not based on two principles: a good candidate should be non-accidental, and any equivalent or better candidate should be kept as well. The degree of non-accidentalness of an observed rectangle r can be measured by how small the probability $\mathbb{P}\big[k(r, \mathbf{X}) \geq k(r, \mathbf{x})\big]$ is, where $\mathbf{X}$ denotes a random set of N dots following H_0. In the same vein, a rectangle r' will be considered at least as good as r given the observation $\mathbf{x}$, if $\mathbb{P}\big[k(r', \mathbf{X}) \geq k(r', \mathbf{x})\big] \leq \mathbb{P}\big[k(r, \mathbf{X}) \geq k(r, \mathbf{x})\big]$. Note that, in the latter probability terms, as well as in the next ones, we implicitly assume that the rectangle's area is known.

Recall that we want to limit the expected number of accidental detections. Given that N_{tests} candidates will be tested, Theorem 2 (Section 2.3) implies that the expected number of rectangles which are as good as r under H_0, is less than

$$N_{tests} \cdot \mathbb{P}\Big[k(r, \mathbf{X}) \geq k(r, \mathbf{x})\Big]. \tag{2.10}$$

The H_0 stochastic model fixes the probability law of the random number of points in the rectangle, $k(r, \mathbf{X})$, which only depends on the total number of dots N. The discrete nature of this law implies that (2.10) is not actually the expected value but an upper bound of it (see Section 2.3). Let us now analyze the two factors in (2.10).

Here the *a contrario* model H_0 assumes that the N points are i.i.d. with uniform density on the domain. Under the *a contrario* hypothesis H_0, the probability that one dot falls into the rectangle r is

$$p = \frac{S_r}{S_D}, \tag{2.11}$$

where S_r is the area of the rectangle and S_D the area of the domain. As a consequence of the independence of the random points, $k(r, \mathbf{X})$ follows a binomial distribution. Thus, the probability term $\mathbb{P}\left[k(r, \mathbf{X}) \geq k(r, \mathbf{x})\right]$ is given by

$$\mathbb{P}\Big[k(r, \mathbf{X}) \geq k(r, \mathbf{x})\Big] = \mathcal{B}\big(N, k(r, \mathbf{x}), p\big) \tag{2.12}$$

where $\mathcal{B}(n, k, p)$ is the tail of the binomial distribution

$$\mathcal{B}(n, k, p) = \sum_{j=k}^{n} \binom{n}{j} p^j (1-p)^{n-j}. \tag{2.13}$$

The *number of tests* N_{tests} corresponds to the total number of rectangles that could show an alignment, which in turn is related to the number of pairs of points defining such rectangles. With a set of N points this gives $\frac{N \times (N-1)}{2}$ different pairs of points. Note that, since each tested rectangle is defined by a pair of points, we should have $\mathbb{P}[k(r, \mathbf{X}) \geq 2] = 1$. Therefore, we shall replace N by $N - 2$ and $k(r, \mathbf{x})$ by $k(r, \mathbf{x}) - 2$ in Eq. 2.12.

The set of rectangle widths to be tested must be specified *a priori* as well. In the *a contrario* approach, a compromise must be found between the number of tests and the precision of the gestalts that are being sought for. The larger the number of tests, the lower the statistical relevance of detections. However, if the set of tests is chosen wisely, gestalts fitting accurately the tests will have a very low probability of occurrence under H_0 and will therefore be more significant.

Following [45], for a given pair of points, a geometric series of $W = 8$ rectangle widths $\{w_{\max}, q \cdot w_{\max}, q^2 \cdot w_{\max}, \ldots, q^7 \cdot w_{\max}\}$ is tested, decreasing from a maximal width $w_{\max}$ with a factor $q = 1/\sqrt{2}$. In turn, the maximal width $w_{\max}$ is equal to $1/10$ of the rectangle's length. This strategy makes the detection algorithm scale invariant. The total number of tested rectangles is then

$$N_{tests} = \frac{N(N-1) \times W}{2}. \tag{2.14}$$

We will define now the fundamental quantity of the *a contrario* framework, the Number of False Alarms (NFA) associated with a rectangle r and a set of dots $\mathbf{x}$:

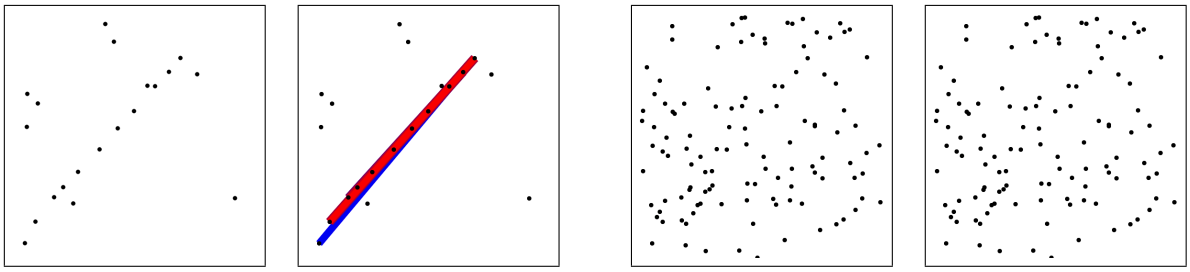
$$\mathrm{NFA}(r, \mathbf{x}) = N_{tests} \cdot \mathbb{P}\Big[k(r, \mathbf{X}) \geq k(r, \mathbf{x})\Big] = \frac{N(N-1) \times W}{2} \cdot \mathcal{B}\Big(N - 2, k(r, \mathbf{x}) - 2, p\Big). \tag{2.15}$$

As said before, this quantity dominates the expected number of rectangles which have a sufficient number of *points* to be as rare as r under H_0. When the NFA associated to a rectangle is small, this means that such an event is rare under the *a contrario* model, and is thus probably meaningful. On the other hand, when the NFA is large, the corresponding event might be expected. A *perceptual threshold* ε must nevertheless be chosen, and rectangles with $\mathrm{NFA}(r, \mathbf{x}) < \varepsilon$ will be called *ε-meaningful rectangles* [9], constituting the detection result of the algorithm.

Theorem 3 ([11],[45])**.**

$$\mathbb{E}\left[\sum_{R \in \mathcal{R}} \mathbb{1}_{\mathrm{NFA}(R, \mathbf{X}) < \varepsilon}\right] \leq \varepsilon$$

where $\mathbb{E}$ is the expectation operator, $\mathbb{1}$ is the indicator function, $\mathcal{R}$ is the set of rectangles considered, and $\mathbf{X}$ is a random set of points on H_0.



(a) (b) (c) (d)

Figure 2.4: Results from the basic dot alignment detector. **(a)** and **(c)** are the input data, and **(b)** and **(d)** are the corresponding results. Each detection is represented by a rectangle and its color indicates the NFA value. In **(b)** the algorithm correctly detects the obvious alignment. Notice that multiple and redundant rectangles were detected; this problem is addressed in [45]. The data set **(c)** contains the same set of points in **(a)** plus added noise dots, thus the aligned dots are still present. However, the algorithm handles correctly the masking by texture or noise and produces no detection.

The theorem states that the average number of ε-*meaningful rectangles* under the *a contrario* model H_0 is bounded by ε. Thus, the number of detections in noise is controlled by ε and it can be made as small as desired. In other words, this shows that this detector satisfies the non-accidentalness principle. The proof of Theorem 2 can be applied here to prove Theorem 3, with $N - k(r, \mathbf{x})$ as measurement conditionally to the rectangle's area (indeed $\mathbb{P}\big[k(r, \mathbf{X}) \geq k(r, \mathbf{x})\big] = \mathbb{P}\big[N - k(r, \mathbf{X}) \leq N - k(r, \mathbf{x})\big]$).

Following Desolneux, Moisan, and Morel [8, 11], we shall set $\varepsilon = 1$ once and for all. This corresponds to accepting on average at most one false detection per image in the *a contrario* model, which is generally reasonable. Also, the detection result is not sensitive to the value of ε, see [11].

Figure 2.4 shows the results of the basic algorithm in two simple cases. The results are as expected: the visible alignment in the first example is detected, while no detection is produced in the second. Actually, the dots in the first example are also present in the second one, but the addition of random dots masks the alignment, in accordance with human perception.

Naturally, this simple model for dot alignment detection has several limitations. First of all, it yields many redundant detections, as shown in of Figure 2.4(b). Additionally, it does not take into account many situations that can arise and significantly affect the perception of alignments. For example: what happens if there are point clusters inside the alignment? What if the background image has a non uniform density? Should not the algorithm prefer alignments where the points are equally spaced? These questions, among others, arise when subjects play the gestaltic game and try to fool the algorithm with new drawings. There are two ways to fool the algorithm: one is by drawing a particular context that prevents the algorithms from detecting a conspicuous alignment. Inversely, the other sort of counterexample is a drawing inducing detections that remain invisible to the human eye. As more counterexamples are found, more sophisticated versions of the algorithm must be developed, and each new version will become harder to falsify than the previous one.

Using this methodology, Lezama *et al.* produced several refined versions of the basic algorithm [45]. The last version, that can be tested online (see Section 2.4.1), is therefore harder to fool. Ideally, the game should end when the *Turing test* [69] is satisfied, namely when a human observer will be unable to distinguish between the detections produced by a machine and by a human.

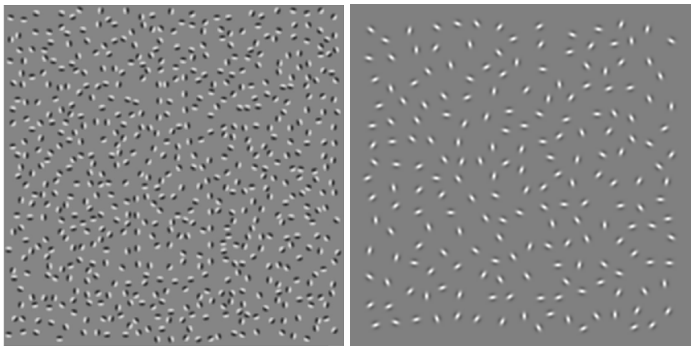


Figure 2.5: **Left:** An image extracted from Nygård et al. [63]. **Right:** Example of an alignment detection experiment to be developed here.

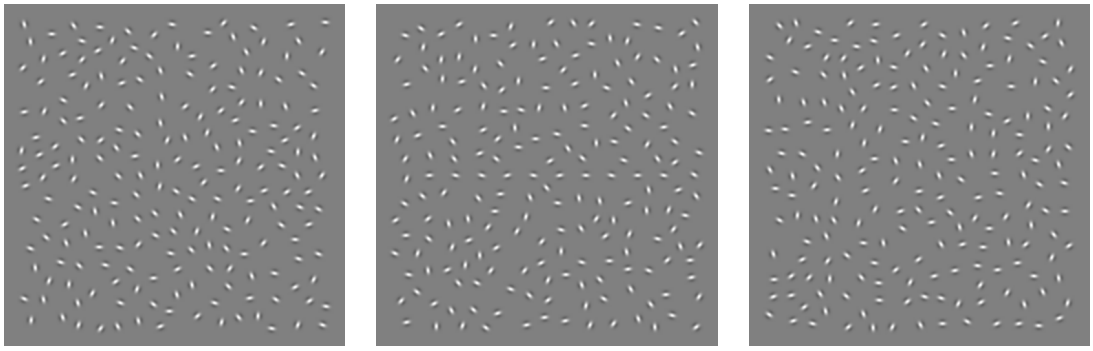
2.5 Detection Theory versus Psychophysics

In this section we leave the question of a quantitative gestaltism and get closer to more classic Psychophysics. The question is whether a quantitative framework like the *a contrario* detection theory can also become a useful addition for psychophysical experiments in human contour perception.

Arrays of Gabor patches have become a classic tool for the study of the influence of good continuation in perceptual grouping [21, 42, 34, 33, 63, 73, 32]. Gabor functions ensure a control on the stimuli spectral complexity and on the spatial scale of the contours. They give a flexible and easy way for building a great variety of stimuli. It has been verified that the more aligned the Gabor patches are to the contour they lie on, the easier their perceptual grouping into a shape's outline [21, 63, 73]. Fig. 2.5 (left), shows an easy example where most subjects recognize a bottle. But the more freedom is left to the Gabor orientation, the harder it is to distinguish such contours from the background. For the influence of other perturbations of the contour such as its motion or its curvature on the object's identifiability, we refer to a recent study [63].

Can we hope for a quantitative interpretation to this experimental framework, namely a function of the stimuli parameters that would predict and explain the evolution of the detection performance? Probabilistic approaches (mainly Bayesian) exist for contour modeling from a perceptual point of view [2, 20], and have sometimes been compared experimentally to human visual perception [97, 16, 13]; but these models usually include several parameters that are optimized to match experimental data. Indeed, in general a Bayesian formulation uses prior probabilities and likelihood terms. This requires hypothesis about the distributions and parameters estimations. In some problems, these are natural and lead to optimal predictions; in others, the optimality is undermined by the absence of sufficient knowledge. The *a contrario* methodology does not require a *probabilistic* model of the sought Gestalt; it only requires a deterministic description of the ideal structure and a background (or *a contrario*) hypothesis H_0, which assumes the maximal entropy for the distribution of the data, and models the absence of the relevant pattern. The Bayesian approach is preferable when all the distributions and their parameters are well known, as it is able to take full advantage of this information. In the lack of such knowledge, the *a contrario* methodology offers a good alternative, as it avoids heuristic choices. In our case, we don't have, *a priori*, perfect knowledge of what a salient alignment is for human perception.

The influence of experimental factors such as the length of the alignment, the density of the patches, and the angular accuracy on human detection is a classic subject of psychophysical inquiry. But the question of whether human performance can be measured with only one adequate quantitative function of the parameters is still open. We shall explore in this thesis if the NFA



(a) (b) (c)

Figure 2.6: Three examples of stimuli used in our experiments. **(a)** A jitter-free alignment with 10 elements. **(b)** A weakly jittered alignment with 10 elements. **(c)** A stimulus with no alignment, containing only elements with random orientations.

furnished by the *a contrario* theory can play this role. Indeed, the NFA retains the remarkable property of being a scalar function of the three psychophysical parameters generally used in this kind of detection experiment. In classic experimental settings, these parameters are varied separately and independently, and no synthetic conclusion can be drawn; only separate conclusions on the influence of each parameter can be reached. If a function like the NFA could play the role of generic *detectability* parameter, the experimental parameters could for example be made to vary simultaneously in the very same experiment. In short, if the hypothesis of a single underlying detection parameter is validated, this would entail a new sort of quantitative analysis of the results, two stimuli being *a priori* considered as equivalent in difficulty if their NFA are similar.

The underlying hypothesis, that the reaction of the subjects to varying stimuli might be predicted as a single scalar function of the stimulus' parameters, is equivalent to the classic hypothesis of a "single mechanism" for contour detection. More precisely, we shall explore if this single mechanism might obey the non-accidentalness principle (the NFA being a probabilistic quantitative expression of it).

To keep the line of the previous section, this one focuses again on the same simple gestalt: straight contours, that is, alignments of Gabor elements, as illustrated in Fig. 2.5 (right). The remainder of this section describes our first attempt to bring together a Turing test and an experiment on visual perception, explaining the main ideas that have been the basis of the work detailed in the next chapters. We briefly present the stimuli we used, a prototypical experiment performed on humans, an *a contrario* algorithm and some results of their comparison. To conclude this tour, we propose to use our stimuli in an online game, in which the subjects should be pushed to the limits of their perception.

2.5.1 The Patterns

Figure 2.6 shows three examples of the stimuli used this experiment. All of them consist of symmetric Gabor elements with varying positions and orientations placed over a gray background. There are two kinds of stimuli: positive stimuli and negative stimuli. Negative stimuli contain elements with random orientations sampled in $[0, \pi)$, e.g. Fig. 2.6 (c). Positive stimuli, see Fig. 2.6 (a) and (b), contain a majority of random elements like in negative stimuli but also a small set of *foreground* elements. The latter lie on a straight line and are uniformly spaced; their orientations are randomly and uniformly sampled from an interval centered on the alignment direction. The size of

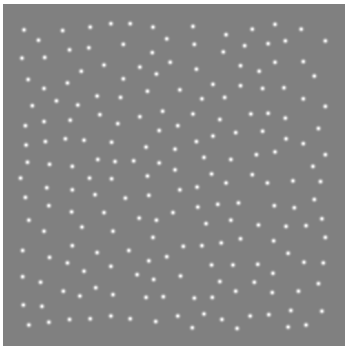


Figure 2.7: Influence of stimuli position. Each dot represents the position where a Gabor element will be placed. The figure includes a perfectly regular and aligned set of dots, surrounded by random placed elements, all generated by the GERT package. It is very difficult to find the alignment, which shows that the position of the elements by itself conveys few cues about the presence of the alignment. (For a comparison, see the same stimulus with Gabor elements, Fig. 2.5 (right), where the alignment is easily spotted.)

this interval gives a measure of the *angular jitter* and will be noted by J. When the jitter is zero, the foreground elements have the exact same orientation as their supporting line. Inversely, a jitter of π leads to completely isotropic elements.

The experiment is designed to study how *angular* jitter affects visibility. Yet, a natural question arises about the contribution to the detection of the accuracy of the alignment and of the regular spacing of the aligned elements. All the stimuli presented in this section, and in the rest of the thesis, were generated with the software GERT (v1.1) that includes special algorithms for the generation of random placed and oriented Gabor elements that mask as much as possible the aligned Gabor elements structure [7]. Figure 2.7 shows an example displaying only the elements position; even if there is in fact a set of perfectly regularly aligned dots, it is very hard to spot them. This suggests that the position of the elements carries few useful cues about the alignment.

2.5.2 The Detection Algorithm

Let us now present our first detection algorithm for alignments of oriented elements, that will be matched to human perception. A more sophisticated version will be featured in Chapter 3. The input to the algorithm is a set of Gabor elements $\mathbf{g} = \{(x_i, \theta_i)\}_{i=1\ldots N}$, defined by the position and orientation of each element. We will further assume that the total number of elements is a fixed quantity N.

A candidate to alignment is defined as a rectangle r, see Fig. 2.8 (left), and the orientation of the Gabor elements inside it will determine whether the candidate is evaluated as a valid alignment or not. The orientation of each Gabor element is compared to the one of the rectangle and when the difference is smaller than a given tolerance threshold τ, the element is said to be τ*-aligned*, see Fig. 2.8 (right). Two quantities will be observed for each rectangle r: the total number of Gabor elements inside it, $n(r)$, and the number among them that are τ-aligned, $k_\tau(r)$. The *a contrario* validation is analogue to the one described in Sect. 2.4.3.

Due to the way the patterns are generated, the only relevant information to evaluate in an alignment is the orientation of the Gabor elements. Consequently, the *a contrario* model H_0 is defined with N random variables corresponding to the orientation of the elements and satisfying the following two conditions:

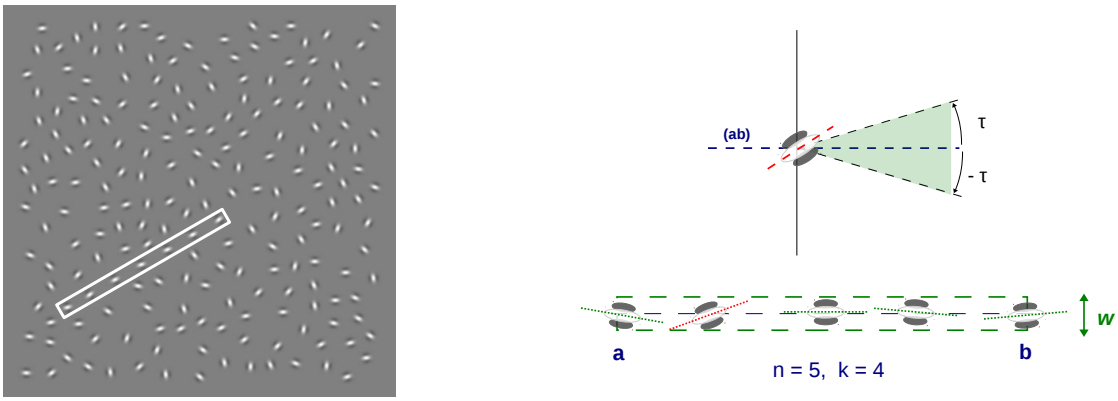


Figure 2.8: **Left:** A candidate to alignment, defined by a rectangle R. **Right-Top:** A Gabor element whose angle with (ab) is larger than τ and thus it is not counted as an aligned point. **Right-Bottom:** A detailed example where we see a total of five Gabor elements inside the rectangle, $n(r) = 5$, being τ-aligned with (ab), i.e. $k_\tau(r) = 4$.

- the orientations Θ_i are independent from each other;
- each orientation Θ_i follows a uniform distribution in $[0, \pi)$.

Under these *a contrario* assumptions, the probability that a Gabor element be τ-aligned to a given rectangle is given by

$$p(\tau) = \frac{2\tau}{\pi}. \tag{2.16}$$

Notice that the symmetric Gabor elements are unaltered by a rotation of π rads. The independence hypothesis implies that the probability term $\mathbb{P}[k_\tau(R) \geq k_\tau(r) | n(R) = n(r)]$, where R is a rectangle in a random set of Gabor elements following H_0, is given by

$$\mathbb{P}\left[k_\tau(R) \geq k_\tau(r) | n(R) = n(r)\right] = \mathcal{B}\Big(n(r), k_\tau(r), p(\tau)\Big), \tag{2.17}$$

where as before $\mathcal{B}(n, k, p)$ is the tail of the binomial distribution.

We still need to specify the family of tests to be performed. Each pair of dots will define a rectangle of fixed width w, so the total number of rectangles is $\frac{N(N-1)}{2}$. The width w is chosen as a small fraction $(1/10)$ of the average distance between nearest neighbors in the set of dots. Also, a finite number of angular precisions τ_i will be tested for each rectangle. Then,

$$N_{tests} = \frac{N(N-1)}{2} \cdot \#\mathcal{T}, \tag{2.18}$$

where $\#\mathcal{T}$ is the cardinality of the set $\mathcal{T}$ of tested precisions. The NFA of a candidate is defined in two steps. First, for any $\tau \in \mathcal{T}$, we set

$$\mathrm{NFA}^{(\tau)}(r) = N_{tests} \cdot \mathcal{B}\Big(n(r), k_\tau(r), p(\tau)\Big), \tag{2.19}$$

and finally

$$\mathrm{NFA}(r) = \min_{\tau \in \mathcal{T}} \mathrm{NFA}^{(\tau)}(r). \tag{2.20}$$

A small NFA value indicates a rare and interesting event in the *a contrario* model, whereas a large NFA value might corresponds to a likely (and therefore insignificant) configuration. The proposed detection method validates a rectangle candidate r whenever $\mathrm{NFA}(r) < \varepsilon$. The non-accidentalness property of Theorem 2 holds for this NFA as well.

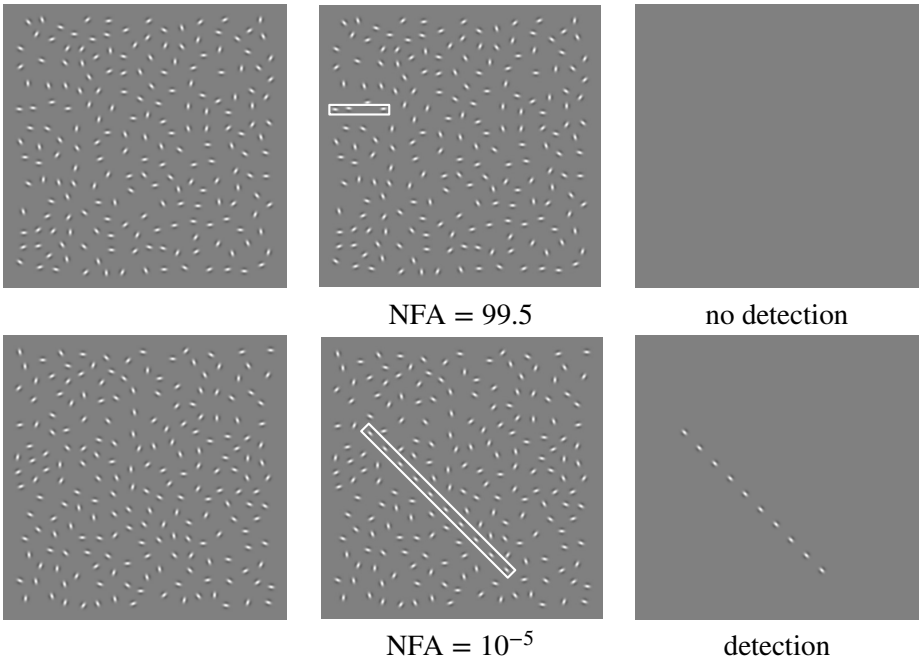


Figure 2.9: Two examples of the proposed validation method for alignment of Gabor elements. The rectangle in the first row has three elements inside, all of them aligned; that number is nevertheless too small to produce a detection, as its NFA value is larger than one. In the second example, all ten Gabor elements are aligned, giving an NFA of 10^{-5} and producing a detection.

Theorem 4.

$$\mathbb{E}\left[\sum_{R\in\mathcal{R}} \mathbb{1}_{\mathrm{NFA}(R)<\varepsilon}\right] \leq \varepsilon.$$

where $\mathbb{E}$ is the expectation operator, $\mathbb{1}$ is the indicator function, $\mathcal{R}$ is the set of rectangles considered, defined by the pairs of dots in a random set of Gabor elements on H_0.

To prove Theorem 4, it is sufficient to apply Theorem 2 to $\mathrm{NFA}^{(\tau)}$ with $n(r) - k_\tau(r)$ as measurement, and then deduce the result for the NFA of equation 2.20.

Once again we follow Desolneux et al. [8, 11] and set $\varepsilon = 1$. In our experiments, we use the NFA as an indication of the visibility of the gestalt according to the proposed theory; a value considerably smaller than 1 is "non-accidental" and should imply a conspicuous gestalt. A value larger than 1 can occur just by chance and might therefore be associated to an irrelevant gestalt. Figure 2.9 shows two examples of detection by this method.

2.5.3 Experiment

An informal psychophysical experiment was performed online by voluntary subjects using an interactive website[4]. Their task was to report on the visibility of the aligned Gabor patterns. The online methodology was necessarily more flexible and less controlled on various aspects than it would be in a laboratory: we had no reliable information about the subjects, their visualization conditions in front of their computers were not controlled, the comprehension of the task by the subjects might vary, etc. Notwithstanding their uncontrolled essence, online experiments give access to a larger

[4] `http://dev.ipol.im/~blusseau/aligned_gabors`

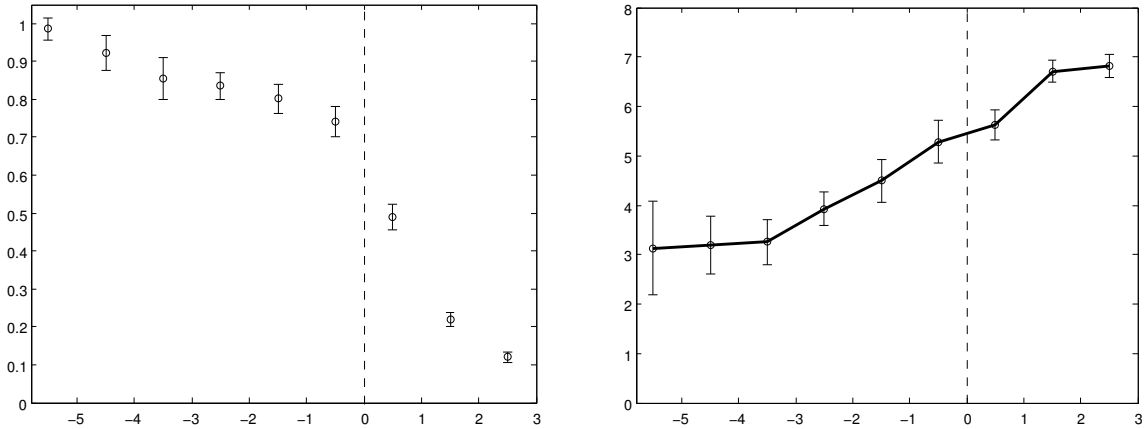


Figure 2.10: Comparison of the subjects' responses to the NFA. **Left:** The average answer rate is plotted relative to $\log_{10}$(NFA). Each point indicates the proportion of positive answers to stimuli with best NFA in the corresponding bin. **Right:** The average response times in seconds per bin. In both cases, the abscissa represents the scale of $\log_{10}$(NFA) divided into 9 bins; the first bin is defined by $\log_{10}(\text{NFA}) < -5$, the last one by $\log_{10}(\text{NFA}) \geq 2$, and the other 7 bins by $k \leq \log_{10}(\text{NFA}) < k+1$ for $k = -5, \ldots, 1$. The error bars give approximately 95 % confidence about the mean values (each interval is defined as $[\overline{x} - 2\frac{s}{\sqrt{n}}, \overline{x} + 2\frac{s}{\sqrt{n}}]$, where $\overline{x}$, s and n are respectively the mean, standard deviation and number of trials of the bin).

number of subjects and bring a great experimental flexibility. In Chapter 3, a better supervised version of this prototypical experiment will be described.

The data set used here is composed of over 14000 stimuli (negative and positive) as the one described in Sect. 2.5.1. Each image has a size of 496×496 pixels and containing $N = 200$ elements. For positive stimuli, 9 levels of jitter ($J \in \{0, \frac{\pi}{5}, \frac{\pi}{4}, \frac{\pi}{3}, \frac{\pi}{2}, \frac{2\pi}{3}, \frac{3\pi}{4}, \frac{4\pi}{5}, \pi\}$) and 8 different segment lengths were used, between 3 and 10 elements. During each session, the subject saw 35 of these images, one after another. The first five images were training stimuli and no results were recorded at this stage of the experiment. The following 30 images were randomly sampled over the data set, with constraints that ensured a balance between negative, positive, hard and easy stimuli. For each stimulus, the subject was asked to answer whether they saw or not a "straight line"; the answer and response time were recorded. There was no time limit to provide the answer but it was suggested to answer as soon as the subject made up their mind. At the end of the session, a feedback was given on false detections and on the consistency of the subject's answer through an "attention score". This score rewarded the fact that the subject answered better on easy stimuli than on hard ones and indicated if the task was well understood or not.

2.5.4 Results

In order to compare human and machine perception we precomputed the NFA for each rectangle on all the images of the data set. Each image was associated to its best (smallest) NFA. The hypothesis to be tested was that the NFA value should be related directly to the visibility by humans; if this is true, the average score given to an image by humans, namely the proportion of "Yes", should be related to the NFA of the most salient structure. In what follows we will analyze the data obtained from 7137 trials.

The NFA scale was divided into bins. To each bin were associated statistics on the trials whose

NFAs belonged to this bin. Figure 2.10 shows the average answer rate and response times for nine $\log_{10}(\text{NFA})$ intervals. Note that NFA < 1 (or $\log_{10}(\text{NFA}) < 0$) means detection of the alignment by the algorithm.

The results significantly support the hypothesis that a single scalar function of all parameters predicts the detectability. Indeed, the answer rate follows a sigmoid shape roughly centered at $\log_{10}(\text{NFA}) = 0$. The second graph, plotting the response time versus the NFA, also agrees with the hypothesis: the less visible the stimuli, the more time is spent searching for valid gestalts. Statistical tests confirm this average tendency.

The experiment seems compatible with the hypothesis that, at least in this restricted perceptual environment (formed of three parameters, the number of Gabor patches, the length of the alignment and its jitter on orientation), the value of NFA may account for the human "detectability" of an alignment. Surprisingly, the human detection (attentive) threshold is close to the best algorithm in this restricted environment. Indeed, alignments with NFA smaller than 1 were detected by a majority of subjects. Alignments with NFA larger than one, which are likely to occur just by chance, were detected by a minority of subjects. Furthermore, the detection curve is steepest when the NFA crosses 1. The curve is not as steep for the mean response time as a function of NFA. This can be simply explained by the fact that the patience of subjects undergoes a rapid temporal erosion; they are not ready to look long for a needle in a haystack. These encouraging results require confirmation and to be more firmly established. This will be the focus of Chapter 3.

2.5.5 Consequence: An Online Game

Online experimentation opens new possibilities that need to be explored farther, and in particular the use of computer games as an experimentation tool. A successful game may attract the attention of subjects and if the resulting mass of results is large enough, it could compensate for the lack of control on other aspects of the experimental setting.

The player of a computer game is usually directed toward an objective and faced with obstacles. To be attractive, a game cannot be too easy, but not too hard either; a good balance of this difficulty is the key to the popularity of the game. To use games for psychophysical purposes, the player should be directed to detect some pattern, the obstacles being the conditions that preclude this perception. Motivated players will do their best effort, revealing the limits of human perception.

To go in this direction we created a prototype version of an alignment game.[5] The player is presented with Gabor stimuli as described before. Only positive stimuli with variable difficulties are used. In this way one knows that there is an alignment gestalt; but its position is unknown and the assignment of the player is to spot it. The subject is asked to click in the image on any point of the straight line. The distance between the clicked point and the actual line segment is recorded, and if the distance is low enough, the player gets a score decreasing exponentially with the reaction time. Otherwise, the player gets a zero score. When the stimulus is quite visible, all subjects are able to point correctly to it; when it is not, the distance to the alignment becomes random. This rash transition should permit to pinpoint the human detection threshold.

The presentation of the stimuli is divided into several sequences of ten images. The first sequence is always supposed to be very easy (long segments with little jitter). Then the difficulty of the following sequences changes according to the performance achieved on the previous one. The collected data could permit us to compare a detection method to human performance in the way described before. This game is only a prototype, but it actually inspired the protocols of our experiments, described in the next chapters.

[5] `http://dev.ipol.im/~blusseau/srv1`

2.6 Conclusion

Needless to be said, the experimental devices and first results that we just described are not sufficient to make any rash conclusion on the existence of quantitative predictions on human perception. They need to be extended to more rigorous protocols, as those presented in the following chapters, an to other gestalts commonly used in Psychophysics, such as for example general good continuation (Chapter 4). In short, this was a preliminary work that laid the basis of the studies detailed in the remaining of the dissertation.

3 Measuring the salience of alignments by their non-accidentalness

The present study builds on the preliminary work of the previous chapter. It introduces a new quantitative approach based on the *a contrario* theory, which formalizes the non-accidentalness principle for good continuation. Like in Chapter 2, this model yields an ideal observer algorithm, able to detect non-accidental alignments in Gabor patterns. To evaluate the approach, we compared this ideal observer with the human attentive performance, on three experiments of straight contours detection in arrays of Gabor patches, generalizing and better formalizing the protocols described in Section 2.5. Confirming our previous observations, the experiments presented here showed a strong correlation between the detectability of the target stimuli and their degree of non-accidentalness, as measured by our model. What is more, the algorithm's detection curves were very similar to the ones of human subjects. This fact seems to validate our proposed measurement method as a convenient way to predict the visibility of alignments. This framework could be generalized to other Gestalts.

3.1 Introduction

The question of how vision integrates a set of elements into the contour of a shape was raised by the Gestaltists. They identified several conditions that favored the emergence of such percepts [90, 58, 37, 88, 89]. Among them is the well known grouping principle (or *Gestalt*) of good continuation, that refers to the perceptual grouping of elements forming smooth curves.

This principle has been further investigated in Psychophysics, with a recurring use of arrays of Gabor patches as stimuli (see [33, 34, 32] for a review). The experiments described in [21] opened a long standing research line dealing with various aspects of the *association field* hypothesis. For example, some studies [44, 84] compared the detection of paths formed by Gabor patches oriented parallel to the contour (*snakes*), to the detection of *ladders*, in which all the elements are orthogonal to the path, and *ropes* (for [44]), where elements are obliquely oriented with respect to the path. In [63], the effect of orientation jitter, analogous to the "α" parameter of [21], was measured on the detection and identification of everyday objects' contours. The combination of the association field grouping effect and other potential cues, such as symmetry [55, 74], closure of contours as well as the surface enclosed in contours [42, 54], was also studied. As far as the detection of snakes is concerned, a result that is common to all the previously mentioned experiments is that the visual system is better at detecting *smooth* paths rather than jagged ones, and all the more so as they are formed by elements that are roughly parallel to the local tangent of the contour.

An important step towards the understanding of the mechanisms underlying these observations

is the definition of a model that could *explain* and *predict* the quantitative results obtained experimentally. Can we predict, for example, as observed in [21], that a path composed of 12 Gabor patches embedded in an array of 244 other elements randomly oriented, is likely to be detected when the absolute angle β between two consecutive edges is less than 60°, and without orientation jitter ($\alpha = 0°$)? What about the decay of the detection performance for a fixed β when $\alpha = 15°$ and 30°? What if the path was composed of, say, only 8 elements? Computational models designed to make such predictions have already been proposed [97, 13]. The framework described in [97] consists in a cortex-inspired network, in which the interaction between cells is ruled by a mathematical formulation of the association field. In accordance with the experimental results of [21], the latter favors the synchronization between cells responding to close, co-circular stimulations, as well as among straight ladders configurations. The resulting algorithm contains 11 parameters that were tuned by optimizing the detection results on a few images. It is compared to the psychophysical data of [21, 42, 39], showing an interesting match between the perceptual results and the model's detections. Equally interesting is the Bayesian approach of [13]. By defining a generative model of contours, the authors could compare human subjects to an ideal observer in their own experiment of contour detection. They used the psychophysical data they collected as a reference, and tuned the parameters of their algorithm by optimizing the correlation between the subjects and the model's responses. Although the obtained *"optimal observer"* still performed significantly better than the average of subjects, the model managed to mimic the dynamics of the subjects' responses, and to reproduce some observed phenomena, such as the early detection of longer contours. Another quantitative approach to contour detection in noise is [91], which does not propose an ideal observer but a Bayesian prediction of the detectability of contours relative to their complexity.

The present chapter introduces a new approach to build a non-parametric predictive model of good continuation, still based on the *non-accidentalness principle*. As explained in Section 2.2, this concept has emerged at the crossroad of human and computer vision research. Indeed, faced with the quantitative questions raised by the qualitative results found by the Gestalt school [90, 37, 88], vision scientists like Witkin and Tenenbaum remarked that *"the appearance of spatiotemporal coherence or regularity is so unlikely to arise by the chance interaction of independent entities that such regular structure, when observed, almost certainly denotes some underlying unified cause or process"*[95, p. 481]. Although it was particularly well addressed in [95], previous or contemporary formulations of non-accidentalness exist. It has inspired computer vision since the 1980s, when researchers relied on the so called *non-accidental features* to perform automatic scene interpretation [76, 77, 35, 3, 53, 92, 49, 50], and it was crucial in the success of the scale-space filtering method [93] which led to actual breakthroughs in computer vision [51, 52]. What is more, the concept of non-accidental features also sparked the interest of psychophysicists, who wanted to test their relevance in perception [86, 87, 18].

The *a contrario* theory is a mathematical formulation of non-accidentalness, as we saw in Section 2.3. It was confronted to human perception in [9]. Although it was not explicitly presented as such, this work somehow tried to investigate the role of non-accidentalness in visual perception. It already suggested that this approach might give an interpretation and quite accurate predictions of the perceptual thresholds. Its most remarkable insight is to translate several detection-affecting parameters into one unique measure of non-accidentalness, the NFA, which correlates well with the subjects detection performance. Yet, the psychophysical protocols in [9] had several limitations that also limited the strength of their conclusions.

First, like in Section 2.5, the experimental set-up in [9] relied on "yes-no" questions to assess the detection by subjects. Here, we used a refined protocol in which subjects indicate *where*, instead of *whether*, a structure is seen. Second, in [9] the authors did not compare *directly* human subjects

to an algorithm. In contrast, here we took inspiration from the innovative work by Fleuret et al. [22] where human and machine performed the same visual categorization task, "side by side", on each stimulus. In a nutshell, the algorithm became an artificial subject, as it was also the case in [13]. Similarly, our experiments compare human and machine vision in a simple perceptual task: the detection of jittered straight contours among a set of randomly oriented Gabor patches (Fig. 3.1). Following the definition of [23, p. 2], "an ideal observer is a hypothetical device that performs a given task at the optimal level possible, given the available information and any specified constraints." In this sense, our *a contrario* detection algorithm is an ideal observer under the constraint that no relevant structure should be detected in noise.

As we said earlier, the non-accidentalness principle can be interpreted as imposing minimal reliability conditions to percepts. A configuration that could have arisen by chance does not provide reliable information and should be rejected. Natural selection evolved reliable perception mechanisms which may obey a similar sound requirement, that no detection should occur in noise [2]. Undoubtedly, this minimal reliability condition is not the only factor involved in the salience of a particular structure. Nevertheless, it is interesting to evaluate whether this factor alone can predict the visibility of a specific Gestalt to some extent. The purpose of this chapter is precisely to study to what extent the *a contrario* formulation of non-accidental alignments can account for their salience in a large set of stimuli. We shall illustrate that this theory permits to summarize into one single measure, the NFA, the many experimental parameters affecting this visibility. By addressing in this chapter the case of alignments, we hope to point out the key features that make the approach applicable to other Gestalts. Indeed, given a grouping law described qualitatively, the *a contrario* framework proposes a way to evaluate quantitatively how strongly the configurations of elements must stick to this grouping law, to be distinguished from a general, accidental configuration.

In Sections 3.2-3.5 we present our experimental set up, consisting in three versions of a straight contours detection task, and expose the results obtained by 32 subjects. After describing our model and detailing a parameterless detection algorithm in Section 3.6, we characterize our set of stimuli by their level of non-accidentalness, that we measure by the NFA. This enables us to evaluate, in Sections 3.7 and 3.8, how non-accidentalness correlates with detection performance, and if an *a contrario* algorithm mimics, and therefore can predict, the subjects behavior on average. Finally, Section 3.9 summarizes our conclusions.

3.2 Methods

This section presents the techniques that are common to all the experiments described in the chapter.

3.2.1 Stimuli

In all the following experiments, subjects were shown arrays of Gabor patches of 496×496 pixels, similar to those represented in the first row of Figure 3.1. They were generated with the software GERT (v1.1) [7]. An array contained a certain number N of patches, each of which represented a symmetrical Gabor function, characterized by its position in the image and its orientation θ. The family of Gabor functions we used is described by

$$\forall (x, y) \in \mathbb{R}^2, \qquad G_{f,\theta,\sigma}(x, y) = \frac{1}{2}\left(1 + e^{-\frac{x^2+y^2}{2\sigma^2}} \cos\left(2\pi f (y\cos\theta - x\sin\theta)\right)\right) \tag{3.1}$$

f being the spatial frequency, σ the space constant set to $\sigma = \frac{1}{4f}$ and θ an angle in $[0°, 360°]$. Note that this function is centered on the origin, and can then be placed at the desired position in the image.

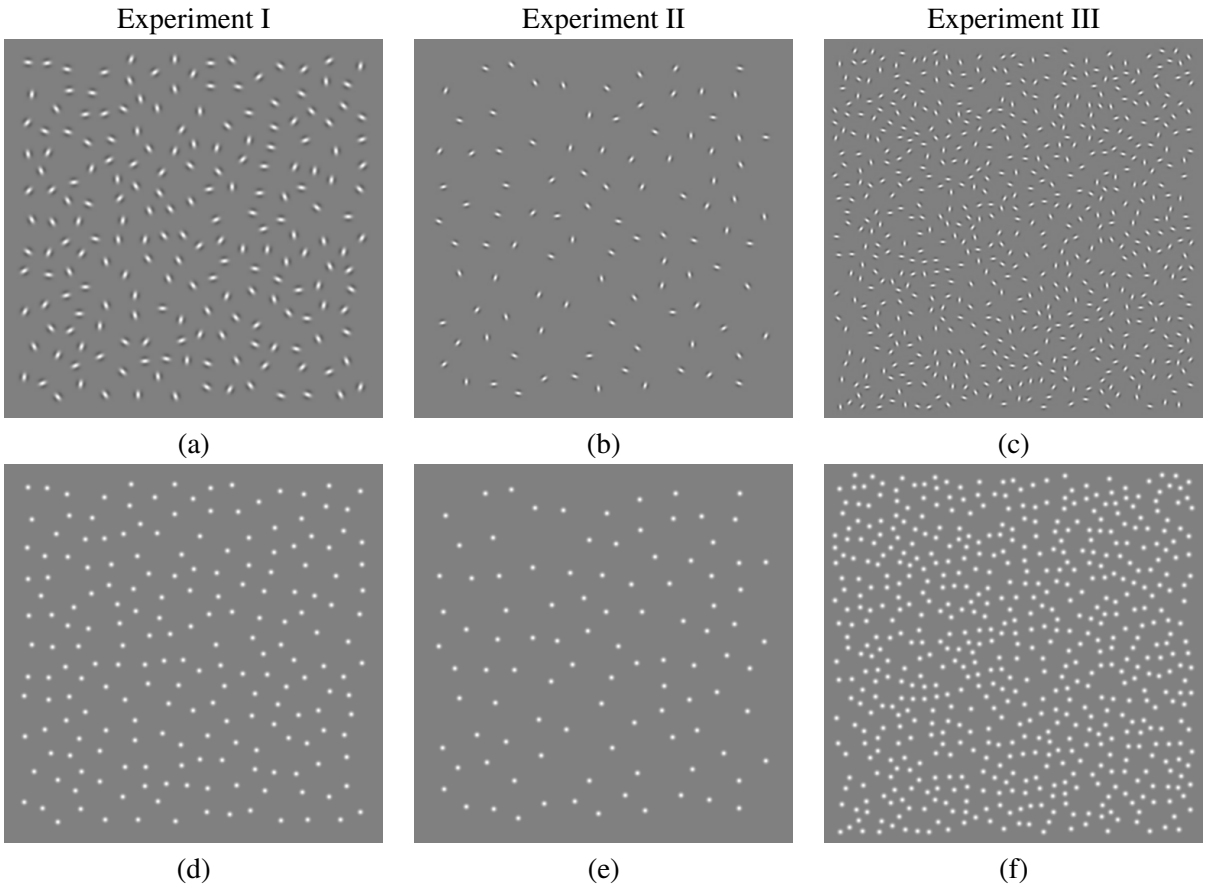


Figure 3.1: First row: three arrays of N Gabor patches with long, moderately jittered straight contours; $N = 200$ in (a), $N = 100$ in (b) and $N = 600$ in (c). Images (d), (e) and (f) show dots representing the coordinates of the Gabor elements of images (a), (b) and (c) respectively.

Each array included n aligned patches, regularly spaced by a distance r_a. Their orientations were chosen as follows: given the line's direction θ_l, the n angles were randomly and independently sampled according to a uniform distribution in $[\theta_l - \alpha, \theta_l + \alpha]$, with $\alpha \in [0°, 90°]$. This is equivalent to adding uniform noise, also called *angular jitter*, to orientations equal to θ_l. The alignment position and direction θ_l were selected randomly and uniformly in each image. In all the chapter, these aligned patches will be referred as *target alignment* or *target elements*. The $N-n$ non-aligned elements, called *background elements*, were randomly placed but respected a minimal distance r_b from each other and from the aligned patches. Their orientations were randomly and independently sampled according to a uniform distribution in $[0, 180°)$. Distances r_a and r_b were set as functions of N to fulfill two requirements. First, r_b was tuned so that all N elements fit into the image and fill it homogeneously, avoiding clusters and empty regions. Second, we chose r_a larger than r_b to make the alignment almost impossible to detect from the coordinates of the elements only, that is to say from the proximity and width constancy cues only (Fig. 3.1 (d)). More precisely, we set $r_b = 1.3 r_{hex}$ and $r_a = 2 r_{hex}$, where $r_{hex} \stackrel{\text{def}}{=} \frac{456}{\sqrt{2N\sqrt{3}}}$ is, in pixels, the maximal radius for N discs to fit in a 456×456 pixels square without intersecting (the number 456 corresponds to the image side minus two margins of 20 pixels, one at each border). Then the difficulty to detect the aligned Gabors was essentially ruled by their number n, and by the angular precision α. The larger n and

the smaller α, the more conspicuous the alignment.

3.2.2 Procedure

Each experimental session consisted of 126 trials divided into three blocks of 42, with pauses after the first and the second block. In each trial, the subject was shown on a screen an array containing N Gabor patches, n of which were aligned ($n \in \{4, \dots, 9\}$) and affected by an angular jitter of fixed intensity $\alpha \in \{9°, 15°, 22.5°, 30°, 45°, 90°\}$. Four trials were conducted for each couple of conditions (n, α) with $\alpha \leq 45°$, and one trial per maximal jitter condition $(n, 90°)$, in a random order. The subjects knew that every stimulus contained an alignment, but did not know where, how long and how jittered it was. They were asked to click on an element they perceived as part of the alignment, and to give their best guess in case they did not see any clear alignment. We wanted the task to be *attentive*. The stimulus remained on the screen until an answer was given. The elapsed time between the presentation of the stimulus and the click is what we call the reaction time, and was measured at each trial. Subjects were advised to spend no more than 20 seconds per stimulus. They were free to follow this qualitative suggestion or not and no time information was provided during the sessions. The coordinates of the click were recorded as well. After the subject's click, a transition gray image was displayed for 500 ms and then the next stimulus appeared.

To help understand the task, subjects were first shown two training sequences of 5 and 10 trials, with the corresponding target alignments at the end of each sequence.

3.2.3 Apparatus

The monitor properties, the distance to the screen and the illumination conditions could vary from one subject to another. Indeed, various computers were used at different places along the sessions. However we can estimate the average distance between the subjects' eyes and the monitor to be about 70 cm, and the average size of a pixel to be approximately 0.02 cm $\times$ 0.02 cm, so that the 496×496 images subtended about $8° \times 8°$ of visual angle in average. The experiments were set up to work on a web browser, and they were taken by selected subjects under our supervision. They are accessible at `http://bit.ly/na_alignments`. Our bet was that, for the scope of our study, some variability in the viewing conditions would not have significant inter-subject effects. What differed between the experiments described hereafter, was the number N and the size of the Gabor patches.

3.3 Experiment I ($N = 200$)

A previous version of this experiment was presented in [4].

3.3.1 Subjects

Twelve subjects, five women and seven men, with age between 20 and 40 years old, took the experiment (accessible at `http://bit.ly/ac_alignments`). All the subjects were naive to the purpose of the experiment and had normal or corrected to normal vision.

3.3.2 Stimuli

The stimuli of this experiment were those described in Section 3.2.1, with $N = 200$ Gabor patches per image, and a spatial frequency $f = 0.077$ cycles per pixel (see Figure 3.1(a) for an example). Four sequences of 126 images were pre-computed, according to the categorization detailed in Sec-

tion 3.2.2, and for each of the twelve sessions one sequence was picked randomly by the computer program. One subject was shown sequence 1, four subjects saw sequence 2, sequence 3 was seen by two subjects and sequence 4 by five subjects.

3.3.3 Results

Subject	1	2	3	4	5	6	7	8	9	10	11	12	all
Det. rate	0.63	0.65	0.65	0.56	0.64	0.52	0.67	0.80	0.61	0.64	0.80	0.63	0.65
Aver. r. t. (s)	7.7	9.5	8.4	12.9	14.1	7.5	9.3	20.9	7.9	8.2	21.0	14.4	11.8

Table 3.1: Individual detection rates and average reaction time per trial, for Experiment I.

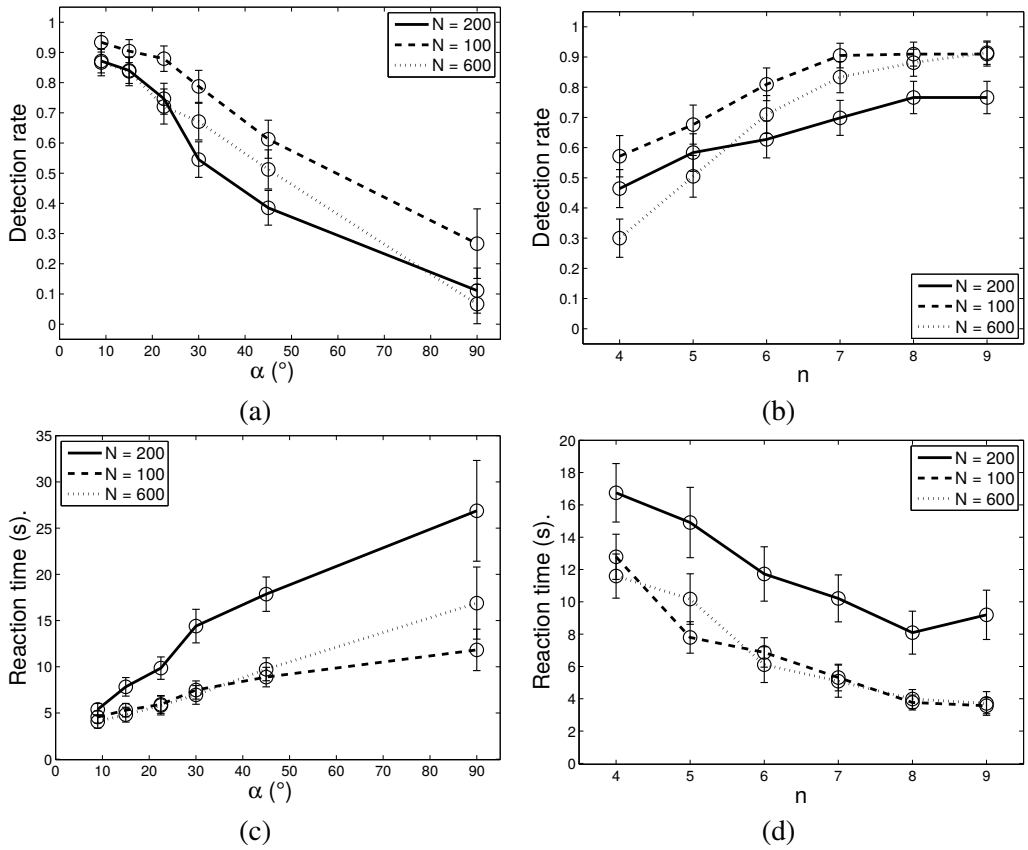


Figure 3.2: Results of the Experiments I-III (see also Tab. 3.1, 3.2 and 3.3). The subjects' detection rates (top row) and reaction times (bottom row) are plotted as functions of the jitter intensity α (left hand column) and the number n of aligned elements (right hand column). The error bars give 95 % confidence; they are defined as $\left[\overline{x} - 2\frac{\sigma}{\sqrt{q}}, \overline{x} + 2\frac{\sigma}{\sqrt{q}}\right]$, where $\overline{x}, \sigma$ and q are respectively the mean, standard deviation and number of trials of the corresponding condition.

Each click made by a subject was associated to the nearest Gabor element in the image, and counted as a valid detection when that element belonged to the target alignment. This count permitted to define the detection rate. Figure 3.2 displays the results of Experiment I (represented by a solid line), as well as those of Experiments II and III, described in the next sections. Fig. 3.2 (a) and (b) plot the detection rate as a function of the jitter level α and the number n of aligned elements,

whereas (c) and (d) plot the reaction times as a function of the same parameters (all the trials were counted; there was no attempt to remove outliers). The error bars give 95 % confidence; they are defined as $\left[\overline{x} - 2\frac{\sigma}{\sqrt{q}}, \overline{x} + 2\frac{\sigma}{\sqrt{q}}\right]$, where $\overline{x}, \sigma$ and q are respectively the mean, standard deviation and number of trials of the corresponding condition. In Table 3.1 are reported detection rates and reaction times achieved by each subject of Experiment I.

3.3.4 Discussion

Subjects 8 and 11 achieved higher rates than the rest of the subjects, but they also dedicated more time to the task. The results plotted in Figure 3.2 are consistent with those obtained in previous studies on the influence of orientation jitter on contour detection [21, 44, 63]. The more jittered the alignments are, the lower the average detection performance, and the longer the subjects took in looking for them (Figs. 3.2 (a) and (c)). It was also harder to detect short alignments than long ones (Figs. 3.2 (b) and (d)). It is worth noting that subjects still achieved about 10% detection rate in the maximal jitter condition (8 detections out of 72 trials), while chance level is $\frac{1}{6}\sum_{k=4}^{9}\frac{k}{N} = 3.3\%$ for $N = 200$. Three factors could explain these 8 clicks on one of the target elements in the maximal jitter conditions. The first one is pure chance, when the subject did not see any alignment and clicked on a random element, that turned out to belong to the target. This kind of event seems to have occurred in one of the eight mentioned cases. Indeed, the corresponding alignment appears to be impossible to detect, and the clicked point does not seem to belong to any other relevant structure (see Figure 3.3, first row). The second factor is the overlapping of the target and a random structure that looks like a jittered alignment. From the observation of the concerned stimuli and the subjects' answers, it seems that this scenario may explain 4 out of the 8 detections in noise (see two of these in Figure 3.3, second row). The three remaining clicks seem to be actual detections of a part of the target alignment composed of nine elements. Indeed, they occurred on the same stimulus, showing an agreement of three subjects on the same linear structure (Figure 3.3, third row). The latter case is the most interesting, since it shows that random orientations are not always enough to mask the alignment.

The *a contrario* detection theory allows the definition of parameterless algorithms that automatically adapt to changes in the size of the data. In order to demonstrate this property and compare it with perception, it was natural to set up experiments with stimuli containing less (see Section 3.4) and more (see Section 3.5) patches than in Experiment I. This is the motivation for Experiments II and III.

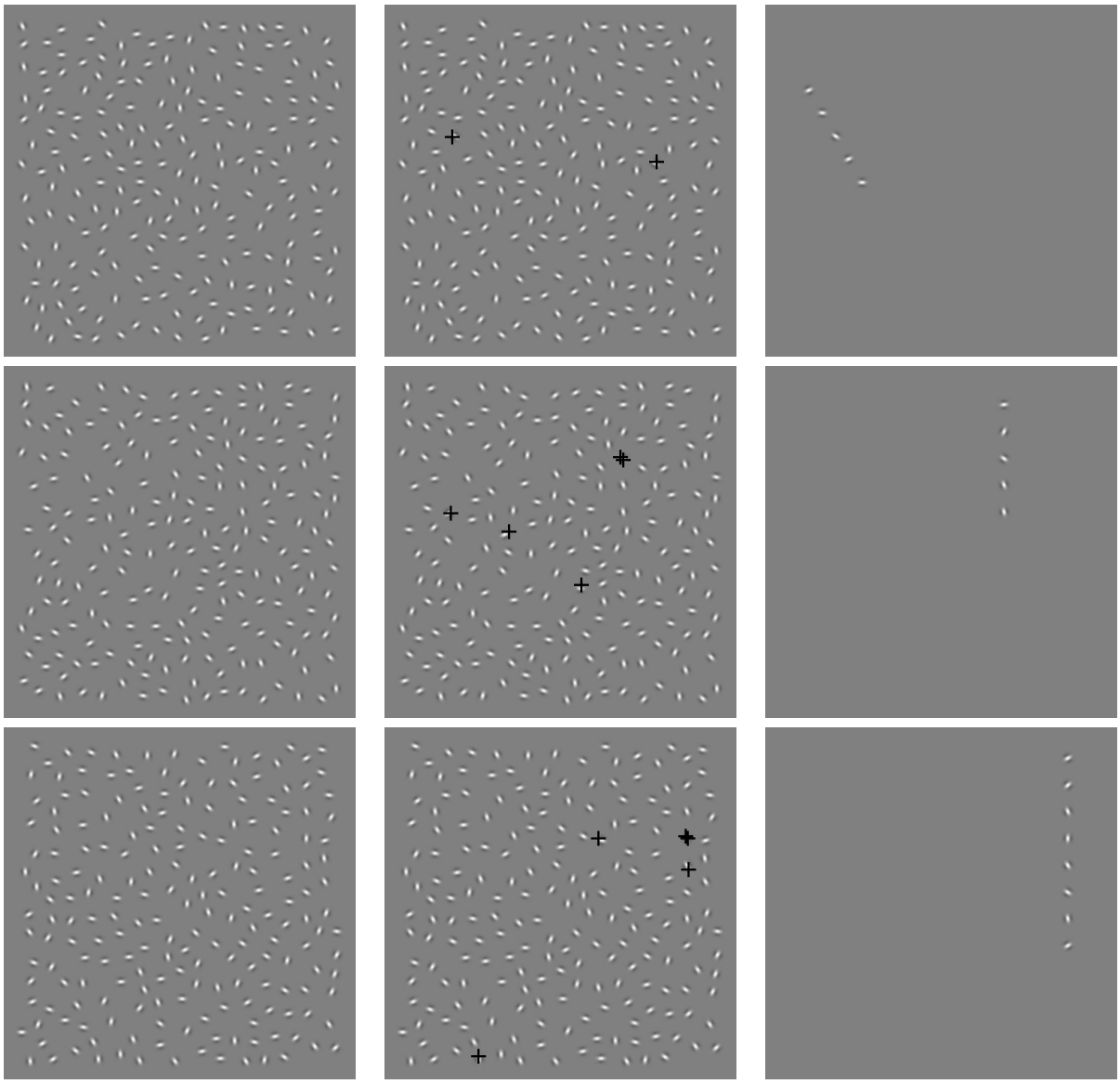


Figure 3.3: Three examples of detections in the maximal jitter condition in Experiment I. Each row represents an example. The left hand column shows the original stimuli, the central column displays the subjects' clicked points, and in the right-hand column are represented the target alignments. First row: one subject seems to have clicked completely randomly on the target alignment. Second row: the two subjects who clicked on the target probably saw a different random structure. Third row: three subjects seem to agree on the detection of a subset of the target, composed of nine elements.

3.4 Experiment II ($N = 100$)

3.4.1 Subjects

Ten subjects participated in Experiment II (accessible at `http://bit.ly/ac_alignments_100elts`). They were three women and seven men, with age between 20 and 60 years old, and had normal or corrected to normal vision. Contrary to Experiment I, here one of the subjects is a researcher who collaborated to the present study (subject 9 in Table 3.2); the others were naive to the purpose of the experiment.

3.4.2 Stimuli

The stimuli for this experiment differed from Experiment I by two parameters: first, they counted $N = 100$ Gabor patches each; second, the spatial frequency was set to $f = 0.12$ cycles per pixel, that is to say approximately 1.5 times the frequency used to build the stimuli of Experiment I. Figure 3.1(b) is an example. In these images, containing half as many elements but with same size as in Experiment I, the patches are more distant from each other. By increasing the spatial frequency f and making Gabor patches look thinner than in Experiment I, we wanted to accentuate the sensation of empty spaces between neighboring elements. Only one pre-computed sequence of 126 images was used for this experiment, because we wanted to compare all the ten subjects on the very same stimuli.

3.4.3 Results and discussion

Subject	1	2	3	4	5	6	7	8	9	10	all
Det. rate	0.75	0.83	0.79	0.78	0.89	0.79	0.66	0.84	0.87	0.77	0.80
Aver. r. t. (s)	5.2	5.0	13.1	5.5	7.0	7.9	6.8	5.9	2.6	7.8	6.7

Table 3.2: Individual detection rates and average reaction time per trial, for Experiment II.

In Experiment II the subjects obtained globally better detection rates than those of Experiment I (see Tab. 3.1 and 3.2 and Fig. 3.2). This is not surprising since in this last experiment, the stimuli were less crowded than in the previous one. As a consequence, the set of stimuli presented less difficulty on average. Here we observe almost 30% for the detection rate in the maximal jitter condition (detection in 16 out of 60 trials), the chance rate being approximately 6.5%. Although the hypothesis formulated in Section 3.3.4 still hold to account for this observation, it seems that in most of the 16 cases the subjects *did* detect the target alignment, thanks to other cues such as the perfect alignment of the elements' coordinates, and their perfectly regular spacing. Since the proportion of aligned elements relative to the total number of elements was higher than in Experiment I, it is not surprising that the masking was sometimes less efficient here.

3.5 Experiment III ($N = 600$)

3.5.1 Subjects

As in the previous experiment, we had ten participants in Experiment III (accessible at `http://bit.ly/ac_alignments_600elts`), three women and seven men. They were between 20 and 60 years old, and had normal or corrected to normal vision. Except for one of them, who collaborated to the present study (the same person as in Experiment II and subject 5 in Table 3.3), the subjects were naive to the purpose of the experiment.

3.5.2 Stimuli

The stimuli were of the same kind as the one of Figure 3.1(c). They counted $N = 600$ Gabor patches each and the spatial frequency was set to $f = 0.12$ cycles per pixel, like in Experiment II. This time, the spatial frequency was chosen to avoid too dense displays and overlaps among patches. Only one pre-computed sequence of 126 images was used for this experiment, like in Experiment II.

3.5.3 Results and discussion

Subject	1	2	3	4	5	6	7	8	9	10	all
Det. rate	0.70	0.68	0.62	0.67	0.72	0.79	0.68	0.71	0.67	0.66	0.69
Aver. r. t. (s)	7.8	6.1	4.4	5.1	3.7	11.4	7.9	7.5	4.8	9.2	6.8

Table 3.3: Individual detection rates and average reaction time per trial, for Experiment III.

Like in Sections 3.3.4 and 3.4.3, we shall discuss the detection performance of almost 7% for the maximal jitter condition (4 out of 60 trials), above chance level (approximately 1%). Here the masking effect seems to have been efficient in 5 out of 6 stimuli with maximally jittered alignments. Indeed, the observed performance is mainly due to one stimulus, in which three subjects seem to have detected a subset of the nine elements forming the target alignment. In the remaining case it is more difficult to guess if the subject perceived the target alignment or another structure overlapping with it.

Since this experiment is the one with the most crowded stimuli, we would expect the detection rates to be lower than in all the previously presented experiments. This holds for the comparison between Experiments II and III, but subjects of Experiment I, in which there were $N = 200$ Gabor patches per stimulus, achieved a generally lower performance than those of Experiment III (see Tab. 3.1, 3.3 and Fig. 3.2). This could be related to the difference of spatial frequency defining the Gabor functions. In Experiment I, this frequency was lower and thus the patches' central blobs looked thicker than in Experiments II and III. This may have created a more efficient crowding effect, explaining lower detection rates in the former experiments, despite a smaller number of patches per image compared to Experiment III.

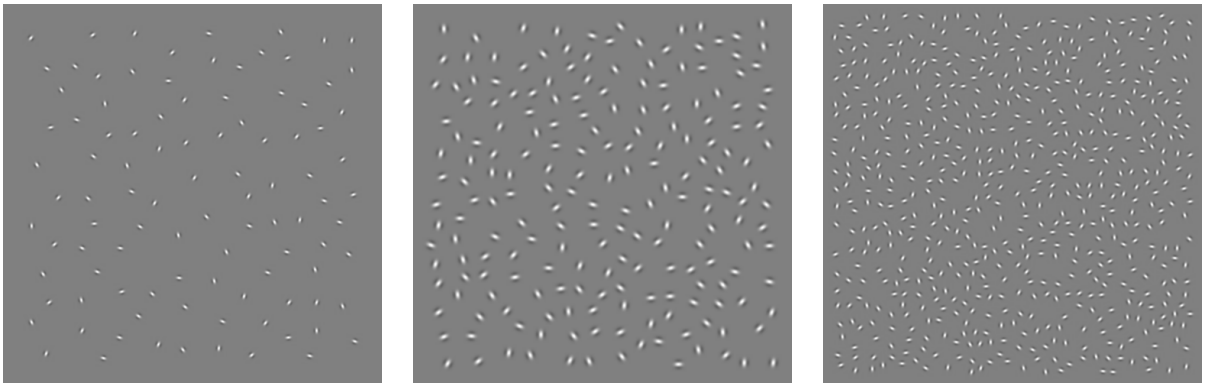
3.6 Model and algorithm

Recall that an *a contrario* method requires two models. On the one hand, a geometric model, which is deterministic and, on the other hand, a probabilistic model. In the following subsections, we detail these models for our study.

3.6.1 The geometric model

For a given tuple $(g_1, \dots, g_n)$ of n Gabor patches, we consider the variables $\alpha_1, \alpha_{2L}, \alpha_{2R}, \dots, \alpha_{(n-1)L}, \alpha_{(n-1)R}, \alpha_n$, as illustrated in Figure 3.5(e). Variable α_{iL} is the absolute angle between the orientation of Gabor patch i and the line joining it to the patch $i-1$, while α_{iR} is the same thing changing $i-1$ for $i+1$. Since the first and last patches in the tuple have no previous and next elements respectively, we simply note $\alpha_1 \stackrel{\text{def}}{=} \alpha_{1R}$ and $\alpha_n \stackrel{\text{def}}{=} \alpha_{nL}$. Then we define $\omega_1 \stackrel{\text{def}}{=} \alpha_1$, $\omega_n \stackrel{\text{def}}{=} \alpha_n$ and for $i = 2, \dots, n-1$, $\omega_i \stackrel{\text{def}}{=} \max(\alpha_{iL}, \alpha_{iR})$, see Figure 3.5(f).

The ideal alignment that will be our reference structure in the present study, is such a tuple of Gabor patches for which all angles ω_is are equal to 0°. In words, it is a set of aligned patches whose orientations are the same as their line's orientation.



(a) (b) (c)

Figure 3.4: Three arrays of Gabor patches containing only background elements and illustrating the *a contrario* model. There are $N = 100$ patches in (a), $N = 200$ in (b) and $N = 600$ in (c).

Then we measure the deviation of a general tuple from an ideal alignment by its maximum angle $\omega^* \stackrel{\text{def}}{=} \max\{\omega_1, \dots, \omega_n\}$.

3.6.2 The *a contrario* model

Now we need to define the *a contrario* model for our stimuli. Figure 3.4 shows three arrays of Gabor patches with *only background elements* and thus no significant alignment. Formally, we model an array of N Gabor patches by a set $\mathbf{g} = \{(x_i, \theta_i)\}_{i=1\dots N}$, where $x_i \in [0,1]^2$ represents the coordinates of patch number i, and $\theta_i \in \mathbb{R}$ its orientation (in this section the variables representing angles will be expressed in radians). We note $\mathbf{x} \stackrel{\text{def}}{=} \{x_1, \dots, x_N\}$ and, for any three points $x, y, z \in \mathbf{x}$, ϕ_{xyz} is the angle between vectors $\overrightarrow{xy}$ and $\overrightarrow{yz}$. Finally, the set of the 6 nearest neighbors of a point $x \in \mathbf{x}$ is noted $\mathcal{N}_6(x)$ (see Figures 3.5 (a) and (b)). Then we define an *a contrario* array of Gabor patches as a random set $\mathbf{G} = \{(X_i, \Theta_i)\}_{i=1\dots N}$ verifying two properties:

1. The random variables $\Theta_1, \dots, \Theta_N$ are independent and uniformly distributed in $[0, 2\pi)$.

2. For any $X \in \mathbf{X}$ and $Y \in \mathcal{N}_6(X)$ the angle $\Phi^*_{XY} \stackrel{\text{def}}{=} \min_{Z \in \mathcal{N}_6(Y)} |\Phi_{XYZ}|$ is uniformly distributed in $[0, \frac{\pi}{6})$ (see Figure 3.5 (d)), and is independent from $\Phi^*_{X'Y'}$ for any $(X', Y') \neq (X, Y)$.

Let's clarify the relation between this definition and the background stimuli of Figure 3.4. Property 1 corresponds exactly to the rule we used to define the background elements' orientations in our arrays of patches (see Section 3.2.1). The relevance of property 2 needs more justification. As explained in Section 3.2.1, we built each stimulus so that N elements fit in it, that two elements were not too close to each other (being distant of at least r_b), and that there were no empty regions in the image. These requirements are somehow a converse to those of the well known problem consisting in filling a region with as many spheres as possible, given their radius r. Indeed, in our case we know the number of discs to fit into the square image, and we want to set their radius r_b in order to get the most homogeneous layout. The most compact way to fit discs in a given region is to lay them on a hexagonal lattice. That is why we set $r_b = 1.3 r_{hex}$ (see Section 3.2.1), the value $2r_{hex}$ corresponding to placing the elements on an exact lattice, while we wanted to allow some randomness in their coordinates. The resulting stimuli look like hexagonal lattices affected by some noise, in which property 2 approximately holds (see Figure 3.5 (a)).

3.6.3 Description of the algorithm

Input The input of the detection algorithm is a set $\mathbf{g} = \{(x_i, \theta_i)\}_{i=1\dots N}$, that models a Gabor array composed of N patches (see Section 3.6.2).

Step 1 List the N distances of each point x_i to its nearest neighbor, and define d_{avg} as the mean value of this list.

Step 2 Define an oriented graph $\gamma = (\mathbf{x}, \mathbf{e})$, $\mathbf{x}$ being the set of vertices of γ, and $\mathbf{e}$ the set of edges, defined as follows: $(x_i, x_j) \in \mathbf{e}$ if and only if x_j is one of the 6 nearest neighbors of x_i and $d(x_i, x_j) \leq 2d_{avg}$, where $d(x_i, x_j)$ is the Euclidean distance between the two points. We will denote by $\mathcal{N}_\gamma(x)$ the set of neighbors of x in γ; see Figure 3.5 (a).

Step 3 Set $\mathcal{C} = \emptyset$. Then for each point $x \in \mathbf{X}$, and for each neighbor $y \in \mathcal{N}_\gamma(x)$, initialize a chain $c = (x, y)$, and add c to $\mathcal{C}$. This initialization step is illustrated in Figures 3.5 (a) and (b).

Step 4 Expand each started chain c trying to keep it as rectilinear as possible. More precisely, denote by x and y the penultimate and last points of c. Set $z^* = \arg\min_{z \in \mathcal{N}_\gamma(y)} |\phi_{xyz}|$ the neighbor of y that minimizes $|\phi_{xyz}|$. Following the notations of Section 3.6.2, $|\phi_{xyz^*}| = \phi^*_{xy}$. If z^* is not already in c then add z^* at the end of c, add c to $\mathcal{C}$ and carry on expanding c as long as it contains less than $\sqrt{N}$ points. By this process, represented in Figures 3.5 (b), (c) and (d), we build all the chains to be tested as candidate alignments. Then the number of tests is approximately

$$N_T \stackrel{\text{def}}{=} 6 \times N \times \sqrt{N}. \tag{3.2}$$

N_T is actually an overestimation of the number of tests, since not all nodes in γ have six neighbours.

Step 5 Compute the Number of False Alarms (NFA) of each chain of $\mathcal{C}$ containing at least three points, as follows. For a given chain $c = (z_1, \dots, z_n)$ with $n \geq 3$, consider the variables $\omega_1, \dots, \omega_n$, as defined in Section 3.6.1 and illustrated in Figures 3.5 (e) and (f). Under the assumption that $\mathbf{g}$ is a realization of an *a contrario* array of Gabor patches, the orientations $\omega_1, \omega_2, \dots, \omega_n$ are samples of n independent random variables $\Omega_1, \Omega_2, \dots, \Omega_n$. The first and last angles Ω_1 and Ω_n are uniformly distributed in $[0, \frac{\pi}{2}]$, and a short development shows that $\Omega_2, \dots, \Omega_{n-1}$ have a cumulative distribution function $F(\omega) = \mathbb{P}(\Omega_2 \leq \omega) = \dots = \mathbb{P}(\Omega_{n-1} \leq \omega)$, defined by

$$F(\omega) = \begin{cases} \frac{12\omega^2}{\pi^2} & \text{if } 0 \leq \omega < \frac{\pi}{12} \\ \frac{12}{\pi^2}(\frac{\pi}{6}\omega - \frac{\pi^2}{12^2}) & \text{if } \frac{\pi}{12} \leq \omega < \frac{5\pi}{12} \\ \frac{12}{\pi^2}\left(\omega^2 - \frac{2\pi\omega}{3} + \frac{\pi^2}{6}\right) & \text{if } \frac{5\pi}{12} \leq \omega \leq \frac{\pi}{2}. \end{cases} \tag{3.3}$$

Note that for a random patch Z_i in a chain, the probability law of $\max(\alpha_{iL}, \alpha_{iR})$ depends on the angle $\Phi^*_{Z_{i-1}Z_i}$. The cumulative distribution function F is obtained by integrating over all possible values for $\Phi^*_{Z_{i-1}Z_i}$, according to the law hypothesized in the *a contrario* model. Noting

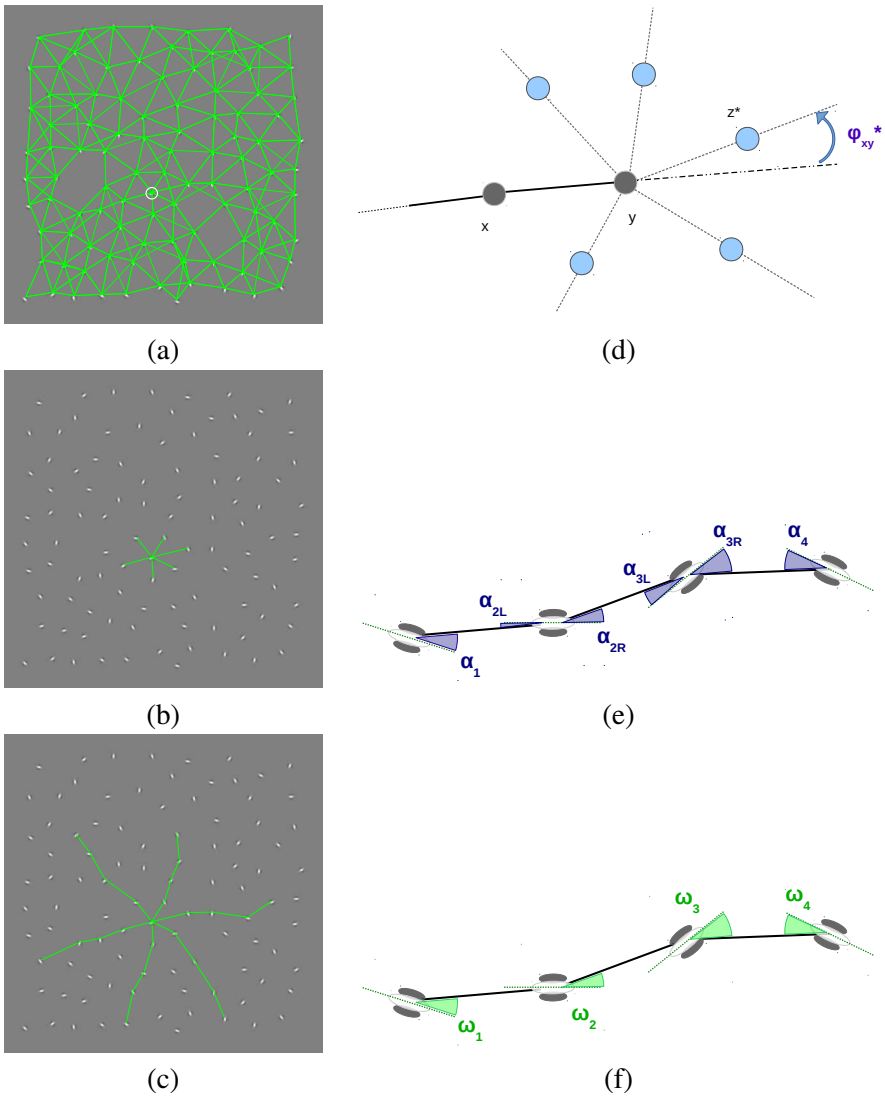
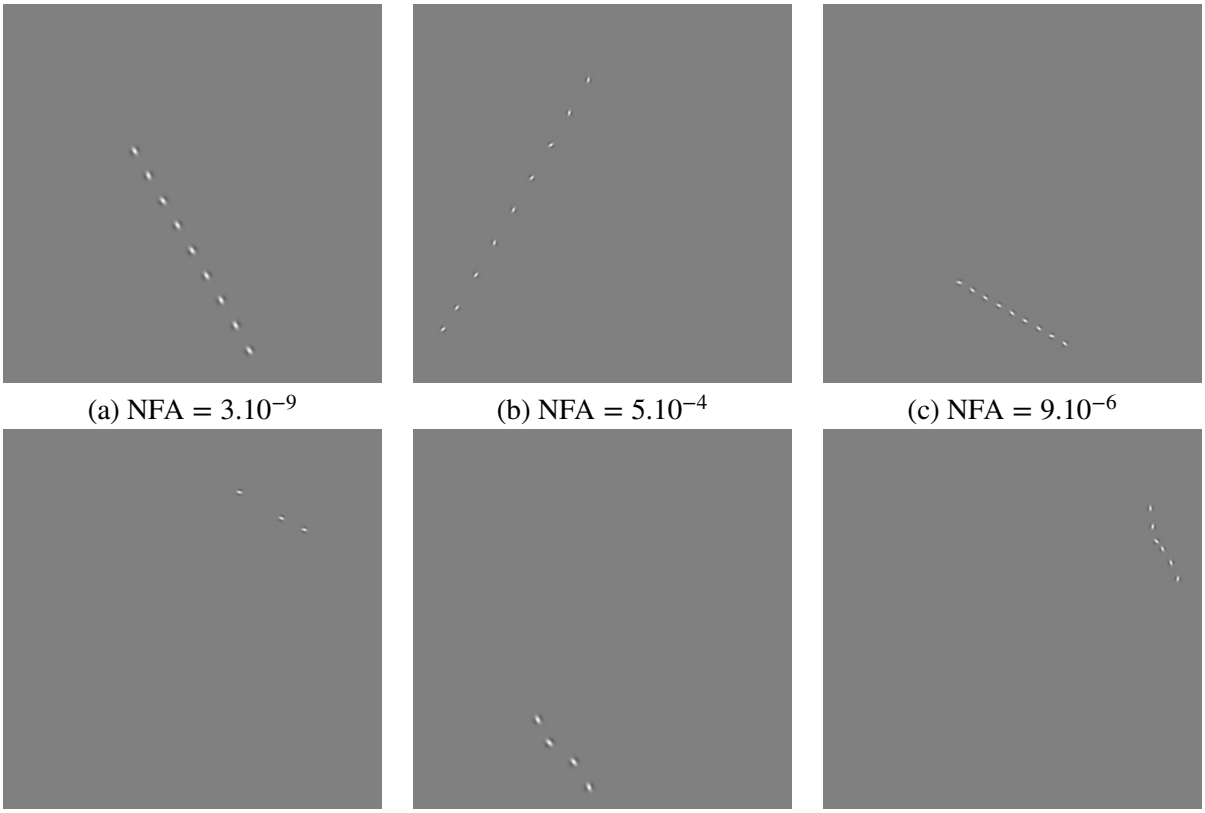


Figure 3.5: (a) On this array of 100 Gabor patches, we represent the associated graph γ (the orientations of the edges were omitted). In this graph, most of the points are linked to their 6 nearest neighbors, except when a neighbor is too remote, as it often happens for elements that are close to the image border. For a point $x \in \mathbf{x}$ (surrounded in white in (a)), a chain (x, y) is started for each $y \in \mathcal{N}_\gamma(x)$ in (b), and expanded in (c) into the most rectilinear possible chain. Picture (c) shows the result after 3 iterations of the expansion process. (d) Given two neighbors x and y, we look for the neighbor z of y that minimizes the absolute angle between $\overrightarrow{xy}$ and $\overrightarrow{yz}$. (e)-(f) The variables the algorithm measures when analyzing a chain containing $n = 4$ points.



(a) NFA = 3.10^{-9} (b) NFA = 5.10^{-4} (c) NFA = 9.10^{-6}

(d) NFA = 11.4 (e) NFA = 4.9 (f) NFA = 42.8

Figure 3.6: Detections by the algorithm in the arrays of Figures 3.1 (first row) and 3.4 (second row). The detected chains in the second row do not look like "good" alignments, and are actually difficult to see. This is consistent with the values of the NFA that are all larger than 1 and thus not considered as significant.

$\omega^* \stackrel{\text{def}}{=} \max\{\omega_1, \dots, \omega_n\}$ and $\Omega^* \stackrel{\text{def}}{=} \max\{\Omega_1, \dots, \Omega_n\}$, the probability that all Ω_i be less than the observed ω^* is

$$\mathbb{P}\left(\Omega^* \leq \omega^*\right) = \left(\frac{\pi\omega^*}{2}\right)^2 \times F(\omega^*)^{n-2}. \tag{3.4}$$

As usual in the *a contrario* theory, we get a natural definition of the NFA of chain c:

$$\begin{aligned} \text{NFA}(c) &\stackrel{\text{def}}{=} N_T \times \mathbb{P}\left(\Omega^* \leq \omega^*\right) \\ &= N_T \times \left(\frac{\pi\omega^*}{2}\right)^2 \times F(\omega^*)^{n-2}. \end{aligned} \tag{3.5}$$

Output The algorithm returns the lowest NFA and the chain that achieves it.

3.6.4 Property of the algorithm

As explained in Section 2.3, a small NFA characterizes an unexpected event, while common events have a large NFA. The NFA as defined in the previous section is a positive real number, upper bounded by N_T. The longer the chain and the smaller the angles ω_i, the smaller the NFA. Consequently, the algorithm is built to detect chains with little direction change and Gabor patches

roughly locally tangent to their chain. The following result is a consequence of Theorem 1 and, like the latter, it justifies the interpretation of the NFA as a measure of non-accidentalness.

Theorem 5. *Let $\varepsilon > 0$, and $\mathbf{G} = \{(X_i, \Theta_i)\}_{i=1\ldots N}$ an* a contrario *array of N Gabor patches. Then the expected number of chains c such that* $\mathrm{NFA}(c) \leq \varepsilon$*, is less than ε.*

Note that, since the variable Ω^* is continuous whatever the number of patches in a chain, this expected number could be equal to ε. The inequality is due to the overestimation of the number of tests.

The proof of Theorem 5 is exactly the same as for Theorem 2, with ω^* as measurement and the number n of elements as size variable. One of its implications is that there is on average less than one chain with NFA lower than 1 in an *a contrario* stimulus. The bottom row of Figure 3.6 shows that no such meaningful structure was found in some stimuli containing only background elements. Since our stimuli are close to the *a contrario* model, a target alignment with NFA < 1 is likely to be detected by the algorithm because it is unlikely that other chains in the background be assigned a lower NFA. Conversely, this is not guaranteed anymore for a target alignment with NFA ≥ 1.

3.7 Subjects compared to the algorithm

In this section we compare the subjects' detection performance to the algorithm described in Section 3.6.3. The question is whether a rigorous mathematical model of non-accidental alignments could match average human detection, and there is no better way to confront the theory to the subjects than by building an artificial observer on that theory.

Recall that each click made by a subject is associated to the nearest Gabor element in the image, and counted as a valid detection when that element belongs to the target. The same is done for the algorithm, by selecting the most central element among the returned ones (that is to say, the closest to the barycenter of the returned elements). This count permits to define the detection rate for the algorithm as well.

In Figure 3.7, each of the first three rows compares the subjects of one experiment to the algorithm. The fourth row presents the same data averaged over all three experiments. The plots of columns A and B show the detection rate as a function of the jitter level and the number of aligned elements respectively, whereas columns C and D display respectively the detection rate and the reaction time as functions of $\log_{10}(\mathrm{NFA})$ of the target alignment. In these two latter sets of plots, we grouped the data into ten bins with equal number of trials. We did not define any reaction time for the algorithm, thus only the subjects' curves appear in column D.

Tables 3.4, 3.5 and 3.6 show the results of a statistical analysis of the agreement between the subjects and the algorithm's behaviors, as well as among subjects. To compare the answers of two observers (whether two subjects or a subject and the algorithm), we adapted an idea exposed in [13]. Suppose two observers i and j were presented with the *same* set of n_c stimuli, and achieved respectively c_i and c_j correct detections in it. We want to measure the similarity in the distributions of their correct answers, beyond their detection rates. For example, imagine both observers detected $c_i = c_j = 5$ alignments out of $n_c = 10$ trial. They may have answered in exact opposite ways, if each observer detected in the trials where the other failed. In that case, we would like to point out how different these behaviors are, despite the identical detection rates. Conversely, if $c_i = 7$, $c_j = 4$ with observer i having answered correctly in the 4 trials where observer j succeeded as well, then observer i did better than j in 3 out of the 10 trials, but both observers agreed on the other 7 trials. Thus they show quite similar behaviors, no matter the difference in their detection rates. Formally,

Subject	1	2	3	4	5	6	7	8	9	10	11	12	Algorithm
1	-	-	-	-	-	-	-	-	-	-	(105, 1.2e-12)	-	(95, 3.5e-06)
2	-	-	-	-	(99, 2.3e-08)	(97, 1.3e-09)	(94, 6.8e-06)	-	-	-	-	(102, 4.9e-10)	(101, 1.8e-08)
3	-	-	-	-	-	-	-	(105, 6.3e-12)	(97, 3.9e-08)	(101, 9.4e-10)	-	-	(93, 4.9e-05)
4	-	-	-	-	-	-	-	-	-	-	-	-	(82, 0.0014)
5	-	-	-	-	-	(98, 4.3e-10)	(93, 1.3e-05)	-	-	-	-	(99, 1.5e-08)	(98, 3.7e-07)
6	-	-	-	-	-	-	(89, 5.6e-06)	-	-	-	-	(97, 1.8e-09)	(90, 2.7e-07)
7	-	-	-	-	-	-	-	-	-	-	-	(88, 0.00053)	(95, 3e-05)
8	-	-	-	-	-	-	-	-	(98, 4.7e-09)	(100, 7.5e-09)	-	-	(108, 1.2e-07)
9	-	-	-	-	-	-	-	-	-	(100, 1.1e-09)	-	-	(100, 7.1e-11)
10	-	-	-	-	-	-	-	-	-	-	-	-	(96, 1.1e-06)
11	-	-	-	-	-	-	-	-	-	-	-	-	(104, 1.5e-05)
12	-	-	-	-	-	-	-	-	-	-	-	-	(99, 6.7e-08)

Table 3.4: The values (k, p_k) are reported for Experiment I. The variable k is the number of identical responses given by two observers in a set of 126 trials, and the p-value p_k measures how significant it is to observe at least k identical responses. Recall that in Experiment I not all the subjects saw the same sequences of stimuli. Therefore, the cells are filled only for pairs of observers presented with the same stimuli.

Subject	1	2	3	4	5	6	7	8	9	10	Algorithm
1	-	(110, 2.3e-11)	(96, 0.00043)	(100, 5.3e-06)	(98, 0.0013)	(106, 5.9e-09)	(93, 2.4e-05)	(108, 9.2e-10)	(102, 1e-05)	(99, 1.1e-05)	(100, 2.9e-05)
2	-	-	(100, 0.00063)	(104, 3.3e-06)	(104, 0.003)	(102, 9.2e-05)	(93, 5.5e-05)	(112, 5.9e-09)	(106, 0.00016)	(95, 0.009)	(104, 8.6e-05)
3	-	-	-	(104, 7.3e-07)	(98, 0.04)	(102, 1.8e-05)	(91, 0.00026)	(102, 0.0002)	(106, 8.7e-06)	(95, 0.0029)	(104, 1.1e-05)
4	-	-	-	-	(98, 0.015)	(106, 5.7e-08)	(95, 5.2e-06)	(104, 6.6e-06)	(106, 1.6e-06)	(103, 6.8e-07)	(98, 0.0014)
5	-	-	-	-	-	(104, 0.00019)	(91, 0.00042)	(104, 0.01)	(112, 2.1e-05)	(101, 0.00052)	(104, 0.003)
6	-	-	-	-	-	-	(91, 0.00026)	(108, 1.4e-07)	(112, 6.8e-10)	(107, 8.4e-09)	(102, 9.2e-05)
7	-	-	-	-	-	-	-	(91, 0.00041)	(95, 3.9e-06)	(92, 8.3e-05)	(89, 0.0019)
8	-	-	-	-	-	-	-	-	(112, 2.5e-07)	(97, 0.0036)	(102, 0.0015)
9	-	-	-	-	-	-	-	-	-	(101, 0.00031)	(110, 9.5e-07)
10	-	-	-	-	-	-	-	-	-	-	(99, 0.00037)

Table 3.5: The values (k, p_k) for Experiment II. The variable k is the number of identical responses given by two observers in a set of 126 trials, and the p-value p_k measures how significant it is to observe at least k identical responses.

let $s_{i,1}, \dots, s_{i,n_c}$ and $s_{j,1}, \dots, s_{j,n_c}$ be the binaries answers (1 for a detection, 0 for a missed trial) given by observers i and j to the same n_c stimuli. Then $k = \sum_{l=1}^{n_c} s_{i,l}s_{j,l} + (1 - s_{i,l})(1 - s_{j,l})$ denotes the number of trials in which the observers agreed, i.e. both detected or both failed. We compare the observed number k to its random counterpart $K = \sum_{l=1}^{n_c} S_{i,l}S_{j,l} + (1 - S_{i,l})(1 - S_{j,l})$, assuming that the c_i and c_j detections are distributed randomly and independently (formally, the random vectors $(S_{i,1}, \dots, S_{i,n_c})$ and $(S_{j,1}, \dots, S_{j,n_c})$ are independent and uniformly distributed over $\{v \in \{0,1\}^{n_c}, \sum_{l=1}^{n_c} v_l = c_i\}$ and $\{v \in \{0,1\}^{n_c}, \sum_{l=1}^{n_c} v_l = c_j\}$ respectively). Under this basic assumption, we get the following law for K:

$$\mathbb{P}\left(K = k|n_c, c_i, c_j\right) = \begin{cases} \binom{c_i}{\frac{1}{2}(k-(n_c-\sigma_{ij}))}\binom{n_c-c_i}{\frac{1}{2}(n_c-\delta_{ij}-k)}\Big/\binom{n_c}{c_j} & \text{if } 0 \le k - |n_c - \sigma_{ij}| = 0 \mod (2) \text{ and } k \le n_c - |\delta_{ij}| \\ 0 & \text{otherwise} \end{cases} \tag{3.6}$$

where $\sigma_{ij} = c_i + c_j$ and $\delta_{ij} = c_i - c_j$. We denote by $p_k = \mathbb{P}\left(K \ge k|n_c, c_i, c_j\right) = \sum_{l=k}^{n_c} \mathbb{P}\left(K = l|n_c, c_i, c_j\right)$ the probability of observing at least k identical answers in the responses of the two observers. The number p_k is a p-value measuring if the observed k is significantly greater than expected under the basic assumption and, consequently, if there is significant agreement. In Tables 3.4, 3.5 and 3.6 we report, for each subject, the values (k, p_k) resulting from the comparison with the algorithm on the 126 trials seen by the subject. The same figures are given for each pair of subjects who saw the same sequence of images. For convenience, we call *agreement test* this analysis, that will also be applied to the results of Chapter 4.

Subject	1	2	3	4	5	6	7	8	9	10	Algorithm
1	-	(108, 1.9e-13)	(110, 6.1e-17)	(109, 2.3e-14)	(103, 1.5e-09)	(105, 3.9e-10)	(108, 1.9e-13)	(106, 1.7e-11)	(104, 4e-11)	(103, 1.1e-10)	(98, 1.6e-07)
2	-	-	(100, 6.9e-10)	(111, 4e-16)	(105, 3.9e-11)	(99, 3.5e-07)	(108, 1.1e-13)	(102, 2.1e-09)	(102, 4.2e-10)	(103, 7.9e-11)	(96, 8.3e-07)
3	-	-	-	(107, 2e-14)	(101, 1.8e-10)	(99, 1.1e-09)	(100, 6.9e-10)	(100, 7.6e-10)	(96, 6.9e-08)	(105, 4.7e-13)	(92, 4.8e-06)
4	-	-	-	-	(112, 1.4e-16)	(104, 1.8e-10)	(109, 1.4e-14)	(109, 3.3e-14)	(109, 7.9e-15)	(106, 7.8e-13)	(103, 1.5e-10)
5	-	-	-	-	-	(102, 1.6e-07)	(111, 1.8e-15)	(103, 3.3e-09)	(99, 4.5e-08)	(100, 8.8e-09)	(101, 9.1e-09)
6	-	-	-	-	-	-	(103, 1.8e-09)	(105, 1.6e-09)	(97, 1.1e-06)	(96, 1.9e-06)	(101, 2.8e-08)
7	-	-	-	-	-	-	-	(106, 7.2e-12)	(98, 5.7e-08)	(97, 1.3e-07)	(102, 7.6e-10)
8	-	-	-	-	-	-	-	-	(106, 2.8e-12)	(103, 1.5e-10)	(96, 2e-06)
9	-	-	-	-	-	-	-	-	-	(101, 7.5e-10)	(98, 5.7e-08)
10	-	-	-	-	-	-	-	-	-	-	(89, 0.0002)

Table 3.6: The values (k, p_k) for Experiment III. The variable k is the number of identical responses given by two observers in a set of 126 trials, and the p-value p_k measures how significant it is to observe at least k identical responses.

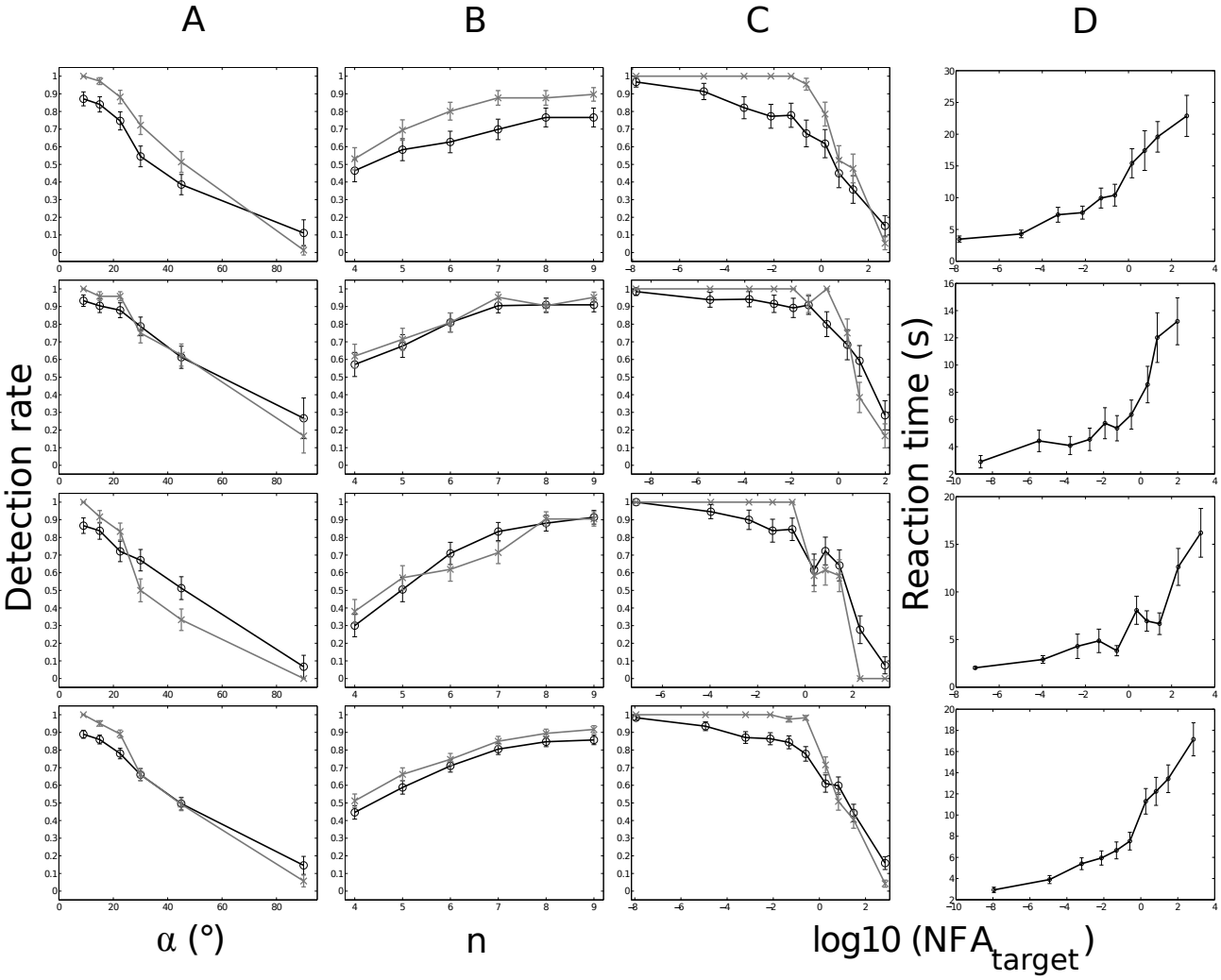


Figure 3.7: Comparison of the subjects of Experiments I-III to the *a contrario* detection algorithm. The solid black lines represent the subjects, whereas the gray solid lines represent the algorithm's results. The first, second and third row display the results of experiments I, II and III respectively, whereas the fourth row shows the same data averaged over all three experiments. The detection rates are plotted as functions of the jitter intensity α (column A), the number n of aligned elements (B) and the $\log_{10}$ of the target's NFA (C). The latter is also the x variable in column D's plots, the y axis representing the reaction time, in seconds.

3.8 Discussion

The first three columns of Figures 3.7 exhibit a strong similarity of shape and values between the subjects' detection curves and the algorithm's ones. The match is obviously not perfect, especially in the comparison with the subjects of Experiment I, in which the algorithm performs generally better than the average of the subjects. The four plots of column A show that the algorithm tends to be more accurate than the average of the subjects on moderately jittered stimuli, whereas subjects are generally better than the algorithm at detecting more jittered alignments, especially in the maximal jitter condition. Column C provides an additional interpretation of the differences between the subjects and the algorithm. In these graphics all the stimuli are ordered by the $\log_{10}(\text{NFA})$ of their target alignment. They show that the algorithm achieves better detection rates in stimuli with lower NFA (non-accidental alignments), and gets worse than the average of the subjects beyond a certain threshold, always larger than 1 (in the plot, $\log_{10}(\text{NFA}) = 0$).

A possible explanation of the discrepancies between the average of the subjects and the algorithm is the following. On the one hand, the algorithm may never miss the alignment that is most distinguishable from noise in a stimulus, because of the exhaustive search it carries out. This is what we expect from an ideal observer under the specified constraints. Subjects are less exhaustive and may lack attention sometimes (recall that they had only a minimal training). Besides, the algorithm takes as input the exact coordinates and orientations of the patches, whereas the subjects may have a lower visual accuracy. On the other hand, although we built the stimuli so as to avoid the detection of the alignments in absence of the orientation cue (see Sect. 3.2.1), it seems that in some cases the background elements could not completely mask the perfect rectilinearity and the regular spacing of the target elements. For example, with a little effort, one can find the hidden alignment by looking only at Figure 3.1 (e). According to the observed performance in the maximal jitter condition (Sect. 3.3-3.5), and the impressions that some of the subjects reported, we may infer that the regularity information was exploited by the subjects in some cases. On the contrary, the *a contrario* model described in Section 3.6.2 does not take into account this information. When the subjects are using information not available to the ideal observer, it is not surprising that they get better detection rates.

Nevertheless, the algorithm does perform quite similarly to the subjects. As tables 3.4, 3.5 and 3.6 show, its responses agree significantly with those of the subjects ($p \leq 0.003$). This agreement is actually comparable to the inter-subjects one. We have thus a statistical confirmation of the good match observed in Figure 3.7.

Furthermore, this similarity is not the result of the optimization of a set of parameters to get the best possible fit. It is the direct output of one unique non-parametric algorithm, that was compared to human subjects on every single trial. In short, while a model fit to the subjects results would be easily obtained by tuning the parameters of a parametric algorithm, here the fit between algorithm and subjects is absolute.

Another observation that must be pointed out is that the NFA of the target alignments is an accurate one-dimensional measure of the stimuli's difficulty. Indeed, in Figures 3.7, columns C and D show decreasing detection rates and increasing reaction times with respect to the NFA. According to Theorem 5, points with $\log_{10}(\text{NFA})$ lower than -2 in these plots represent stimuli in which the target alignment is expected to occur less than once every hundred *a contrario* array of Gabor patches. At the other end of the NFA axis, on the right of $\log_{10}(\text{NFA}) = 2$ for example, stand the images whose hidden alignment is an event that could occur, on average, more than a hundred times in a single *a contrario* Gabor array, and is thus indistinguishable from noise. Consequently, what columns C and D show in Figures 3.7, is a strong correlation between detectability and non-accidentalness.

Moreover, these results raise the hope that the *a contrario* framework might become a useful tool for other psychophysical studies of the non-accidentalness principle in perceptual grouping. It allows the translation of many parameters used to build stimuli, into one interpretable measure that permits comparing all the stimuli with each other. In Figure 3.7, columns C and D are an example of such flexibility, since they present the detection rates and reaction times as a function of the NFA, over a wide range of stimuli ruled by three parameters. In the present study we focused on the detection of straight contours in order to start with a model easy to define and to explain. But the same technique could be adapted to the study of symmetry, motion, or more general contour integration.

To illustrate this, we made a minor modification in our algorithm permitting to measure the non-accidentalness of stimuli proposed in [21] (Figure 3.8). In this experiment, for each image extracted from [21], the coordinates and orientations of the patches were computed by smoothing the image, selecting the pixels darker than 0.6 times its mean value, and by calculating the center and main direction of each resulting connected component. As described in [21], in this kind of arrays the Gabor patches are approximately set on a square grid. This is why we changed the number of visited neighbors by the algorithm from 6 to 4. We changed the number of tests N_T accordingly. The detection results again show a good fit between NFA and detectability of curves.

The limitations of the present work also require a discussion. First, in the experimental protocol most aspects of the viewing conditions were not normalized. What is more, the difference in the Gabor patches' spatial frequency between Experiment I on one side, and Experiments II and III on the other side, may have induced effects that we did not expect and that were not the scope of our study. Also, the fact that the subjects were only minimally trained may add variability among their response criteria, and overall reduce the detection rate in comparison to the algorithm. However, our intention was not to uncover new psychophysical phenomena, but to present a new interpretation of well known effects. In fact, despite the just mentioned limitations, the behavior observed in our experiments is consistent with previous studies on contour integration and the association field [21, 44, 13, 63].

Comparing our results to those of the latter studies may also be questionable, though, beyond the fact that we focused on straight contours. Indeed, contrary to the referred experiments, in ours, attention played an important part in the perceptual process, since the subjects were presented with the stimuli without time limitation. Thus, the good fit between the ideal observer's and the subjects' performance is explainable, as the subjects had all time to find the good continuations. Reproducing our experiments according to a preattentive paradigm could be the object of a future work.

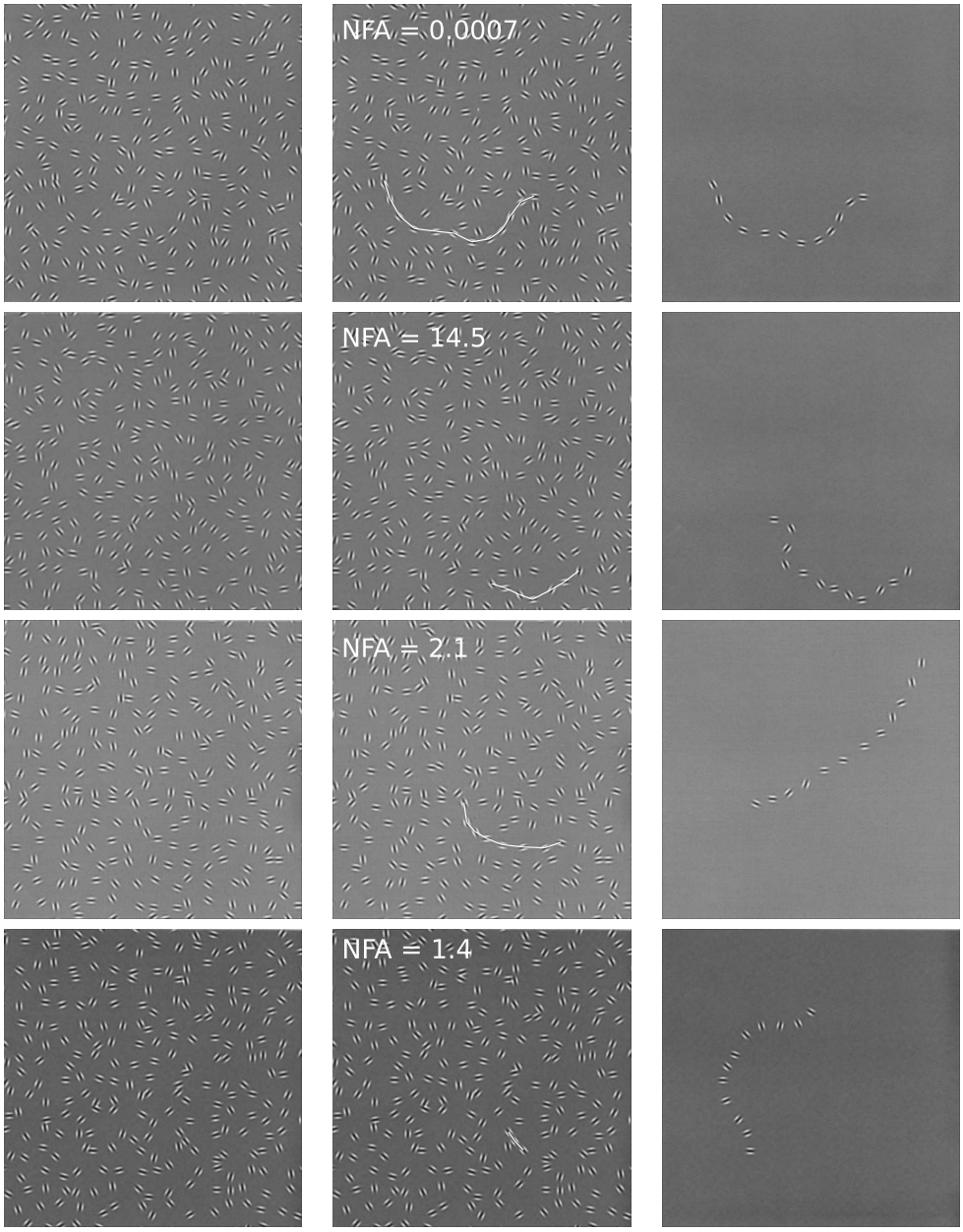


Figure 3.8: Some examples of contour detection by a slightly modified version of our algorithm. In the left hand are images extracted from [21] (from top to bottom row: Figs. 3, 6, 10 and 8). The coordinates and orientations of the patches were computed by smoothing the image, selecting the pixels darker than 0.6 times its mean value, and by calculating the center and main direction of each resulting connected component. As described in [21], in this kind of arrays, the Gabor patches are approximately set on a square grid. This is why we changed the number of visited neighbors by the algorithm from 6 to 4, and modified the number of tests N_T accordingly. The middle column displays the chain with smallest NFA found by our algorithm. In the right hand column we report the corresponding target path, extracted from [21] as well.

3.9 Conclusion

This chapter has presented an attempt to interpret and quantify a specific class of perceptual grouping, thanks to a mathematical model of the non-accidentalness principle. We exposed a way to apply the *a contrario* theory to implement a parameterless algorithm, adapted to the detection of a fixed Gestalt in a masking background. The strong correlation between the subjects' detection performance and the NFA, as well as the match between the algorithm and the subjects' curves, seem to argue in favor of the non-accidentalness principle as a way to interpret and predict the perceptual grouping of jittered alignments. We believe that the presented method could be adapted to address other questions of quantitative Gestalt. The next chapter addresses the case of general good continuation.

4 Measuring the salience of curves by their non-accidentalness

This chapter presents a formulation of the non-accidentalness principle for the perceptual grouping of oriented elements composing a smooth path. Good continuation is intended in a broad sense and is no longer restricted to alignments. As previously, a Number of False Alarms is derived from the *a contrario* theory to measure the non-accidentalness of linear structures, and is implemented in a detection algorithm. This ideal observer is then compared to human visual perception through two experiments of attentive contours detection in arrays of Gabor patches. Not only the NFA showed a strong correlation with detection performance, but we also found an equivalence between the detectability of the target stimuli and their NFA. Furthermore, the algorithm's answers matched accurately those of human subjects, on average as well as on a trial-by-trial basis.

4.1 Introduction

The qualitative definition of the good continuation grouping law emerged from the Gestalt school of psychology [90, 58, 37, 88, 89], and can be rephrased in Palmer's words: "All else being equal, elements that can be seen as smooth continuations of each other tend to be grouped together" [64, p. 259]. The challenge of building a quantitative formulation of good continuation has already been introduced at the beginning of the previous chapter. Since it was first enunciated [90], various aspects of this Gestalt law have been examined, including amodal completion and contour integration of basic oriented and non-oriented elements. In line with the works presented earlier in the dissertation, here we will focus again on the case of oriented elements, and more precisely on arrays of Gabor patches like those of Figure 4.2. For recent studies on the *a contrario* modeling of good continuation for non oriented elements, we refer to [47, 46].

In Psychophysics, methods using Gabor patches to investigate the perception of good continuation inherit from seminal works [21, 42], and have triggered results until recently [34, 33, 63, 32]. Among the many contributions reviewed in Chapter 3 [44, 63, 55, 54, 84, 74] and that inspired the experiments of the latter, we would like to get back to [21] and the still open quantitative question of the *association field*. As illustrated in Figure 4.1, this term refers to a tendency of our perception to group, or *associate*, two elements that are approximately oriented like the line joining them. Whenever the two elements deviate too much from this configuration, the association effect vanishes. When the association field is activated within the pairs of adjacent elements in a chain, the perception of a so called *path* occurs. The subsequent quantitative question is about the threshold angles that allow the association and the emergence of a path, and how they depend on the number

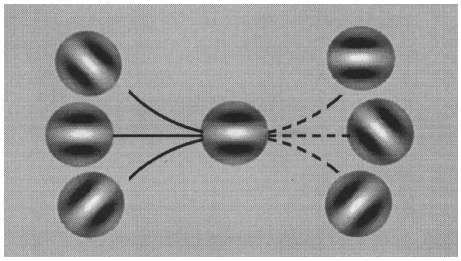


Figure 4.1: Principle of the association field. Figure extracted from [21].

of elements in the chain and the number of distractors.

In Chapter 3, we proposed a first attempt to address partially these questions, restricting ourselves to the case of alignments. In the short computational experiment described in Section 3.8, we stepped towards a generalization to smooth paths.

We shall now complete this step. As a mathematical formulation of non-accidentalness, the *a contrario* framework will be applied to interpret quantitatively the perception of general contours in arrays of random Gabor patches. This study will add to the set of already existing (mostly Bayesian) computational models that are compared to perception for this visual task, such as [97] and [13], commented in the previous chapter.

In Sections 4.2-4.4 we present our experimental setup, consisting in two versions of an attentive contour detection task. Although inspired by [21] and others, the design of our stimuli have several specificities. One is that we do not follow the two-steps procedure consisting in defining a target path first, and then filling the background of the stimulus with other patches. Instead, a unique process rules the definition of *all* the elements' positions in an image, and the target elements are subsequently chosen among this set of points. This prevents from undesired cues in the target path, and the need to mask them. Another specificity is that the target elements are chosen to minimize direction changes, starting from a random point and expanding the chain deterministically. Therefore, although we do not control the exact shape of the contours, the resulting paths are never too jagged. However, their visibility among the cluttered background is affected by three varying parameters: the number of target elements, the amount of angular noise deviating their orientations from the local tangents to the contour, and the total number of patches in the image.

Both experiments reported in Sections 4.3-4.4 use this kind of stimuli. However, they differ in their procedures and apparatus. In particular, the first experiment was run on a web interface on different computer devices, whereas the second experiment was set in a unique controlled environment. This was a way to assess the reliability of the method used in the first experiment (as well as in the experiments of the previous chapter), despite the variability in its experimental conditions.

In Section 4.5, we detail our model for the detection of paths in good-continuation. Contrary to other computational approaches of curve detection [100, 24, 97], ours is not related to biological considerations. Our proposal is to investigate the links between the salience of target curves and their non-accidentalness in a background model. If accurate predictions can be made based on the non-accidentalness principle, this might give evidence for a relevant cognitive mechanism. However, the *a contrario* detection algorithm developed here is not intended to reflect any realistic neural process.

Our model stands rather at a quantitative Gestalt level. To predict the salience of curves, it relies on the good-continuation *and* proximity laws. Indeed, although in our stimuli and many similar ones [21, 42, 63, 55, 84], the homogeneous spatial distribution of the patches is meant to minimize the role of proximity in the contour integration task [88], this grouping law might actually

still play an important part. The distribution of the distances between neighbors is concentrated in a narrow interval $[d_{\min}, d_{\max}]$. Still, a smooth path composed of elements oriented in good continuation, might be hard to see if all (or most) pairs of adjacent points in the path are distant of about $d_{\max}$. Conversely, if these distances are as small as $d_{\min}$, the contour may be visible even with more jittered orientations. Concerning the good continuation law, it is often characterized by the cocircularity relationship between oriented elements [66]. In our detection algorithm, we don't use explicitly this criterion, despite its relevance. Instead, we require each element in the path to be oriented tangentially to the contour. Given the considered smooth paths, in practice this criterion is an approximation of cocircularity, but it has the convenient property of handling only independent variables.

Considering good continuation and proximity raises the issue of cue combination, that is, how to merge the orientation and distance observations to infer the salience of a contour. To address it, we assume that all variables are independent, and focus on a unique quantity: the sum of these variables, after their appropriate normalization.

The resulting algorithm provides a characterization of our set of stimuli by their level of non-accidentalness, that we measure by the NFA. This enables us to evaluate, in Sections 4.6 and 4.7, how non-accidentalness correlates with detection performance, and if an *a contrario* algorithm mimics, and therefore can predict, the subjects behavior on average. Finally, Section 4.8 summarizes our conclusions.

4.2 Methods

4.2.1 Stimuli

Design of a stimulus In all the following experiments, subjects were shown arrays of Gabor patches of 500×500 pixels, like those represented in the first row of Figure 4.2. Once again, they were generated with the software GERT (v1.1) [7]. Like in the previous chapters, an array contained a certain number N of patches, each of which represented a symmetrical Gabor function, characterized by its position in the image and its orientation $\theta \in [0°, 360°]$ and described by

$$\forall (x, y) \in \mathbb{R}^2, G_{f,\theta,\sigma}(x, y) = \frac{1}{2}\left(1 + e^{-\frac{x^2+y^2}{2\sigma^2}} \cos\left(2\pi f (y\cos\theta - x\sin\theta)\right)\right). \tag{4.1}$$

The spatial frequency was set to $f = 0.12$ cycles per pixel, the space constant to $\sigma = \frac{1}{4f}$. Each array included a target path formed by n patches. In the whole chapter, these n patches will be referred as *target elements*, and we will call *background elements* the other $N - n$ Gabor patches of a stimulus. As previously, in addition to N and n, a stimulus was characterized by an *angular jitter* $\alpha \in [0°, 90°]$. These three parameters N, n and α, were used to generate our stimuli in four steps.

1. *Setting the N patches' coordinates.* The method applied here is similar to the one used to set the coordinates of the background elements in Chapter 3. All N elements are randomly placed but respect a minimal distance $d_{\min}$ from each other and from the edges of the image. This value $d_{\min}$ is set as a function of N and the size of the image, so that all N elements fit into the image and fill it homogeneously, avoiding clusters and large empty regions (Figure 4.2, second row). For an $I \times I$ square image, the most compact configuration fulfilling these requirements consists in a hexagonal lattice, in which we would have $d_{\min} = \frac{2(I-2d_{\min})}{\sqrt{2N\sqrt{3}}} = \frac{2I}{\sqrt{2N\sqrt{3}}+4}$ pixels. In order to achieve a random layout for the elements' coor-

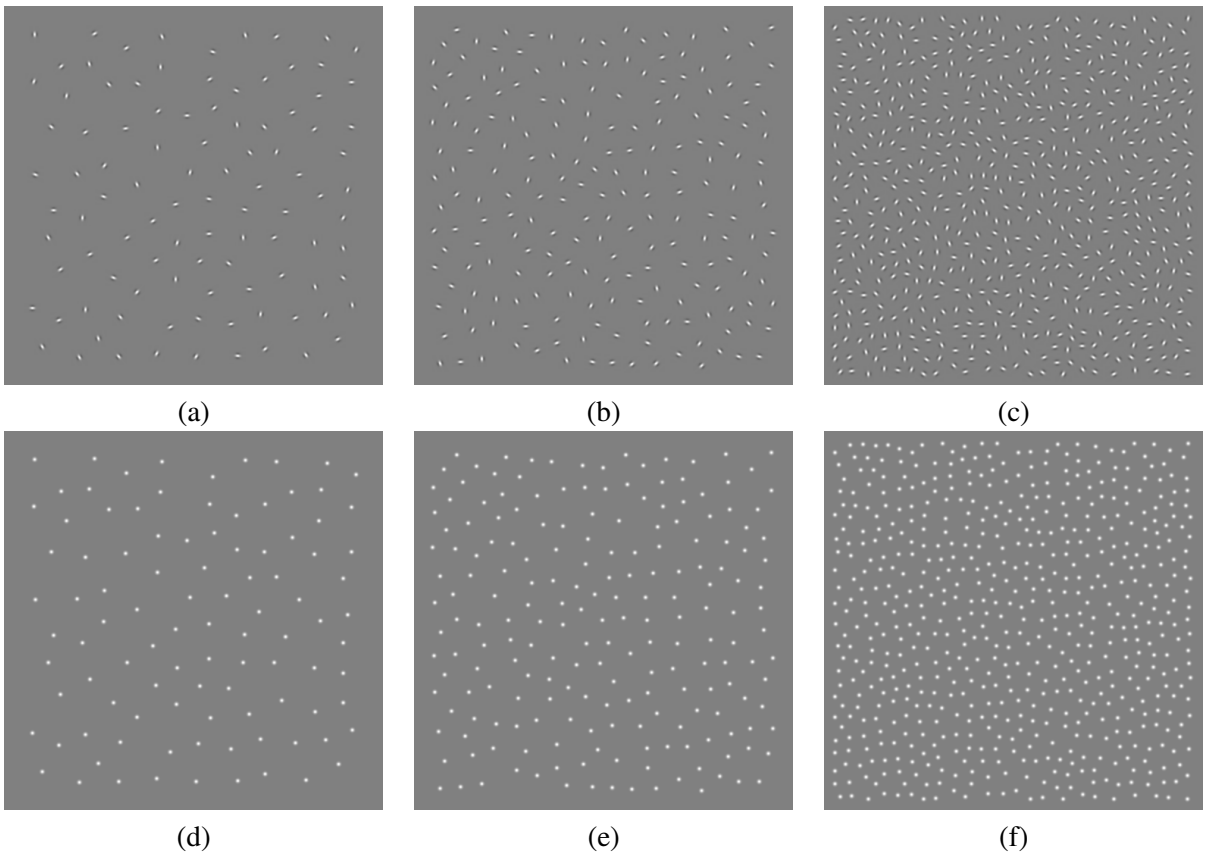


Figure 4.2: First row: three arrays of N Gabor patches with long, moderately jittered contours; $N = 100$ in (a), $N = 200$ in (b) and $N = 600$ in (c). Images (d), (e) and (f) show dots representing the coordinates of the Gabor elements of images (a), (b) and (c) respectively.

dinates (and not a regular lattice), we need a smaller value $d_{\min} = \frac{\lambda(I-2d_{\min})}{\sqrt{2N\sqrt{3}}} = \frac{\lambda I}{\sqrt{2N\sqrt{3}+2\lambda}}$ pixels with $\lambda < 2$. For our stimuli, we set $\lambda = 1.5$, and $I = 500$ pixels.

2. *Choosing the n target elements.* Given a set of N points $\mathbf{x} = \{x_1, \dots, x_N\}$ sampled as explained above, the target path is defined as a smooth chain of n adjacent points.

   Therefore, we first need to specify a neighborhood system on $\mathbf{x}$, that we choose as the 6-nearest neighbors oriented graph, from which we remove some long edges. To do so, we apply the same method as the one implemented in the detection algorithm of Chapter 3 (Section 3.6.3): first, we list the N distances of each point x_i to its nearest neighbor in $\mathbf{x}$, and note d_{avg} the mean value of this list. Then the modified oriented graph γ is obtained by removing from the initial 6-nearest neighbors graph, the edges that are longer than $2d_{avg}$. We denote by $\mathcal{N}_\gamma(x)$ the set of neighbors of x in γ.

   Now, we define a target path by following the edges of γ, so that each element in the path (except the first one) is a neighbor of its antecedent. What is more, we want the path to be smooth, which is why we try to minimize the direction changes between adjacent edges. More precisely, for a fixed integer $n \geq 4$ we build a target path composed of n elements through the following procedure:

   - Choose randomly in $\mathbf{x}$ the first element of the path, denoted x_1.
   - Set x_2, the second element of the path, as the point in $\mathcal{N}_\gamma(x_1)$ which is the closest to the image's center.
   - While the path $x_1, \dots, x_k$ counts less than n elements ($k < n$) set x_{k+1} as the point in $\mathcal{N}_\gamma(x_k)$ which maximizes $\cos(\overrightarrow{x_{k-1}x_k}, \overrightarrow{x_k x_{k+1}})$[1].

3. *Setting the target elements' orientations according to α.* Each element in the target path is assigned an orientation corresponding to the direction of the local tangent to the path, affected by some random angular noise.

   Formally we define, for each x_i in the path, a reference direction τ_i^{ref} approximating the discrete tangent to the path at this point:

   - $\tau_1^{ref} = \arg(\overrightarrow{x_1 x_2})$
   - $\tau_i^{ref} = \arg(\overrightarrow{x_{i-1} x_{i+1}})$ for $2 \leq i \leq n-1$
   - $\tau_n^{ref} = \arg(\overrightarrow{x_{n-1} x_n})$

   where the $\arg(\vec{u})$ denotes the angle between $\vec{u}$ and the abscissa axis (that is, for a vector $\vec{u} = (u_1, u_2)$, $\arg(\vec{u}) = \arctan(\frac{u_2}{u1})$ if $u_1 \neq 0$, and 90° otherwise).

   Then given the fixed angle $\alpha \in [0°, 90°]$, each orientation θ_i is randomly sampled from a uniform distribution on $[\tau_i^{ref} - \alpha, \tau_i^{ref} + \alpha]$.

4. *Setting the background elements' orientations.* They were uniformly and independently sampled in $[0, 180°)$.

[1] In all the chapter, we will note $\cos(\vec{u}, \vec{v})$ the cosine of the angle between vectors $\vec{u}$ and $\vec{v}$.

Sequences of stimuli During an experimental session, a subject was presented with a sequence of precomputed stimuli. Such a sequence was defined by the number N of Gabor patches per image, the same for all the images in the sequence, and by varying the number n of target elements and α, the angular jitter.

In what we call the *main sequences*, the parameters n and α took their values in $\{4, 5, \ldots, 12\}$ and $\{9°, 18°, 36°, 45°, 72°, 90°\}$ respectively, defining $9 \times 6 = 54$ conditions. There were four stimuli per condition, so that each of the main sequences counted $m = 216$ stimuli.

In the *training sequences*, there were 9 conditions ($n \in \{4, 7, 10\}$ and $\alpha \in \{9°, 36°, 90°\}$) and three stimuli per condition, totaling 27 stimuli.

Three values of N were considered in our experiments: $N = 100, 200$ and 600. For each value of N, three main sequences labeled "A", "B" and "C", and one training sequence were computed. Therefore, we had a stimuli dataset composed of 9 main sequences of 216 stimuli each and 3 training sequences of 27 stimuli each.

4.2.2 Experimental sessions

During an experimental session, every stimulus of *one* sequence was presented once to the subject, on a computer screen. Therefore, we call *main* (resp. *training*) session an experimental session in which the subject was presented with one of the main (resp. training) sequences.

The 216 trials of a main session were divided into four blocks of 54, so that every condition described in Section 4.2.1 appeared once in a block. The order of presentation of the conditions within a block was shuffled. There was a pause after each of the first three blocks. On the other hand, a training session contained only one block of 27 trials, in which the 27 conditions of a training sequence were presented in random order.

In each trial, the subjects knew the stimulus included a target path and were asked to click on a Gabor element they perceived as part of it. In case they did not see any relevant path they were advised to give their best guess. As in Chapter 3, we wanted the task to be *attentive*. Therefore, the stimulus remained on the screen until an answer was given. The elapsed time between the presentation of the stimulus and the click, that we call *reaction time*, was measured for each trial. Subjects were advised to spend no more than about 20 seconds per stimulus. They were free to follow this qualitative suggestion or not and no time information was provided during the sessions. The coordinates of the click were recorded as well. After the subject's click, a transition gray image was displayed for 500 ms and then the next stimulus appeared. Subjects were never given any feedback on the exactness of their answers.

4.3 Experiment I

4.3.1 Subjects

Nine subjects, including four females, took the experiment. They were all between 20 and 30 years old, and had normal or corrected to normal vision. All the subjects were naive to the purpose of the experiment.

4.3.2 Procedure

In this experiment, each subject performed four experimental sessions: the training session and then the three main sessions associated to one single value of N (100, 200 or 600). For each value of N, three subjects took the four corresponding sessions, with at least one representative of each gender (two females for $N = 200$ and one female for each of the two other conditions).

Subject		1		2		3		All	
		d.r.	r.t.	d.r.	r.t.	d.r.	r.t.	d.r.	r.t.
	bl. 1	28	4.3	28	10	43	6.4	**33**	**7**
	bl. 2	31	5.1	31	11	43	8.2	**35**	**8.1**
A	**bl. 3**	24	5.5	39	8.8	46	6.6	**36**	**7**
	bl. 4	33	5.5	35	12	37	10	**35**	**9.2**
	all bl.	**29**	**5.1**	**33**	**11**	**42**	**7.9**	**35**	**7.8**
	bl. 1	33	4.8	37	5.8	48	11	**40**	**7**
	bl. 2	26	4.3	35	6	44	8	**35**	**6.1**
B	**bl. 3**	28	5.8	22	6.6	41	7.6	**30**	**6.7**
	bl. 4	35	5.7	26	6.2	35	8.3	**32**	**6.7**
	all bl.	**31**	**5.1**	**30**	**6.2**	**42**	**8.6**	**34**	**6.6**
	bl. 1	39	6.1	30	8.6	43	4.7	**37**	**6.4**
	bl. 2	28	4.7	37	11	35	5	**33**	**6.8**
C	**bl. 3**	28	4.1	35	8.7	35	4.3	**33**	**5.7**
	bl. 4	20	4.5	30	9.4	30	5.4	**27**	**6.4**
	all bl.	**29**	**4.8**	**33**	**9.3**	**36**	**4.9**	**32**	**6.3**
Tot.		**29**	**5**	**32**	**8.7**	**40**	**7.1**	**34**	**6.9**

Table 4.1: Detection rates (noted d.r.) in percents and average reaction times (noted r. t.) in seconds for $N = 100$ in Experiment I. The results are reported for each subject, and each block of each main session. In the last two columns and the last line we also report average values computed over the different subjects and different sequences.

The different sessions were taken over several days, depending on the availability of the subjects. The only constraint was to complete any started session, without taking other pauses than those provided by the experimental program. No feedback was given to the subjects on their responses.

4.3.3 Apparatus

Like in the experiments presented in Chapter 3, the monitor properties, the distance to the screen and the illumination conditions could vary from one subject to another. Indeed, various computers were used at different places along the sessions. However we can estimate the average distance between the subjects' eyes and the monitor was about 70 cm, and the average size of a pixel was approximately 0.02 cm $\times$ 0.02 cm, so that the 500×500 images subtended about $8° \times 8°$ of visual angle in average. In this experiment, the sessions were set up to work on a web browser, and they were taken by selected subjects under our supervision. They are accessible at the following link `http://bit.ly/ac_curves`.

4.3.4 Results

Only the results of the main sessions are reported here. Each click made by a subject was associated to the nearest Gabor element in the image, and counted as a valid detection when that element belonged to the target path. Therefore, the detection rate was computed like in Chapter 3. Figure 4.3 displays the results of Experiment I. Plots (a) and (b) show the detection rate as a function of the jitter level α and the number n of target elements, whereas (c) and (d) plot the reaction time as

Subject		**1**		**2**		**3**		**All**	
		d.r.	r.t.	d.r.	r.t.	d.r.	r.t.	d.r.	r.t.
	bl. 1	20	8.7	22	7	26	11	**23**	**9**
	bl. 2	19	7.1	24	9.6	33	15	**25**	**11**
A	**bl. 3**	22	6	28	6.6	28	15	**26**	**9.3**
	bl. 4	26	10	33	6.1	33	13	**31**	**9.9**
	all bl.	**22**	**8**	**27**	**7.3**	**30**	**14**	**26**	**9.7**
	bl. 1	20	5.3	35	4.5	26	20	**27**	**9.9**
	bl. 2	20	6.4	22	7.3	19	22	**20**	**12**
B	**bl. 3**	24	5.9	20	5.8	35	17	**27**	**9.5**
	bl. 4	28	5.9	28	5.7	31	21	**29**	**11**
	all bl.	**23**	**5.9**	**26**	**5.9**	**28**	**20**	**26**	**11**
	bl. 1	24	8.8	35	7.7	46	19	**35**	**12**
	bl. 2	22	12	39	5.6	31	11	**31**	**9.4**
C	**bl. 3**	22	8	26	6.2	35	16	**28**	**10**
	bl. 4	20	7.7	24	7.3	30	18	**25**	**11**
	all bl.	**22**	**9**	**31**	**6.7**	**36**	**16**	**30**	**11**
Tot.		**22**	**7.6**	**28**	**6.6**	**31**	**17**	**27**	**10**

Table 4.2: Detection rates (noted d.r.) in percents and average reaction times (noted r. t.) in seconds for $N = 200$ in Experiment I. The results are reported for each subject, and each block of each main session. In the last two columns and the last line we also report average values computed over the different subjects and different sequences.

Subject		1		2		3		All	
		d.r.	r.t.	d.r.	r.t.	d.r.	r.t.	d.r.	r.t.
	bl. 1	9.3	4.2	11	4.4	9.3	4.8	**9.9**	**4.5**
	bl. 2	13	6.6	13	3.9	11	5.1	**12**	**5.2**
A	**bl. 3**	15	4	15	4.8	19	4.7	**16**	**4.5**
	bl. 4	11	5.6	24	4.3	13	5.7	**16**	**5.2**
	all bl.	**12**	**5.1**	**16**	**4.4**	**13**	**5.1**	**14**	**4.9**
	bl. 1	19	4.5	30	4.6	19	6	**22**	**5**
	bl. 2	11	9.1	11	4.2	15	5.7	**12**	**6.4**
B	**bl. 3**	17	5	15	3.6	13	6.1	**15**	**4.9**
	bl. 4	13	4.1	13	2.9	11	5.5	**12**	**4.2**
	all bl.	**15**	**5.7**	**17**	**3.8**	**14**	**5.8**	**15**	**5.1**
	bl. 1	11	6.5	20	3.1	17	6	**16**	**5.2**
	bl. 2	20	6.6	11	3.8	13	7.5	**15**	**6**
C	**bl. 3**	22	5.8	15	2.6	20	5.3	**19**	**4.6**
	bl. 4	17	5.4	20	2.8	17	6.5	**18**	**4.9**
	all bl.	**18**	**6.1**	**17**	**3.1**	**17**	**6.3**	**17**	**5.2**
Tot.		**15**	**5.6**	**17**	**3.8**	**15**	**5.8**	**15**	**5**

Table 4.3: Detection rates (noted d.r.) in percents and average reaction times (noted r. t.) in seconds for $N = 600$ in Experiment I. The results are reported for each subject, and each block of each main session. In the last two columns and the last line we also report average values computed over the different subjects and different sequences.

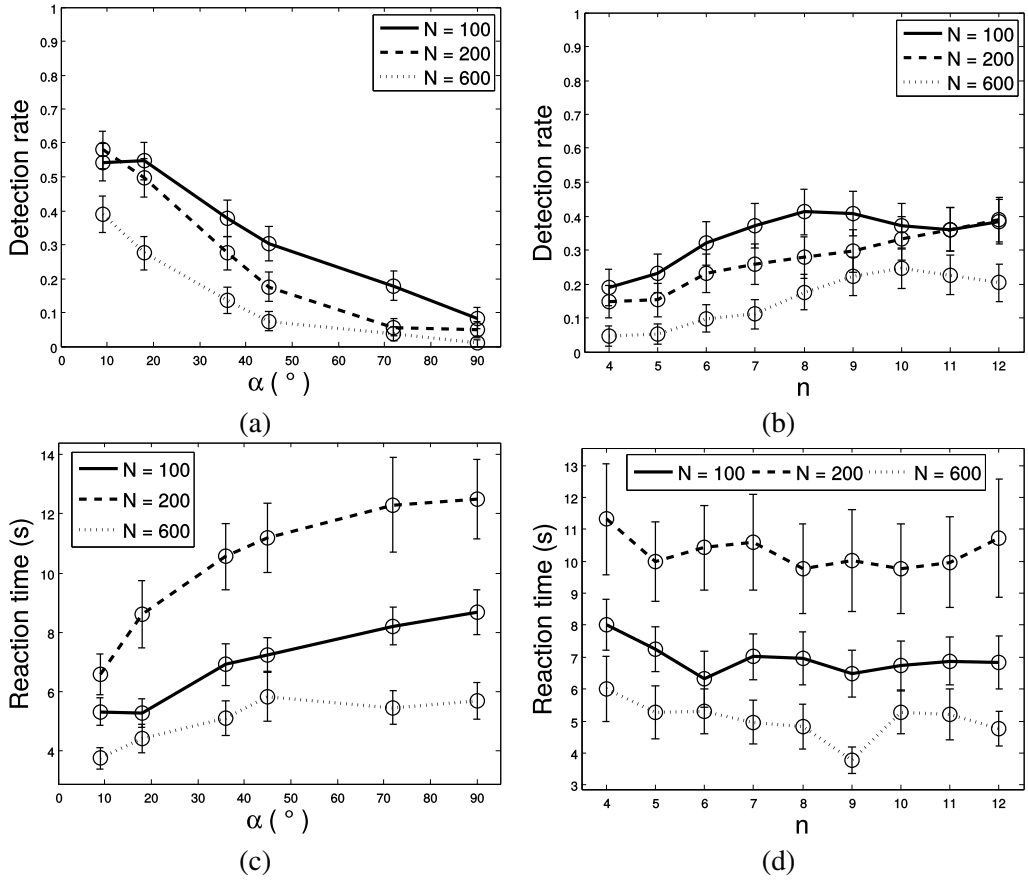


Figure 4.3: Results of the Experiment I (see also Tab. 4.1, 4.2 and 4.3). The subjects' detection rates (top row) and reaction times (bottom row) are plotted as functions of the jitter intensity α (left hand column) and the number n of aligned elements (right hand column). The error bars give 95 % confidence; they are defined as $[\overline{x} - 2\frac{\sigma}{\sqrt{q}}, \overline{x} + 2\frac{\sigma}{\sqrt{q}}]$, where $\overline{x}, \sigma$ and q are respectively the mean, standard deviation and number of trials of the corresponding condition.

a function of the same parameters (all the trials were counted; there was no attempt to remove outliers). The error bars give 95 % confidence; they are defined as $[\overline{x} - 2\frac{\sigma}{\sqrt{q}}, \overline{x} + 2\frac{\sigma}{\sqrt{q}}]$, where $\overline{x}, \sigma$ and q are respectively the mean, standard deviation and number of trials of the corresponding condition.

In Tables 4.1, 4.2 and 4.3 are reported detection rates and reaction times achieved by each subject of Experiment I, for $N = 100, 200$ and 600 respectively.

4.3.5 Discussion

As expected, Figures 4.3 (a) and (b) show that the detection rate is an increasing function of the number n of target elements, and drops with increasing jitter to reach chance level for the maximal jitter condition. (For a given N, chance level is $c_N = \frac{1}{9}\sum_{n=4}^{12}\frac{n}{N}$, yielding $c_{100} = 8\%$, $c_{200} = 4\%$ and $c_{600} = 1.33\%$). Additionally, for a fixed number of target elements and a fixed level of jitter, the detection difficulty increases with N.

Concerning reaction times, the subjects took longer to answer to stimuli with more jittered target paths, which was also expected, although this effect is less clear for $N = 600$. It was less

predictable, however, that reaction time would not vary significantly with n, and that the lowest reaction times were achieved for $N = 600$. The latter observation may be explained by a difference of strategy between subjects. All subjects in the $N = 600$ sessions seem to have dedicated roughly the same time to all stimuli, namely the time they needed to detect the path in easy conditions. On the contrary, for $N = 100$ and $N = 200$, subjects kept searching the target when they did not find it right away, especially Subject 2 for $N = 200$.

Finally, it is interesting to note the absence of a clear trade-off between reaction time and detection rate. Indeed, except for Subject 2 in the condition $N = 200$, who performed better but took also longer in responding, higher detection rates are not explained by longer reaction times.

4.4 Experiment II

4.4.1 Subjects

Seven subjects, including four females, took part in this experiment. They were aged between 21 and 39 and had normal or corrected to normal vision. All the subjects were naive to the purpose of the experiment, except one, named Subject 7 in Tables 4.5-4.7, and who collaborated to the present study.

4.4.2 Procedure

Contrary to Experiment I, here each subject tested the three conditions $N = 100$, $N = 200$ and $N = 600$. All subjects but Subject 7, took one training session and one main session for each condition. Subject 7 took all the training and main sessions. There were two reasons for such a procedure.

First, since we observed in Experiment I that there was no strong learning effect and that one sequence seemed enough to represent the behavior of a subject on a given condition, we preferred to test more different subjects for every value of N. This was a way to ensure that the results of Experiment I on a given condition were representative, and not determined by the peculiarity of few individuals.

Second, we wanted a non-naive subject as a reference, to measure an "asymptotic" performance in the case of a perfectly understood task.

As presented in Table 4.4, the experimental sequences were assigned to the naive subjects so that each sequence was seen by two of them, and every permutation in the order of the three conditions was represented once.

4.4.3 Apparatus

The sessions were done in a controlled environment, on a PsychToolBox interface. The images were displayed in a darkened room on a cathodic monitor of size $28.65 \times 21.49\ cm^2$, resolution 1024×768 pixels and 60 Hz refresh rate. The subject was placed $1m$ away from the screen, so that

Subjects	1	2	3	4	5	6
1st sequence	$N = 600$, A	$N = 200$, A	$N = 200$, A	$N = 100$, A	$N = 100$, A	$N = 600$, A
2nd sequence	$N = 100$, B	$N = 100$, B	$N = 600$, B	$N = 200$, B	$N = 600$, B	$N = 200$, B
3rd sequence	$N = 200$, C	$N = 600$, C	$N = 100$, C	$N = 600$, C	$N = 200$, C	$N = 100$, C

Table 4.4: Assignation of the different main sequences to the 6 naive subjects.

the stimuli subtended 8×8°, and each Gabor patch 0.24×0.24° of visual angle. In these conditions, the spatial frequency f was equal to 6.54 cycles per degree.

4.4.4 Results

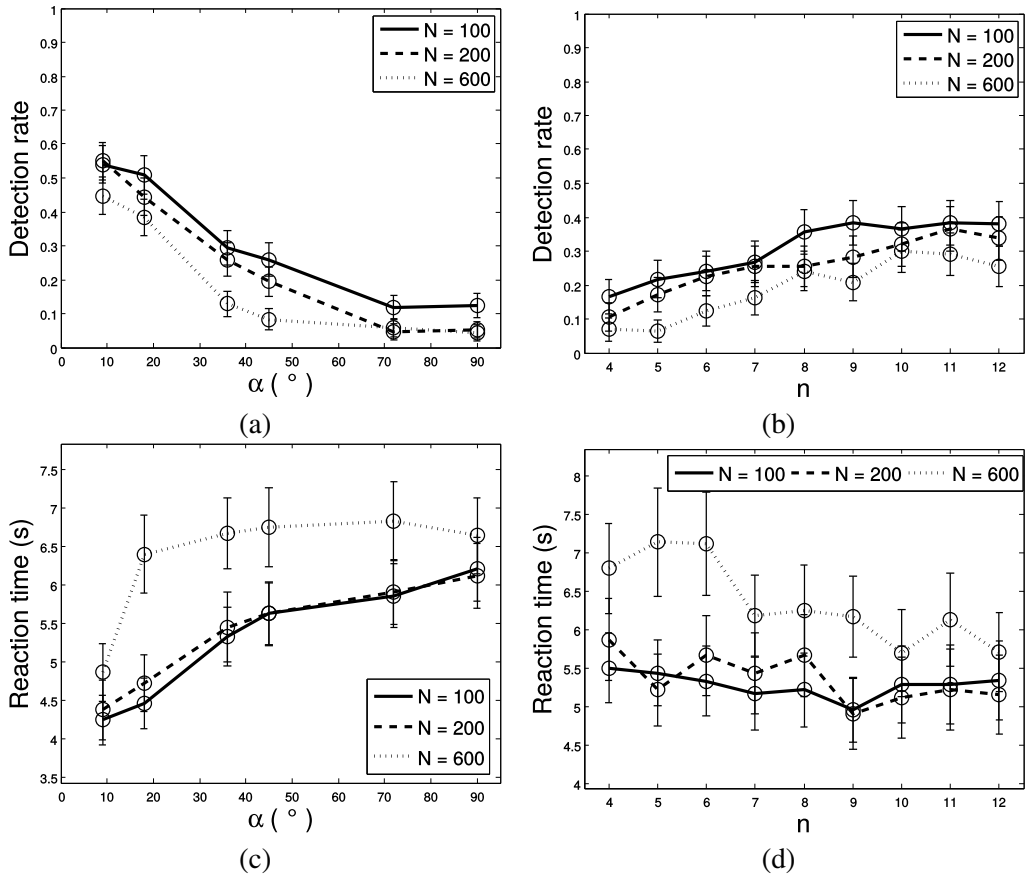


Figure 4.4: Results of the Experiment II (see also Tab. 4.5, 4.6 and 4.7). The subjects' detection rates (top row) and reaction times (bottom row) are plotted as functions of the jitter level α (left hand column) and the number n of aligned elements (right hand column). The error bars give 95 % confidence; they are defined as $[\overline{x} - 2\frac{\sigma}{\sqrt{q}}, \overline{x} + 2\frac{\sigma}{\sqrt{q}}]$, where $\overline{x}, \sigma$ and q are respectively the mean, standard deviation and number of trials of the corresponding condition.

Figure 4.4 displays the results of Experiment II, in the same way as for Experiment I.

In Tables 4.5, 4.6 and 4.7 are reported detection rates and reaction times achieved by each subject of Experiment II, for $N = 100, 200$ and 600 respectively.

4.4.5 Discussion

A one-way analysis of variance (ANOVA) shows that the order in which the naive subjects took the main sessions had no significant effect on detection performance. For each value of N, we tested if the differences in the number of detections among the six subjects could be explained by the order in which the corresponding session was taken. Thus, for a given N, the six individual numbers of detections were labeled according to whether the subjects experimented the condition N during their first, second or third session. As a sum of $m = 216$ independent binary digits, an

Subject		1		2		3		4		5		6		7		All	
		d.r.	r.t.	d.r.	r.t.	d.r.	r.t.	d.r.	r.t.	d.r.	r.t.	d.r.	r.t.	d.r.	r.t.	d.r.	r.t.
	bl. 1	-	-	-	-	-	-	31	4.2	33	4.6	-	-	43	6.9	**36**	**5.2**
	bl. 2	-	-	-	-	-	-	44	2.8	28	4.8	-	-	41	5.8	**38**	**4.5**
A	**bl. 3**	-	-	-	-	-	-	35	2.5	28	4	-	-	31	6.6	**31**	**4.4**
	bl. 4	-	-	-	-	-	-	26	3	19	5.4	-	-	37	3.4	**27**	**3.9**
	all bl.	**-**	**-**	**-**	**-**	**-**	**-**	**34**	**3.1**	**27**	**4.7**	**-**	**-**	**38**	**5.7**	**33**	**4.5**
	bl. 1	37	5.7	26	7.5	-	-	-	-	-	-	-	-	54	5	**39**	**6.1**
	bl. 2	35	4.9	26	9.1	-	-	-	-	-	-	-	-	43	4.6	**35**	**6.2**
B	**bl. 3**	28	5.2	15	6.4	-	-	-	-	-	-	-	-	37	6.9	**27**	**6.2**
	bl. 4	39	5	26	7.7	-	-	-	-	-	-	-	-	41	5.2	**35**	**6**
	all bl.	**35**	**5.2**	**23**	**7.7**	**-**	**-**	**-**	**-**	**-**	**-**	**-**	**-**	**44**	**5.4**	**34**	**6.1**
	bl. 1	-	-	-	-	46	6.9	-	-	-	-	26	3.7	46	5	**40**	**5.2**
	bl. 2	-	-	-	-	28	6.4	-	-	-	-	7.4	5.4	39	5.1	**25**	**5.6**
C	**bl. 3**	-	-	-	-	31	7.4	-	-	-	-	13	3.7	37	4.9	**27**	**5.3**
	bl. 4	-	-	-	-	35	6.4	-	-	-	-	28	3.2	33	5.2	**32**	**4.9**
	all bl.	**-**	**-**	**-**	**-**	**35**	**6.8**	**-**	**-**	**-**	**-**	**19**	**4**	**39**	**5**	**31**	**5.3**
Tot.		**35**	**5.2**	**23**	**7.7**	**35**	**6.8**	**34**	**3.1**	**27**	**4.7**	**19**	**4**	**40**	**5.4**	**33**	**5.3**

Table 4.5: Detection rates (noted d.r.) in percents and average reaction times (noted r. t.) in seconds for $N = 100$ in Experiment II. The results are reported for each subject, and each block of each main session. In the last two columns and the last line we also report average values computed over the different subjects.

Subject		1		2		3		4		5		6		7		All	
		d.r.	r.t.	d.r.	r.t.	d.r.	r.t.	d.r.	r.t.	d.r.	r.t.	d.r.	r.t.	d.r.	r.t.	d.r.	r.t.
	bl. 1	-	-	20	11	20	7.6	-	-	-	-	-	-	37	6.3	**26**	**8.5**
	bl. 2	-	-	17	7.8	13	5.4	-	-	-	-	-	-	37	6.7	**22**	**6.7**
A	**bl. 3**	-	-	24	4.9	20	5.9	-	-	-	-	-	-	28	5.7	**24**	**5.5**
	bl. 4	-	-	26	8.7	24	5.9	-	-	-	-	-	-	48	5.6	**33**	**6.7**
	all bl.	**-**	**-**	**22**	**8.2**	**19**	**6.2**	**-**	**-**	**-**	**-**	**-**	**-**	**38**	**6.1**	**26**	**6.8**
	bl. 1	-	-	-	-	-	-	35	3.5	-	-	15	4.7	37	4.4	**29**	**4.2**
	bl. 2	-	-	-	-	-	-	24	3.4	-	-	7.4	5.1	30	3.6	**20**	**4**
B	**bl. 3**	-	-	-	-	-	-	20	3.7	-	-	13	3.8	26	4.1	**20**	**3.9**
	bl. 4	-	-	-	-	-	-	22	4.5	-	-	24	4.2	31	3.8	**26**	**4.2**
	all bl.	**-**	**-**	**-**	**-**	**-**	**-**	**25**	**3.8**	**-**	**-**	**15**	**4.4**	**31**	**4**	**24**	**4.1**
	bl. 1	35	7.4	-	-	-	-	-	-	28	5	-	-	33	5	**32**	**5.8**
	bl. 2	24	6	-	-	-	-	-	-	26	2.9	-	-	37	4	**29**	**4.3**
C	**bl. 3**	30	7.8	-	-	-	-	-	-	24	3.5	-	-	43	4.9	**32**	**5.4**
	bl. 4	24	6.5	-	-	-	-	-	-	19	4.1	-	-	30	5.2	**24**	**5.3**
	all bl.	**28**	**6.9**	**-**	**-**	**-**	**-**	**-**	**-**	**24**	**3.9**	**-**	**-**	**36**	**4.8**	**29**	**5.2**
Tot.		**28**	**6.9**	**22**	**8.2**	**19**	**6.2**	**25**	**3.8**	**24**	**3.9**	**15**	**4.4**	**35**	**4.9**	**26**	**5.4**

Table 4.6: Detection rates (noted d.r.) in percents and average reaction times (noted r. t.) in seconds for $N = 200$ in Experiment II. The results are reported for each subject, and each block of each main session. In the last two columns and the last line we also report average values computed over the different subjects.

Subject		1		2		3		4		5		6		7		All	
		d.r.	r.t.	d.r.	r.t.	d.r.	r.t.	d.r.	r.t.	d.r.	r.t.	d.r.	r.t.	d.r.	r.t.	d.r.	r.t.
	bl. 1	15	8	-	-	-	-	-	-	-	-	1.9	5	20	9.3	**12**	**7.5**
	bl. 2	20	8.8	-	-	-	-	-	-	-	-	1.9	4.9	26	9.2	**16**	**7.7**
A	**bl. 3**	24	6.8	-	-	-	-	-	-	-	-	3.7	7.3	20	8.5	**16**	**7.5**
	bl. 4	28	9	-	-	-	-	-	-	-	-	7.4	5.8	24	9.9	**20**	**8.2**
	all bl.	**22**	**8.2**	**-**	**-**	**-**	**-**	**-**	**-**	**-**	**-**	**3.7**	**5.7**	**23**	**9.2**	**16**	**7.7**
	bl. 1	-	-	-	-	20	6.1	-	-	15	3.5	-	-	28	7.1	**21**	**5.6**
	bl. 2	-	-	-	-	19	7.1	-	-	19	4.1	-	-	28	8	**22**	**6.4**
B	**bl. 3**	-	-	-	-	13	7.2	-	-	20	4.1	-	-	28	5.8	**20**	**5.7**
	bl. 4	-	-	-	-	9.3	6.1	-	-	11	3.8	-	-	19	7.7	**13**	**5.9**
	all bl.	**-**	**-**	**-**	**-**	**15**	**6.6**	**-**	**-**	**16**	**3.9**	**-**	**-**	**25**	**7.1**	**19**	**5.9**
	bl. 1	-	-	11	6.7	-	-	24	5.9	-	-	-	-	22	4.8	**19**	**5.8**
	bl. 2	-	-	24	6.1	-	-	15	4.4	-	-	-	-	17	6	**19**	**5.5**
C	**bl. 3**	-	-	11	4.8	-	-	20	4	-	-	-	-	24	6.2	**19**	**5**
	bl. 4	-	-	20	4.4	-	-	20	4.6	-	-	-	-	33	7.9	**25**	**5.6**
	all bl.	**-**	**-**	**17**	**5.5**	**-**	**-**	**20**	**4.7**	**-**	**-**	**-**	**-**	**24**	**6.2**	**20**	**5.5**
Tot.		**22**	**8.2**	**17**	**5.5**	**15**	**6.6**	**20**	**4.7**	**16**	**3.9**	**3.7**	**5.7**	**24**	**7.5**	**18**	**6.4**

Table 4.7: Detection rates (noted d.r.) in percents and average reaction times (noted r. t.) in seconds for $N = 600$ in Experiment II. The results are reported for each subject, and each block of each main session. In the last two columns and the last line we also report average values computed over the different subjects.

appropriate statistical model for the number of detections achieved by a subject in a session, is a binomial distribution $\mathcal{B}(m, p)$, where p is the expected proportion of good answers. In our case, the number m and the observed detection rates are large enough for the binomial distribution to be well approximated by a Gaussian one. Therefore, it was reasonable to test the hypothesis assuming that the six observed numbers of detections were independent samples of the Gaussian distribution $\mathcal{N}(mp, mp(1-p))$. Thus, a one-way analysis of variance is a suitable method for this test. Its results are reported in Table 4.8. For each N, the corresponding p-value is too large to reject the above hypothesis, showing no statistical evidence of a significant effect of the order in presentation.

The effects observed in Experiment II look very similar to those of Experiment I. The major difference with respect to the results of the previous experiment, concerns reaction times. Here, they are higher for $N = 600$ than for the two other conditions, but they are roughly the same for $N = 100$ and $N = 200$. This non-decreasing tendency of reaction times as a function of N is more consistent with what one would expect.

As far as detection rates are concerned, we found no significant difference between the results of Experiment II and those of Experiment I. Indeed, a one way ANOVA was performed for each main sequence, on the detection rates achieved by the subjects of Experiments I and II. Here again, this analysis was well suited for our data, for the same reasons we previously explained. Its results suggest that the difference of environment in the two experiments did not produce significant effects on detection performance (see Table 4.9). One explanation may be that the subjects' attention allowed by unrestricted response times in the procedures of both experiments, could compensate most negative effects of both environments. From now on, we will no longer distinguish between the two experiments in the presentation of the data.

N	100	200	600
$F(2,3)$	0.098	0.91	0.29
p	0.91	0.49	0.77

Table 4.8: Results of the one-way analysis of variance to test the effect of the presentation order, for each value of N. In the second row are reported the computed $F(df_1, df_2)$ values of a Fisher test for data concerning six individuals divided into three groups, yielding $df_1 = 3 - 1 = 2$ and $df_2 = 6 - 3 = 3$ as degrees of freedom. In the third row, the corresponding p-values are given.

	N = 100			**N = 200**			**N = 600**		
	A	B	C	A	B	C	A	B	C
$F(1,4)$	0.14	0.0043	0.055	0	0.16	0.0035	0.15	1.1	2.2
p	0.73	0.95	0.83	1	0.71	0.96	0.71	0.35	0.21

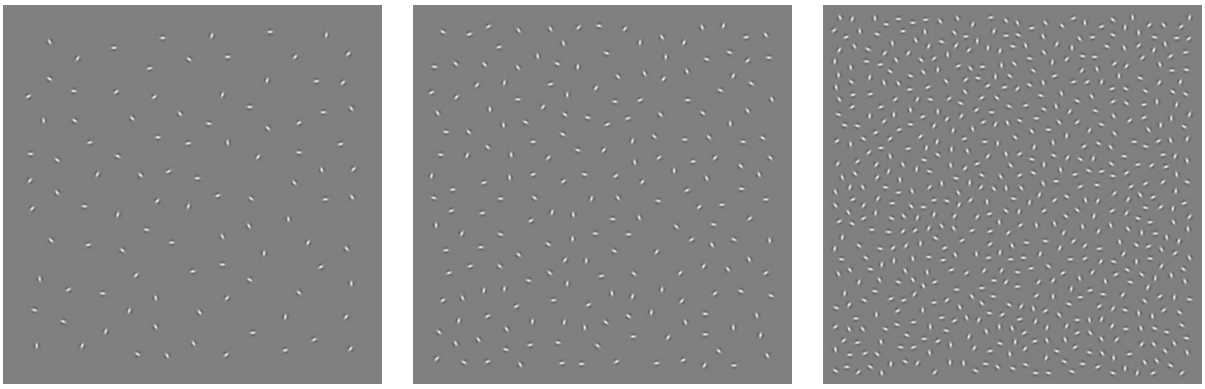
Table 4.9: Results of the one-way analysis of variance to test the effect of the difference of environment between Experiments I and II. In the second row are reported the computed $F(df_1, df_2)$ values of the Fisher test. Here we analyze, for each sequence, six individuals divided into two groups. Hence, the degrees of freedom are $df_1 = 2 - 1 = 1$ and $df_2 = 6 - 2 = 4$. In the third row, we report the corresponding p-values.

4.5 Model and algorithm

In the following subsections, we detail the geometric and *a contrario* models for our study.

4.5.1 The *a contrario* model

Figure 4.5 shows three arrays of Gabor patches with *only background elements* and thus no significant structure. This is what we want to model here. Formally, an array of N Gabor patches is seen as a set $\mathbf{g} = \{(x_i, \theta_i)\}_{i=1\ldots N}$, where $x_i \in [0,1]^2$ represents the coordinates of patch number i, and $\theta_i \in \mathbb{R}$ its orientation (in this section the variables representing angles will be expressed in radians). We note $\mathbf{x} \stackrel{\text{def}}{=} \{x_1, \ldots, x_N\}$. Then we define an *a contrario* array of Gabor patches as a random set $\mathbf{G} = \{(X_i, \Theta_i)\}_{i=1\ldots N}$ verifying the following properties:



(a) (b) (c)

Figure 4.5: Three arrays of Gabor patches containing only background elements and illustrating the *a contrario* model.

1. The random variables $\Theta_1, \dots, \Theta_N$ are independent and uniformly distributed in $[0, 2\pi)$.

2. The coordinates $X_1, \dots, X_N$ are independent from the Θ_is and follow the random process that was used to generate the stimuli, described in Section 4.2.1: X_1 follows a uniform distribution on $[0, 1]^2$, and for $k \geq 2$, X_k follows a uniform distribution on $[0, 1]^2 \backslash \cup_{j=1}^{k-1} \mathcal{D}(X_j, d_{\min}(N))$, with $d_{\min}(N) \stackrel{\text{def}}{=} \frac{1.5}{\sqrt{2N\sqrt{3}+3}}$ and $\mathcal{D}(c, r)$ the open disc of center c and radius r.

We note $\Gamma_6(\mathbf{X})$ the 6-nearest neighbors oriented graph on $\mathbf{X}$: (X_i, X_j) is an edge of $\Gamma_6(\mathbf{X})$ if and only if X_j is one of the six nearest neighbors of X_i. The lengths of the edges of $\Gamma_6(\mathbf{X})$ are identically distributed random variables, and we note F_ℓ their cumulative distribution function.

4.5.2 The geometric model

The family of tests Given a set $\mathbf{g} = \{(x_i, \theta_i)\}_{i=1\dots N}$ of Gabor patches, we need to define the family of tuples $((x_1, \theta_1), \dots, (x_n, \theta_n))$ of $\mathbf{g}$ that will be tested as possible regular curves.

This definition only takes into account the coordinates $\mathbf{x} = \{x_1, \dots, x_N\}$ of the Gabor elements, and focuses on smooth paths, that is to say chains of elements that minimize the direction changes between adjacent edges.

Analogously to Section 4.5.1, $\gamma_6(\mathbf{x})$ denotes the 6-nearest neighbors oriented graph on $\mathbf{x}$, and for any $x \in \mathbf{x}$, we note $\mathcal{N}_6(x)$ the neighbours of x in $\gamma_6(\mathbf{x})$, that is to say the 6 nearest neighbours of x. The family of tests T is defined by the set of all tuples $(x_1, x_2, \dots, x_n)$ such that

- $3 \leq n \leq \sqrt{N}$
- $\forall k \in \{1, \dots, n-1\}, x_{k+1} \in \mathcal{N}_6(x_k)$
- $x_{k+1} = \arg\max_{y \in \mathcal{N}_6(x_k)} \cos(\overrightarrow{x_{k-1}x_k}, \overrightarrow{x_k y})$.

As we will see in Section 4.5.3, the family T can be exhaustively covered by initiating, at each node x_i, 6 chains (x_i, y_j), with $y_j \in \mathcal{N}_6(x_i), \forall j = 1, \dots, 6$, and then expanding these chains along the edges of $\gamma_6(\mathbf{x})$ minimizing changes in direction, as long as they count less than $\sqrt{N}$ elements. This algorithm allows to count that the number of tests in T is

$$N_T = 6 \times N \times (\lfloor \sqrt{N} \rfloor - 2) \tag{4.2}$$

where $\lfloor \cdot \rfloor$ denotes the lower integer part.

The ideal structure For a given a tuple $((x_1, \theta_1), \dots, (x_n, \theta_n))$ of n Gabor patches, we consider the variables $\alpha_1, \dots, \alpha_n, \ell_1, \dots, \ell_{n-1}$, illustrated in Figure 4.6. For each i, ℓ_i is simply the length of segment $[x_i, x_{i+1}]$, and α_i is the absolute angle between the orientation θ_i of Gabor patch i and the reference direction τ_i^{ref} defined like in Section 4.2.1:

- $\tau_1^{ref} = \arg(\overrightarrow{x_1 x_2})$
- $\tau_i^{ref} = \arg(\overrightarrow{x_{i-1} x_{i+1}})$ for $2 \leq i \leq n-1$
- $\tau_n^{ref} = \arg(\overrightarrow{x_{n-1} x_n})$.

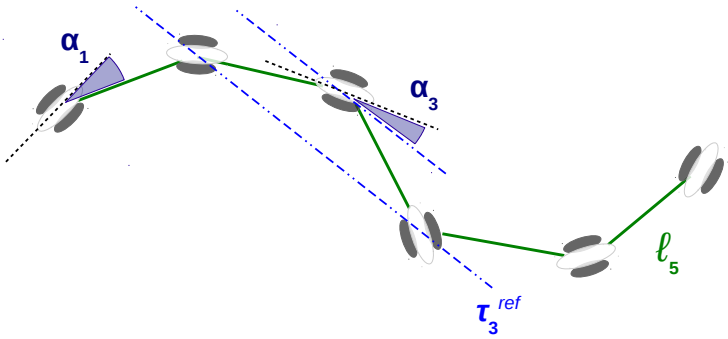


Figure 4.6: Illustration of the variables of interest in the definition of the ideal curve.

The ideal curve that will be our reference structure in the present study, is such a tuple of Gabor patches for which α_is are equal to zero and the ℓ_is are as small as possible. We measure the deviation of a general tuple from an ideal curve by the sum:

$$s = \sum_{i=1}^{n} \frac{2\alpha_i}{\pi} + \sum_{i=1}^{n-1} \tilde{F}_\ell(\ell_i). \tag{4.3}$$

where $\tilde{F}_\ell$ is an estimate of the cumulative distribution function introduced at the end of Section 4.5.1. Each term in s is an increasing function of α_i or ℓ_i, with values in $[0, 1]$. Therefore, a small value for s does correspond to a small deviation from the ideal structure as defined above.

More specifically, if $((x_1, \theta_1), \ldots, (x_n, \theta_n))$ belongs to the family of tests of an *a contrario* array of Gabor patches, then each α_i is the sample of a uniform distribution on $[0, \frac{\pi}{2}]$, and each ℓ_i is a sample of a law with F_ℓ as cumulative distribution function. If we assume that the lengths of two successive edges are independent variables (which is not true, more on this later), then s can be seen as the sum of $2n - 1$ realizations of independent variables uniformly distributed in $[0, 1]$.

4.5.3 Description of the algorithm

Input The input of the detection algorithm is a set $\mathbf{g} = \{(x_i, \theta_i)\}_{i=1\ldots N}$, that models a Gabor array composed of N patches.

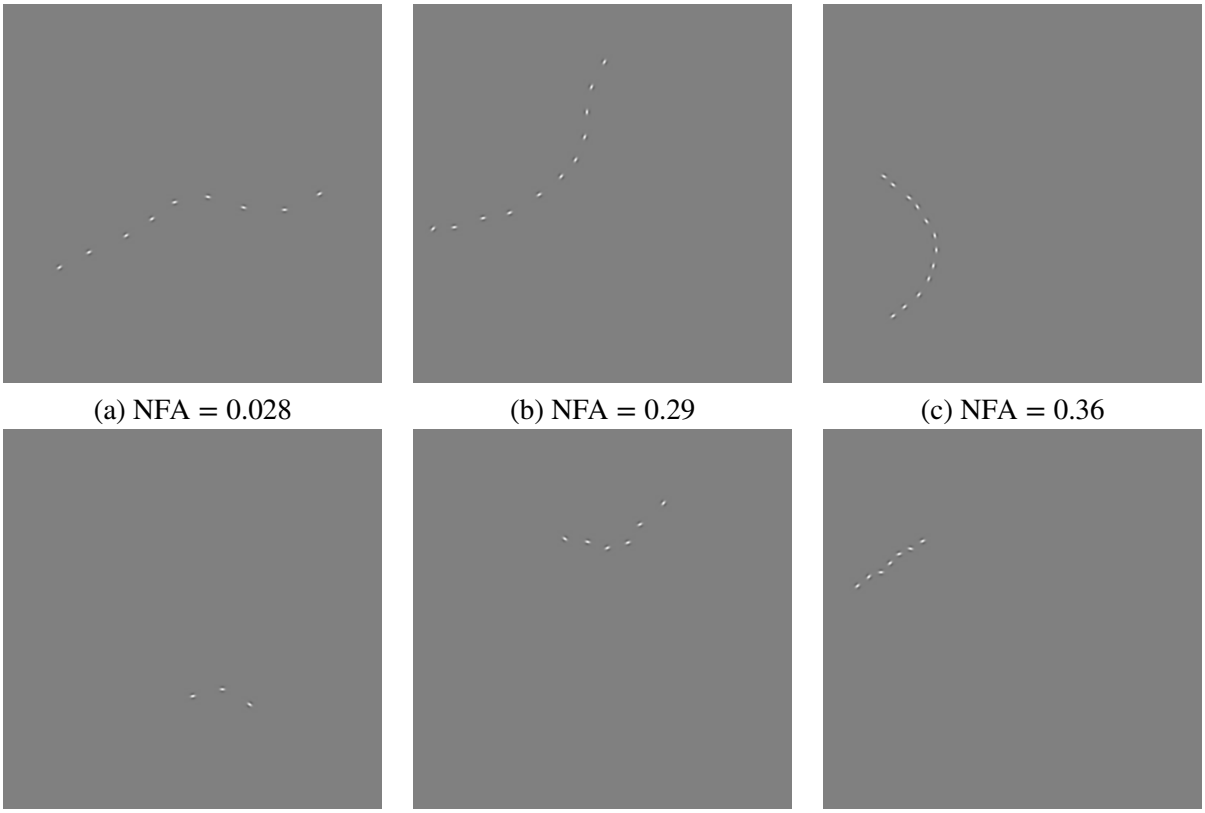
Step 1 Define $\gamma_6(\mathbf{x})$, the 6-nearest neighbors oriented graph on $\mathbf{x}$: (x_i, x_j) is an edge of $\Gamma_6(\mathbf{x})$ if and only if x_j is one of the six nearest neighbors of x_i.

Step 2 Compute $\tilde{F}_\ell$ the normalized cumulative histogram of the edges' lengths in $\gamma_6(\mathbf{x})$. To do so, note $\mathcal{E}$ the set of edges of $\gamma_6(\mathbf{x})$ and calculate for each edge $e \in \mathcal{E}$

$$\tilde{F}_\ell(||e||) \stackrel{\text{def}}{=} \frac{\#\{e' \in \mathcal{E}, ||e'|| \leq ||e||\}}{\#\mathcal{E}}, \tag{4.4}$$

which is the proportion of edges that are not longer than e. It is an estimation of the values $F_\ell(||e||)$ for the observed edges. The reason for this approximation is that a closed form expression of F_ℓ would be non trivial to compute, and would probably not improve significantly the algorithm.

Step 3 Set $\mathcal{C}_{init} = \emptyset$ and $T = \emptyset$. Then for each point $x \in \mathbf{X}$, and for each neighbor $y \in \mathcal{N}_6(x)$, initialize a chain $c = (x, y)$ and add c to $\mathcal{C}_{init}$.



(a) NFA = 0.028 (b) NFA = 0.29 (c) NFA = 0.36

(d) NFA = 2.4 (e) NFA = 3.0 (f) NFA = 2.2

Figure 4.7: Output of the algorithm for the stimuli of Figures 4.2 (first row) and 4.5 (second row).

Step 4 Expand each started chain $c \in C_{init}$ trying to keep it as smooth as possible. More precisely, denote by x and y the penultimate and last points of c. Set $z^* = \arg\max_{z \in \mathcal{N}_6(y)} \cos(\overrightarrow{xy}, \overrightarrow{yz})$. If z^* is not already in c then add z^* at the end of c, add c to T and carry on expanding c as long as it contains no more than $\sqrt{N}$ points.

Step 5 Compute the NFA of each chain of T, as follows. For a given chain $c = ((x_1, \theta_1), \dots, (x_n, \theta_n)) \in T$, consider the variables $\alpha_1, \dots, \alpha_n, \ell_1, \dots, \ell_{n-1}$, as defined in Section 4.5.2, and compute s, the sum of Eq. 4.3. Then we define the NFA of c as

$$\mathrm{NFA}(c) \stackrel{\text{def}}{=} N_T \cdot \mathbb{P}(S_{2n-1} \leq s) \tag{4.5}$$

where N_T is the number of tests introduced in Eq. 4.2 and S_{2n-1} is a random variable following the Irwin-Hall distribution of order $2n-1$, which is simply the law of the sum of $2n-1$ independent variables identically and uniformly distributed in $[0, 1]$.

Output The algorithm returns the lowest NFA and the chain that achieves it.

4.5.4 Properties of the algorithm

Recall that in general a small NFA should characterize an unexpected event in a background model, while common events are supposed to have a large NFA (see Section 2.3). Here, a NFA is used to

measure the non-accidentalness of chains of Gabor patches. The longer the chain and the smaller the α_is and the ℓ_is, the smaller the NFA. Consequently, the algorithm is built to detect chains with little direction change, elements that are close to each other, and Gabor patches roughly locally tangent to their chain.

Some examples of detections are presented in Figure 4.7. In the bottom row, we see that no structure with NFA < 1, was found in the stimuli of Figure 3.4, containing only background elements. Since our stimuli are close to the *a contrario* model, if the NFA defined in the algorithm follows the non-accidentalness property introduced in Section 2.3, a target path with NFA < 1 is likely to be detected by the algorithm because it is unlikely that other chains in the background be assigned a lower NFA. Conversely, this is not guaranteed anymore for a target curve with NFA ≥ 1.

Yet, we should discuss if the non-accidentalness property holds for this NFA. If the deviation measure s (eq. 4.3) followed an Irwin-Hall distribution, then the algorithm's NFA would show the latter property. Two approximations led us to model s as a sample of such a distribution. First, we took $\tilde{F}_\ell$ as an estimation of the actual cumulative distribution function F_ℓ, and thus considered the $\tilde{F}_\ell(\ell_i)$s as sampled from a uniform distribution in $[0, 1]$. Second, we assumed that the random lengths of consecutive edges of a path were independent. This second assumption is not accurate because consecutive edges link one same point to two of its 6 nearest neighbors. Since it is not trivial to predict theoretically the error introduced by these approximations, we tried to measure it empirically.

To do so, we ran the algorithm on three *background* sequences, that is to say sequences of stimuli containing only background elements. There was one sequence for each value of N, and every sequence was composed of 216 stimuli. In this special experiment, for every input stimulus, the algorithm returned the NFA values of all the tested chains. Then for each N we could average the results over the 216 corresponding stimuli, and therefore estimate the expected number of chains with NFA smaller than any $\varepsilon > 0$. The results are reported in Figure 4.8. In the first row, the average number of tests with NFA $\leq \varepsilon$, noted $A(\varepsilon)$, is plotted in blue as a function of ε, for the three values of N. If the non-accidentalness property of Theorem 2 held, $A(\varepsilon)$ should be very close to the identity $id(\varepsilon)$, plotted in red, since in the present case the measurement is a continuous variable. In the second row of Figure 4.8, we represent in blue the ratio $\rho(\varepsilon) \stackrel{\text{def}}{=} \frac{A(\varepsilon)}{\varepsilon}$, that is expected to be approximately equal to 1 if Theorem 2 holds. Here again, the reference value $\rho(\varepsilon) = 1$ is plotted in red.

As we can see, the error introduced by our approximations is not negligible. Indeed, the differences between the blue and the red lines follow the same pattern for the three values of N, and the number of samples used for these experiments is large enough to ensure that the empirical average values $A(\varepsilon)$ estimates reliably the expectation $\mathbb{E}\left[\sum_{C\in T} \mathbb{1}_{\mathrm{NFA}(C)\leq\varepsilon}\right]$. Therefore, the observed gap between the theory and our implementation comes from the inaccuracy of our assumptions. Another experiment suggested that the independence assumption is mainly responsible for this gap, since the approximation of F_ℓ by $\tilde{F}_\ell$ was sufficient to get empirically the results predicted by Theorem 2 when we simulated independent variables for the edges' lengths.

According to Figure 4.8, the algorithm seems to produce a limited number of false detections on average, with an empirical upper bound which appears to be

$$\mathbb{E}\left[\sum_{C\in T} \mathbb{1}_{\mathrm{NFA}(C)\leq\varepsilon}\right] \leq 2\varepsilon, \tag{4.6}$$

at least for the tested values of N. Note that this is only an empirical result, and we have no theoretical proof to confirm it. If it was mathematically established, then multiplying the NFA by 2 in Equation 4.5, would be sufficient to get the inequality of Theorem 2. However, what really

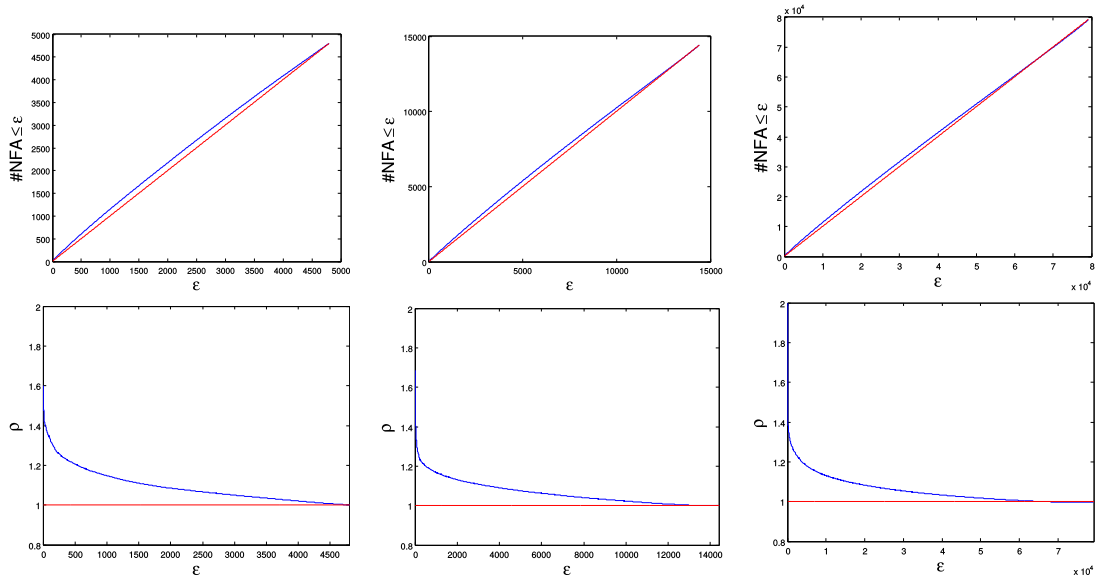


Figure 4.8: Property of the algorithm's NFA in background stimuli. In the first row, the average number of tests with NFA $\leq \varepsilon$, noted $A(\varepsilon)$, is plotted in blue as a function of ε, for the three values $N = 100, 200$ and 600 in the first, second and third column respectively. If Theorem 2 held, this function should be almost equal to, and not greater than the identity function $id(\varepsilon) = \varepsilon$, plotted in red. In the second row, we represent in blue the ratio $\rho(\varepsilon) \stackrel{\text{def}}{=} \frac{A(\varepsilon)}{\varepsilon}$, that is expected to be approximately equal to 1 if Theorem 2 holds. Here again, the reference value $\rho(\varepsilon) = 1$ is plotted in red.

matters is to be able to make the expected number of detections in the background model as small as desired. The inequality of Equation 4.6, if it were proved, would already guarantee this control on false alarms.

4.6 Subjects compared to the algorithm

In this section we compare the subjects' detection performance to the algorithm described in Section 4.5.3. Similarly to the previous chapter, we assess whether our algorithm matches the average human detections.

The comparison relies on the same kind of data as in Chapter 3. On the one hand, recall that the subjects' detection performance is measured by associating each click to the nearest Gabor element in the image, and counting it as a valid detection when that element belongs to the target (see Sections 4.3.4 and 4.4.4). On the other hand, for each stimulus, we take as the algorithm's clicked point the most central element among the returned ones (that is to say, the closest to the barycenter of the output elements), and we use the same criterion as for the subjects to decide whether the algorithm detected or missed the target. Figure 4.9 shows some examples of stimuli along with the algorithm's output and the subjects' clicks.

In Figure 4.11, each of the first three rows compares the subjects to the algorithm for one value of N. The fourth row presents the same data averaged over all experimental sessions. The plots of columns A and B show the detection rate as a function of the jitter level and the number of target elements respectively, whereas columns C and D display respectively the detection rate and the reaction time as functions of $\log_{10}$(NFA) of the target path. In these two latter sets of plots, we

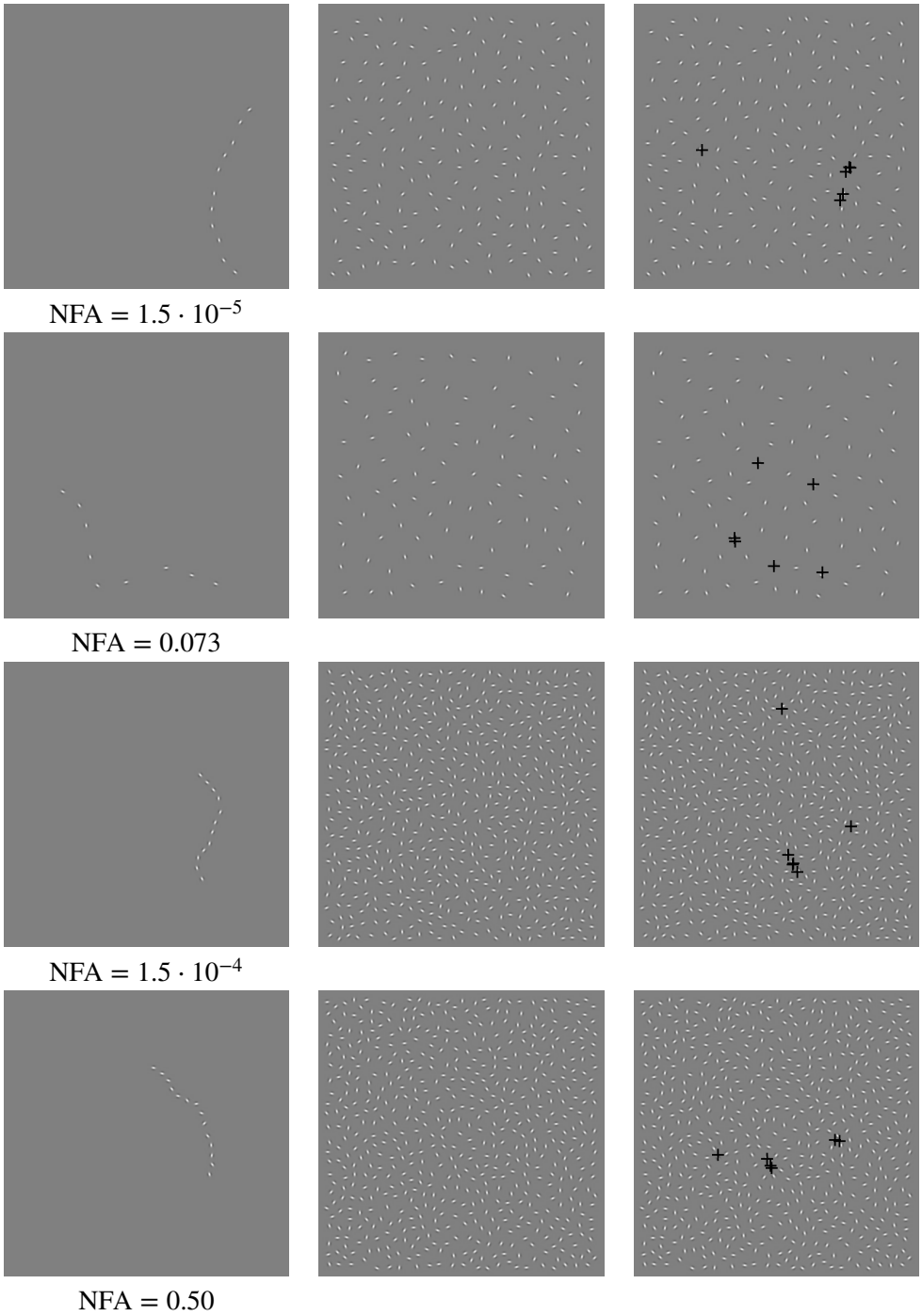


Figure 4.9: Comparison of detections by the algorithm and the clicked points by the subjects. Each row corresponds to one stimulus. In the central column are displayed the original stimulus; the left-hand column shows the detections by the algorithm, and the clicked points are represented by black crosses in the right-hand column. These examples illustrate some typical cases: when most subjects agree with the algorithm (first row); when they partly agree (second row); when the algorithm detects a salient structure but most subjects preferred another one, also salient (third row); when the algorithm detects a structure that is not salient (or only partly), thus contradicting perception (last row).

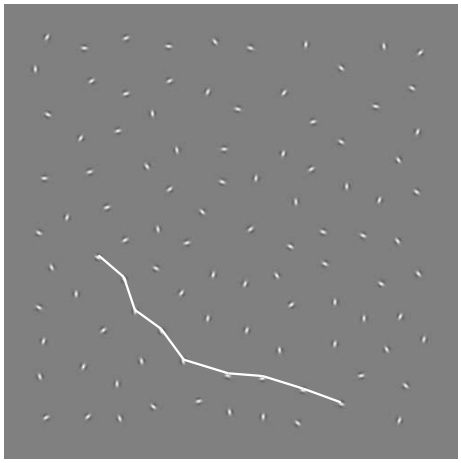


Figure 4.10: An example of chain that was not part of the family of paths tested by the algorithm, but that would have been detected otherwise, since NFA = 0.002 for this chain. Note that this path seems to be the one perceived by four out of six subjects, as the second row of Figure 4.9 shows.

α/n	4	5	6	7	8	9	10	11	12
9°	36 , 42	40 , 50	58 , 83	64 , 75	53 , 75	57 , 92	65 , 83	67 , 83	56 , 67
18°	26 , 0	32 , 25	50 , 50	56 , 58	68 , 83	61 , 83	62 , 58	65 , 92	69 , 1e+02
36°	17 , 8.3	35 , 33	24 , 17	31 , 8.3	51 , 50	46 , 67	22 , 58	29 , 42	57 , 58
45°	14 , 8.3	17 , 17	21 , 17	32 , 17	31 , 25	40 , 50	43 , 33	35 , 50	32 , 42
72°	14 , 0	9.7 , 8.3	12 , 8.3	6.9 , 0	14 , 0	28 , 25	21 , 0	17 , 8.3	14 , 25
90°	2.8 , 0	5.6 , 0	8.3 , 8.3	8.3 , 8.3	18 , 25	12 , 8.3	17 , 8.3	17 , 8.3	6.9 , 17

Table 4.10: Detection rates, in percents, achieved by the subjects and the algorithm in each (α, n) condition, for $N = 100$. In each case, the left hand digit is the average performance over all subjects, and the right hand one is the algorithm's detection rate. In all conditions, at least one subject achieved a higher rate than the algorithm, and at least one achieved a lower rate.

grouped the data into ten bins with equal number of trials. We did not define any reaction time for the algorithm, thus only the subjects' curves appear in column D. Finally, the subjects' curves of column C's first three plots are superimposed in Figure 4.12.

An "unfolded" version of Figure 4.11 plots is presented in Tables 4.10, 4.11, 4.12 and 4.13. Each of these four tables correspond to one row of Figure 4.11. For each value of N, level of angular jitter α and number n of target elements, the detection rates achieved on the corresponding stimuli are reported for the subjects and the algorithm (Tables 4.10-4.12). Like the fourth row of Figure 4.11, Table 4.13 sums up the same figures averaged over the three values of N. To refine the comparison, in Tables 4.10-4.12 we put the detection scores in bold font any time the algorithm's detection rate was strictly lower than the minimum, or strictly greater than the maximum of the subjects' rates in the same condition. This happened only twice over all 166 conditions: in $N = 600$, the cases $(\alpha = 9°, n = 6)$ and $(\alpha = 45°, n = 6)$. In the former case, a detection rate of 50% was reached by subjects in 4 experimental sessions.

The first statistical test we performed to check the apparent similarities observed in Figure 4.11, is a one-way ANOVA. For each value of N, 18 main sessions were taken by the subjects, and the algorithm was applied to the corresponding three main sequences. Thus, we could analyze 21 detection rates divided into 2 groups: on the one hand, the group of the 18 detection rates

α/n	4	5	6	7	8	9	10	11	12
9°	43 , 25	26 , 17	53 , 75	54 , 75	60 , 75	61 , 67	58 , 75	81 , 92	78 , 1e+02
18°	15 , 17	29 , 25	44 , 50	47 , 33	46 , 50	54 , 58	62 , 67	54 , 58	75 , 58
36°	5.6 , 0	19 , 8.3	18 , 8.3	33 , 17	33 , 33	38 , 33	24 , 42	32 , 42	40 , 33
45°	5.6 , 0	15 , 0	15 , 8.3	11 , 17	22 , 33	14 , 17	32 , 42	39 , 42	17 , 42
72°	6.9 , 0	4.2 , 8.3	4.2 , 0	2.8 , 0	0 , 0	6.9 , 8.3	9.7 , 0	5.6 , 17	6.9 , 25
90°	2.8 , 0	4.2 , 0	4.2 , 0	5.6 , 0	2.8 , 0	1.4 , 0	11 , 17	8.3 , 0	5.6 , 8.3

Table 4.11: The detection rates of the subjects (left hand digit) and the algorithm (right hand digits) are reported like in Table 4.10, for $N = 200$. Again, higher and lower rates than the algorithm were achieved by at least one subject, in all conditions.

α/n	4	5	6	7	8	9	10	11	12
9°	12 , 17	9.7 , 8.3	**26 , 58**	40 , 42	49 , 42	56 , 58	60 , 83	68 , 92	53 , 58
18°	12 , 25	11 , 8.3	25 , 25	25 , 33	47 , 50	39 , 33	51 , 67	39 , 50	42 , 67
36°	4.2 , 0	9.7 , 17	8.3 , 17	11 , 25	17 , 17	17 , 17	12 , 25	18 , 25	18 , 33
45°	4.2 , 0	0 , 0	**0 , 8.3**	1.4 , 0	4.2 , 0	4.2 , 0	22 , 17	21 , 8.3	12 , 8.3
72°	1.4 , 0	0 , 0	2.8 , 0	1.4 , 0	2.8 , 8.3	8.3 , 8.3	14 , 8.3	4.2 , 0	6.9 , 0
90°	0 , 0	1.4 , 0	2.8 , 0	2.8 , 0	4.2 , 0	2.8 , 8.3	1.4 , 0	1.4 , 0	2.8 , 0

Table 4.12: Detection rates of the subjects and the algorithm, like in Tables 4.10 and 4.11, but for $N = 600$. We put the detection scores in bold font any time the algorithm's detection rate was strictly lower than the minimum, or strictly greater than the maximum of the subjects' rates in the same condition. This happened only twice: for ($\alpha = 9°, n = 6$) and ($\alpha = 45°, n = 6$). In the former case, a detection rate of 50% was reached by subjects in 4 experimental sessions.

α/n	4	5	6	7	8	9	10	11	12
9°	31 , 28	25 , 25	46 , 72	53 , 64	54 , 64	58 , 72	61 , 81	72 , 89	62 , 75
18°	18 , 14	24 , 19	40 , 42	43 , 42	54 , 61	51 , 58	59 , 64	53 , 67	62 , 75
36°	8.8 , 2.8	21 , 19	17 , 14	25 , 17	34 , 33	33 , 39	19 , 42	26 , 36	38 , 42
45°	7.9 , 2.8	11 , 5.6	12 , 11	15 , 11	19 , 19	19 , 22	32 , 31	31 , 33	20 , 31
72°	7.4 , 0	4.6 , 5.6	6.5 , 2.8	3.7 , 0	5.6 , 2.8	14 , 14	15 , 2.8	8.8 , 8.3	9.3 , 17
90°	1.9 , 0	3.7 , 0	5.1 , 2.8	5.6 , 2.8	8.3 , 8.3	5.6 , 5.6	9.7 , 8.3	8.8 , 2.8	5.1 , 8.3

Table 4.13: Detection rates, in percents, achieved by the subjects and the algorithm in each (α, n) condition, averaged over all values of N. In each case, the left hand digit is the average performance over all subjects, and the right hand one is the algorithm's detection rate.

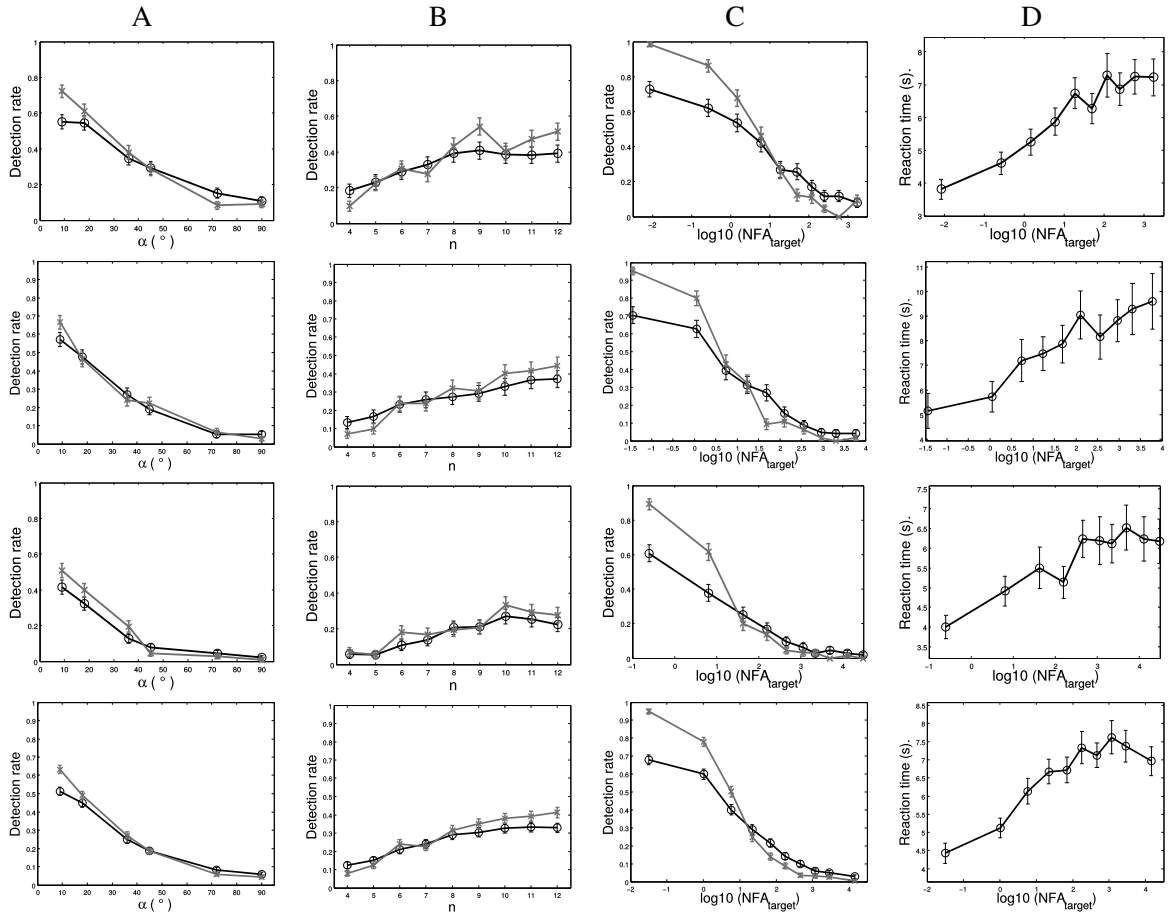


Figure 4.11: The subjects of Experiments I and II are compared to the algorithm. The solid black lines represent the subjects, whereas the gray solid lines represent the algorithm's results. The first, second and third rows display the results corresponding to the three values $N = 100, 200$ and 600 respectively. The fourth row shows the same data averaged over all three conditions. The detection rates are plotted as functions of the jitter intensity α (column A), the number n of aligned elements (B) and the $\log_{10}$ of the target's NFA (C). The latter is also the x variable in column D's plots, the y axis representing the reaction time, in seconds.

achieved by human subjects and, on the other hand, the group of the 3 detection rates achieved by the algorithm. The purpose of this analysis was to measure, for each N, if the algorithm's detection rates differed significantly from the subjects ones. The results are reported in Table 4.14.

Tables 4.15, 4.16 and 4.17 show the results of an *agreement test*, as defined in Section 3.7, between the subjects and the algorithm, as well as among subjects. This analysis measures the similarity in the distributions of the correct answers of two observers who saw the very same stimuli, beyond their detection rates. The variable k denotes the number of trials in which the observers agreed, i.e. both detected or both failed. The p-value p_k indicates if the observed k is significantly greater than expected under the basic assumption that the detections of the two observers are distributed randomly and independently. The lower p_k, the more significant the agreement. In Tables 4.15, 4.16 and 4.17 we report, for each main sequence of stimuli, the values (k, p_k) resulting from the comparisons among the six subjects who saw that sequence. In the last column, each subject is compared to the algorithm through the same test. For convenience, in the tables the subjects are

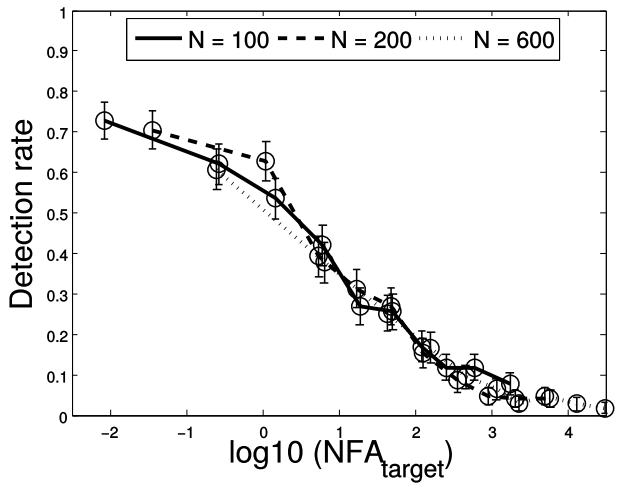


Figure 4.12: Detection rates achieved by the subjects, as functions of $\log_{10}$(NFA) of the target paths, for the three different values of N. The superimposition of the three curves indicate a perceptual equivalence among stimuli characterized by the same NFA. Indeed, for a given NFA value, the detection rates are substantially the same no matter the value of N.

N	100	200	600
$F(1,19)$	0.60	0.12	0.95
p	0.44	0.73	0.34

Table 4.14: Results of the one-way analysis of variance to test the difference between the subjects and the algorithm. In the second row are reported the computed $F(df_1, df_2)$ values of a Fisher test for data concerning 21 individuals divided into two groups, yielding $df_1 = 2 - 1 = 1$ and $df_2 = 21 - 2 = 19$ as degrees of freedom. In the third row, the corresponding p-values are given. p-values above a certain threshold (typically 0.05) indicate that differences between the subjects and the algorithm are not statistically more significant than those among subjects nor than those among the answers given by the algorithm.

	Subject	1	2	3	4	5	6	Algorithm
seq. A	1	-	(158, 6.2e-08)	(150, 0.00037)	(154, 4.4e-05)	(147, 1.4e-06)	(154, 8.7e-08)	(160, 1.1e-08)
	2	-	-	(146, 0.00043)	(158, 1.2e-07)	(157, 2.3e-10)	(154, 3.6e-08)	(158, 1e-08)
	3	-	-	-	(160, 8.4e-07)	(151, 5.4e-08)	(158, 2.7e-09)	(158, 8.5e-08)
	4	-	-	-	-	(145, 6.2e-06)	(162, 5.1e-11)	(158, 1.2e-07)
	5	-	-	-	-	-	(167, 3.6e-15)	(159, 3.2e-11)
	6	-	-	-	-	-	-	(166, 2.7e-13)
seq. B	1	-	(157, 2.8e-07)	(151, 3.9e-06)	(156, 9.2e-08)	(160, 1.1e-11)	(159, 1.3e-11)	(152, 1.1e-07)
	2	-	-	(150, 0.0011)	(155, 6.3e-05)	(149, 1.9e-07)	(154, 1.9e-10)	(145, 0.00012)
	3	-	-	-	(161, 1.2e-08)	(155, 1.6e-09)	(142, 2e-05)	(145, 3.7e-05)
	4	-	-	-	-	(152, 2.4e-08)	(153, 3.4e-09)	(148, 5.4e-06)
	5	-	-	-	-	-	(167, 1.2e-15)	(148, 3.1e-07)
	6	-	-	-	-	-	-	(155, 5.7e-10)
seq. C	1	-	(162, 9e-08)	(164, 6.2e-10)	(151, 0.02)	(163, 1.3e-10)	(159, 7e-10)	(164, 1.3e-11)
	2	-	-	(152, 1.6e-05)	(157, 0.00071)	(149, 2e-05)	(149, 3.8e-06)	(154, 1.9e-07)
	3	-	-	-	(143, 0.043)	(159, 1.9e-09)	(171, 5.2e-16)	(162, 4.5e-11)
	4	-	-	-	-	(142, 0.0095)	(134, 0.057)	(141, 0.0061)
	5	-	-	-	-	-	(152, 8.7e-08)	(155, 1.2e-08)
	6	-	-	-	-	-	-	(159, 1.3e-10)

Table 4.15: Values (k, p_k) of the agreement test for $N = 100$. The variable k is the number of identical responses given by two observers in the corresponding sequence, and the p-value p_k measures how significant it is to observe at least k identical responses. A p-value under a certain threshold (typically 0.01) indicates a significant agreement between two observers. The indexes identifying the subjects are different from those used in tables 4.1-4.7, and one index may correspond to different people from a sequence to the other. For example, Subject 5 for sequence A is not necessarily the same person as Subject 5 for sequence B.

identified by an index between 1 and 6 for each sequence. These indexes are different from those used in tables 4.1-4.7, and one index may correspond to different people from a sequence to the other. For example, Subject 5 for sequence A is not necessarily the same person as Subject 5 for sequence B.

	Subject	1	2	3	4	5	6	Algorithm
seq. A	1	-	(171, 7.2e-07)	(168, 1.8e-06)	(152, 1.5e-06)	(167, 5.9e-08)	(166, 6.5e-09)	(163, 9.8e-07)
	2	-	-	(171, 7.2e-07)	(147, 8.3e-05)	(164, 3e-06)	(157, 4.1e-05)	(162, 6.7e-06)
	3	-	-	-	(156, 4.3e-08)	(169, 8.1e-09)	(158, 8.3e-06)	(161, 5.2e-06)
	4	-	-	-	-	(167, 2.7e-13)	(174, 2.7e-17)	(175, 5.9e-18)
	5	-	-	-	-	-	(169, 2.9e-11)	(166, 3.8e-09)
	6	-	-	-	-	-	-	(171, 1.9e-12)
seq. B	1	-	(163, 2.9e-06)	(157, 0.012)	(176, 4.5e-15)	(170, 4.3e-10)	(169, 4.1e-10)	(169, 6.6e-09)
	2	-	-	(166, 0.00027)	(177, 2.3e-15)	(165, 2.5e-07)	(162, 1.1e-06)	(164, 3.5e-06)
	3	-	-	-	(155, 0.0003)	(161, 0.00043)	(160, 0.00022)	(160, 0.0064)
	4	-	-	-	-	(174, 4.3e-14)	(173, 7.9e-14)	(167, 3.8e-10)
	5	-	-	-	-	-	(165, 1.3e-08)	(167, 2.2e-08)
	6	-	-	-	-	-	-	(174, 3.3e-12)
seq. C	1	-	(159, 1.4e-05)	(163, 2.1e-07)	(156, 8.6e-08)	(168, 2.2e-11)	(160, 2.3e-09)	(161, 9.3e-09)
	2	-	-	(160, 0.00013)	(163, 2.7e-10)	(167, 7.3e-10)	(155, 6.6e-07)	(164, 3.8e-09)
	3	-	-	-	(159, 1.4e-08)	(171, 4.2e-12)	(161, 1.9e-09)	(160, 1e-07)
	4	-	-	-	-	(174, 1.3e-16)	(170, 4e-15)	(167, 4.9e-13)
	5	-	-	-	-	-	(166, 2.3e-12)	(165, 5.5e-11)
	6	-	-	-	-	-	-	(169, 4.8e-14)

Table 4.16: Values (k, p_k) of the agreement test for $N = 200$. The variable k is the number of identical responses given by two observers in the corresponding sequence, and the p-value p_k measures how significant it is to observe at least k identical responses. A p-value under a certain threshold (typically 0.01) indicates a significant agreement between two observers. The indexes identifying the subjects are different from those used in tables 4.1-4.7, and one index may correspond to different people from a sequence to the other.

	Subject	1	2	3	4	5	6	Algorithm
seq. A	1	-	(169, 0.00062)	(173, 0.0016)	(193, 9.8e-19)	(188, 1.4e-17)	(181, 1.2e-09)	(179, 5.5e-11)
	2	-	-	(188, 0.058)	(182, 1e-05)	(175, 6.3e-07)	(184, 5.3e-05)	(172, 0.00028)
	3	-	-	-	(184, 0.0029)	(169, 0.016)	(188, 0.011)	(176, 0.0011)
	4	-	-	-	-	(185, 1.5e-13)	(192, 7.8e-12)	(182, 4.2e-10)
	5	-	-	-	-	-	(179, 3.1e-09)	(185, 5.4e-15)
	6	-	-	-	-	-	-	(180, 4.4e-08)
seq. B	1	-	(177, 0.00012)	(180, 1.5e-06)	(174, 0.00015)	(174, 5.2e-09)	(174, 0.0021)	(178, 1.3e-06)
	2	-	-	(183, 9e-08)	(179, 2.3e-06)	(181, 4.8e-13)	(181, 7.5e-06)	(183, 7.5e-09)
	3	-	-	-	(176, 1.1e-05)	(172, 3.3e-08)	(178, 2.8e-05)	(180, 6.3e-08)
	4	-	-	-	-	(176, 2e-10)	(178, 9.7e-06)	(174, 1.1e-05)
	5	-	-	-	-	-	(172, 7.7e-08)	(176, 1.2e-10)
	6	-	-	-	-	-	-	(180, 3.9e-07)
seq. C	1	-	(178, 5.8e-07)	(171, 3.7e-05)	(178, 1.6e-06)	(174, 1.8e-08)	(180, 1.9e-07)	(177, 4e-08)
	2	-	-	(185, 2.6e-12)	(174, 2.7e-05)	(184, 2.3e-14)	(190, 1.1e-13)	(171, 7.3e-06)
	3	-	-	-	(179, 1.3e-08)	(181, 3.9e-13)	(181, 1.2e-09)	(178, 7.5e-10)
	4	-	-	-	-	(182, 7.1e-13)	(178, 1.6e-06)	(173, 2.4e-06)
	5	-	-	-	-	-	(182, 7.1e-13)	(177, 3.5e-11)
	6	-	-	-	-	-	-	(173, 2.4e-06)

Table 4.17: Values (k, p_k) of the agreement test for $N = 600$. The variable k is the number of identical responses given by two observers in the corresponding sequence, and the p-value p_k measures how significant it is to observe at least k identical responses. A p-value under a certain threshold (typically 0.01) indicates a significant agreement between two observers. The indexes identifying the subjects are different from those used in tables 4.1-4.7, and one index may correspond to different people from a sequence to the other.

4.7 Discussion

We shall first analyze the subjects' results as functions of the NFA assigned by the algorithm to each target path. Therefore we refer to the black curves in column C and D of Figure 4.11, and to the plot of Figure 4.12. In accordance with the observations of the previous chapters, detection rates and reaction times are respectively decreasing and increasing functions of the NFA. Once again, this monotony is clear and steady across the three N conditions, indicating that the NFA of the target paths measures consistently their difficulty to be detected. In addition, Figure 4.12 shows that there is more than a correlation between detectability and NFA: for a given NFA value, the detection rates are substantially the same no matter the value of N. This seems to draw a perceptual equivalence among stimuli characterized by the same NFA, which was one of the most important claims to test.

Now let's compare *directly* the algorithm to the subjects. Figure 4.11 displays a strong similarity between the subjects and the algorithm's detection curves, especially columns A and B. Remarkably, the differences between the algorithm and the subjects appear to follow approximately the same pattern in all the considered plots: in general, the algorithm achieves higher rates than the subjects in the "easiest" half of the stimuli (with less angular jitter, longer target paths, smaller NFA), whereas the contrary occurs in the other half. These differences may seem insignificant when looking at each plot of columns A and B separately, but they appear relevant through their repetition in almost all these plots and, more important, they are obvious in the plots of column C.

The localization of the major discrepancies in the range of small NFAs ($\log_{10}(\text{NFA}) < 1$) yields the following interpretation. The algorithm performs an ideal search strategy, during which the target path is very likely to be tested. Indeed, the search process is analogous to the method used to define target paths (see Section 4.2.1), that is why the family of tests defined in the algorithm almost always contains the target path. Whenever the considered variables in this path deviate from the *a contrario* assumptions, the target is assigned a small NFA and is likely to be detected. Yet, in some cases, a path with a low NFA is not actually salient, as for example in the last row of Figure 4.9. In some other cases, such as in the third row of Figure 4.9, the target path was seen by some subjects, but it was more often missed. This might be explained by a non-exhaustive and possibly less attentive search, or the presence of another relevant path, preferred by a majority of subjects (second and third rows of Figure 4.9, and Figure 4.10). These cases, although they are a minority, reveal a limitation of the model, either in the variables we consider or the way we process them. Indeed, paths with low NFAs, typically $\log_{10}(\text{NFA}) < 0$, are expected to yield effortless detections among the subjects. The inaccuracy of the model could result from several flaws. For example, the balance in the importance given to the proximity cue and the orientation cues, might not be realistic. Also, the independence assumption on the distances between adjacent points might be responsible for low NFAs without salience.

The same weaknesses could explain that the algorithm's performance drops down faster than the subjects'. Furthermore, the visual system probably relies on other cues, such as convexity, and many other mechanisms involved in the task are not handled by our model.

Nevertheless, it is to be noticed that in Figure 4.11 and Tables 4.10-4.12, the algorithm is compared to the *average* of the subjects, while it actually represents only *one* artificial observer. If we isolate the results of one single human subject it is likely that they will not match the average of the other subjects either. Therefore, another comparison criterion is necessary to evaluate our algorithm in a Turing test prospective.

To start with, it is worth noting that, despite the differences observed in Tables 4.10-4.12, between the subjects' and the algorithm's detection rates, at least one subject performed equally or better, and at least one performed equally or worse than the algorithm, in all but two conditions

out of 166. That is to say that the results of the algorithm are almost always within the boundaries of perception.

Secondly, concerning the detection rates achieved on the complete main sequences, the ANOVA detailed in the previous section indicates that the differences between subjects and the algorithm are not more significant than the variations *among* subjects (Table 4.14). Indeed, all p-values are above 0.34, whereas a significant difference between subjects and the algorithm would be reported by a much lower p-value, typically $p < 0.05$.

Finally, the purpose of the *agreement test* is to draw a comparison *trial by trial*, instead of doing it on the detection rates. Here again, Tables 4.15-4.17 reveal no more significant difference between the algorithm and the subjects than among subjects. In these tables, a small p-value (typically $p < 0.01$) indicates a significant agreement between two observers. As we can see, in most cases the p-values are much smaller than this threshold. In particular, $p < 0.007$ for any comparison between a subject and the algorithm, and the few cases showing p-values greater than 0.05 concern the agreement between two human subjects.

4.8 Conclusion

In this chapter we applied the *a contrario* theory to provide a quantitative interpretation of perceptual grouping by good continuation, intended in a more general sense than in the previous chapters, since it is no longer restricted to alignments. Following the method introduced in Chapter 3, we implemented a parameterless algorithm, designed to detect non-accidental smooth paths in a masking background, and compared this algorithm to the human visual perception through psychophysical experiments. This comparison revealed an equivalence between the NFA of the target paths and their detectability by subjects. What is more, the algorithm's detection curves match well the subjects ones, and none of the statistical tests we performed could tell our artificial observer and the human subjects apart. Of course, the match is not perfect, and we spotted several cases in which the detection algorithm was not consistent with perception. However, the overall results give new credit to the non-accidentalness principle, as a way to interpret and predict the perceptual grouping in masking conditions.

5 Conclusion and perspectives

This dissertation has presented a methodology to investigate the role of non-accidentalness in perceptual grouping. Non-accidentalness was modeled by the *a contrario* theory, and we addressed the case of the good continuation of oriented elements.

As a first step, we studied the perception of alignments of Gabor patches, of different lengths and affected by several intensities of angular jitter. The detection rates achieved by subjects were proved to be a decreasing function of the NFA computed for each target alignment by our detection algorithm, and their reaction times an increasing function of the same NFA. Additionally, the algorithm's performance on the same task was very close to that of the human subjects. Of course, differences appeared between humans and our artificial observer. Besides the exhaustive search and almost infinite precision of the algorithm, which are unrealistic in a model of a cognitive system, part of differences might be explained by the existence, in some stimuli, of regularity cues that were insufficiently masked, and thus accessible to the subjects, whereas they were not processed by the algorithm. It goes without saying that this is only one among the many parameters that are not taken into account by our model.

In the generalization of this approach to a broader class of stimuli, in which the targets were smooth paths of Gabor patches, there was no masking problem anymore. Indeed, each path was defined by choosing a chain of elements among the background points and then setting their orientations to form a more or less jittered discrete curve. Another improvement was that one of the presented experiments was run in a fully controlled environment, and its results were not significantly different from those obtained in more variable conditions. The observations retrieved from the comparison of perceptual data to our second *a contrario* algorithm, were in accordance with the study on alignments. This time, the equivalence between the detectability of the targets and their NFA was even clearer. Our model for curves salience relied on the proximity and good-continuation cues, that we combined by normalizing and summing angular deviations and distances between adjacent points. A natural extension of this work would be to include other cues, such as the smoothness of curves, in our geometric model, and to try other strategies of cue combinations. Besides, the proposed NFA for curves was built on approximations and inaccurate assumptions. Much can be done at the mathematical level to improve these aspects and design an algorithm that would comply more precisely with the non-accidentalness property.

Contrary to many well known experiments on good continuation, which inspired our work, in our experiments attention played an important part in the perceptual process, since the subjects were presented with the stimuli without time limitation. Expanding our framework to preattentive paradigms could be the object of a future work. What is more, we believe that the presented method could be adapted to address other questions of quantitative Gestalt such as symmetry and motion perception.

So far, the present thesis has provided support to the *a contrario* theory as a relevant approach to

handle perceptual tasks in computer vision. On the other hand, the predictions achieved by this theory on the results of humans subjects, seem to underline the importance of the non-accidentalness principle in perception.

Bibliography

[1] N. Ahuja and M. Tuceryan. Extraction of early perceptual structure in dot patterns: integrating region, boundary, and component gestalt. *Comput. Vision Graph. Image Process.*, 48(3):304–356, December 1989.

[2] F. Attneave. Some informational aspects of visual perception. *Psychological Review*, 61(3):183–193, May 1954.

[3] Thomas O Binford. Inferring surfaces from images. *Artificial Intelligence*, 17(1):205–244, 1981.

[4] S. Blusseau, A. Carboni, A. Maiche, J.M. Morel, and R. Grompone von Gioi. A psychophysical evaluation of the a contrario detection theory. In *Image Processing (ICIP), 2014 IEEE International Conference on*, pages 1091–1095, Oct 2014.

[5] F. Cao, J.L. Lisani, J.-M. Morel, P. Musé, and F. Sur. *A theory of shape identification*, volume 1948 of *Lecture Notes in Mathematics*. Springer, 2008.

[6] J. Cardelino, V. Caselles, M. Bertalmio, and G. Randall. A contrario hierarchical image segmentation. In *Image Processing (ICIP), 2009 16th IEEE International Conference on*, pages 4041–4044, 2009.

[7] M. Demeyer and B. Machilsen. The construction of perceptual grouping displays using GERT. *Behavior Research Methods, online first.*, pages 1–8, 2011.

[8] A. Desolneux, L. Moisan, and J. M. Morel. Meaningful alignments. *International Journal of Computer Vision*, 40(1):7–23, 2000.

[9] A. Desolneux, L. Moisan, and J. M. Morel. Computational gestalts and perception thresholds. *Journal of Physiology - Paris*, 97:311–324, 2003.

[10] A. Desolneux, L. Moisan, and J. M. Morel. A grouping principle and four applications. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(4):508–513, April 2003.

[11] A. Desolneux, L. Moisan, and J.M. Morel. *From Gestalt Theory to Image Analysis, a Probabilistic Approach*, volume 34 of *Interdisciplinary Applied Mathematics*. Springer, 2008.

[12] W.D. Ellis, editor. *A Source Book of Gestalt Psychology*. Humanities Press, 1967 (originally 1938).

[13] Udo A. Ernst, Sunita Mandon, Nadja Schinkel-Bielefeld, Simon D. Neitzel, Andreas K. Kreiter, and Klaus R. Pawelzik. Optimality of human contour integration. *PLoS Comput Biol*, 8(5):e1002520, 05 2012.

[14] Jacob Feldman. Curvilinearity, covariance, and regularity in perceptual groups. *Vision Research*, 37(20):2835–2848, October 1997.

[15] Jacob Feldman. Regularity-based perceptual grouping. *Computational Intelligence*, 13(4):582–623, 1997.

[16] Jacob Feldman. Bayesian contour integration. *Attention, Perception, & Psychophysics*, 63:1171–1182, 2001.

[17] Jacob Feldman. Perceptual grouping by selection of a logically minimal model. *International Journal of Computer Vision*, 55(1):5–25, 2003.

[18] Jacob Feldman. Formation of visual "objects" in the early computation of spatial relations. *Perception & Psychophysics*, 69(5):816–827, 2007.

[19] Jacob Feldman. Bayes and the simplicity principle in perception. *Psychological Review*, 116(4):875, 2009.

[20] Jacob Feldman and Manish Singh. Information along contours and object boundaries. *Psychological Review*, 112(1):243–252, 2005.

[21] D. J. Field, A. Hayes, and R. F. Hess. Contour integration by the human visual system: Evidence for a local "association field". *Vision Research*, 33(2):173 – 193, 1993.

[22] François Fleuret, Ting Li, Charles Dubout, Emma K. Wampler, Steven Yantis, and Donald Geman. Comparing machines and humans on a visual categorization test. *Proceedings of the National Academy of Sciences*, 108(43):17621–17625, 2011.

[23] Wilson S. Geisler. Contributions of ideal observer theory to vision research. *Vision Research*, 51(7):771 – 781, 2011. Vision Research 50th Anniversary Issue: Part 1.

[24] Ziv Gigus and Jitendra Malik. Detecting curvilinear structure in images. Technical Report 91/619, U. C. Berkeley CSD, 1991.

[25] Joshua I Gold and Takeo Watanabe. Perceptual learning. *Current biology: CB*, 20(2), 2010.

[26] R. Grompone von Gioi and J. Jakubowicz. On computational Gestalt detection thresholds. *Journal of Physiology – Paris*, 103(1-2):4–17, 2009.

[27] R. Grompone von Gioi, J. Jakubowicz, J. M Morel, and G. Randall. Lsd: A fast line segment detector with a false detection control. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 32(4):722–732, 2010.

[28] R. Grompone von Gioi, J. Jakubowicz, J. M. Morel, and G. Randall. LSD: a line segment detector. *Image Processing On Line*, 2012.

[29] Bénédicte Grosjean and Lionel Moisan. A-contrario detectability of spots in textured backgrounds. *Journal of Mathematical Imaging and Vision*, 33(3):313–337, 2009.

[30] S. Grossberg and E. Mingolla. Neural dynamics of perceptual grouping: Textures, boundaries, and emergent segmentations. *Attention, Perception, & Psychophysics*, 38(2):141–171, 1985.

[31] Feng Han and Song-Chun Zhu. Bottom-up/top-down image parsing with attribute grammar. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(1):59–73, January 2009.

[32] R. F. Hess, K. A. May, and S. O. Dumoulin. Contour integration: Psychophysical, neurophysiological and computational perspectives. In J. Wagemans, editor, *Oxford Handbook of Perceptual Organization*. Oxford University Press, 2014.

[33] R.F Hess, A Hayes, and D.J Field. Contour integration and cortical processing. *Journal of Physiology-Paris*, 97(2 - 3):105 – 119, 2003. Neurogeometry and visual perception.

[34] Robert Hess and David Field. Integration of contours: new insights. *Trends in cognitive sciences*, 3(12):480–486, 1999.

[35] Takeo Kanade. Recovery of the three-dimensional shape of an object from a single view. *Artificial Intelligence*, 17(1-3):409 – 460, 1981.

[36] Gaetano Kanizsa. *Organization in vision: Essays on Gestalt perception*. Praeger New York:, 1979.

[37] Gaetano Kanizsa. *Grammatica del vedere*. Il Mulino, 1980.

[38] Gaetano Kanizsa. *Vedere e pensare*. Il Mulino, 1991.

[39] Mitesh K. Kapadia, Minami Ito, Charles D. Gilbert, and Gerald Westheimer. Improvement in visual sensitivity by changes in local context: Parallel studies in human observers and in V1 of alert monkeys. *Neuron*, 15(4):843 – 856, 1995.

[40] D. Kersten, P. Mamassian, and A. Yuille. Object perception as bayesian inference. *Annual Review of Psychology*, 55(1):271–304, 2004.

[41] W. Köhler. *Gestalt Psychology*. Liveright, 1947.

[42] Ilona Kovacs and Bela Julesz. A closed curve is much more than an incomplete one: Effect of closure in figure-ground segmentation. *Proceedings of the National Academy of Sciences*, 90(16):7495–7497, 1993.

[43] Y.G. Leclerc. Constructing simple stable descriptions for image partitioning. *International journal of computer vision*, 3(1):73–102, 1989.

[44] Timothy Ledgeway, Robert F Hess, and Wilson S Geisler. Grouping local orientation and direction signals to extract spatial contours: Empirical tests of "association field" models of contour integration. *Vision research*, 45(19):2511–2522, 2005.

[45] J. Lezama, J. Morel, G. Randall, and R. Grompone von Gioi. A contrario 2D point alignment detection. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 37(3):499–512, March 2015.

[46] José Lezama. *On grouping theory in dot patterns, with applications to perception theory and 3D inverse geometry*. PhD thesis, École normale supérieure de Cachan-ENS Cachan, 2015.

[47] José Lezama, Rafael Grompone von Gioi, Gregory Randall, and Jean-Michel Morel. A contrario detection of good continuation of points. In *Image Processing (ICIP), 2014 IEEE International Conference on*, pages 4757–4761. IEEE, 2014.

[48] D. Lowe. *Perceptual Organization and Visual Recognition*. Kluwer Academic Publishers, 1985.

[49] David G. Lowe. *Perceptual Organization and Visual Recognition*. PhD thesis, Standford University, 1985.

[50] David G Lowe. Three-dimensional object recognition from single two-dimensional images. *Artificial intelligence*, 31(3):355–395, 1987.

[51] David G Lowe. Object recognition from local scale-invariant features. In *Computer vision, 1999. The proceedings of the seventh IEEE international conference on*, volume 2, pages 1150–1157. IEEE, 1999.

[52] David G Lowe. Distinctive image features from scale-invariant keypoints. *International journal of computer vision*, 60(2):91–110, 2004.

[53] David G. Lowe and Thomas O. Binford. The interpretation of three-dimensional structure from image curves. In *Proceedings of the 7th international joint conference on Artificial intelligence - Volume 2*, IJCAI'81, pages 613–618, San Francisco, CA, USA, 1981. Morgan Kaufmann Publishers Inc.

[54] B. Machilsen and J. Wagemans. Integration of contour and surface information in shape detection. *Vision Research*, 51:179–186, 2011.

[55] Bart Machilsen, Maarten Pauwels, and Johan Wagemans. The role of vertical mirror symmetry in visual shape detection. *Journal of Vision*, 9(12):11, 2009.

[56] P. Mamassian, M. S. Landy, and L. T. Maloney. Bayesian modelling of visual perception. In Rajesh Rao, Bruno Olshausen, and Michael Lewicki, editors, *Probabilistic Models of the Brain: Perception and Neural Function*, pages 13–36. Cambridge, MA: MIT Press, March 2002.

[57] D. Marr. *Vision*. Freeman and co., 1982.

[58] W. Metzger. *Gesetze des Sehens*. Verlag Waldemar Kramer, Frankfurt am Main, third edition, 1975.

[59] W. Metzger. *Laws of Seeing*. The MIT Press, 2006 (originally 1936). English translation of the first edition of [58].

[60] D. Mumford. Pattern theory: the mathematics of perception. *Proceedings of the International Congress of Mathematicians, Beijing*, I:401–422, 2002.

[61] David Mumford and Agnès Desolneux. *Pattern theory: the stochastic analysis of real-world signals*. AK Peters, 2010.

[62] P. Musé, F. Sur, F. Cao, and Y. Gousseau. Unsupervised thresholds for shape matching. In *IEEE Int. Conf. on Image Processing, ICIP 2003*, Barcelona, Spain, September 2003.

[63] G.E. Nygård, T. Van Looy, and J. Wagemans. The influence of orientation jitter and motion on contour saliency and object identification. *Vision Research*, 49:2475–2484, 2009.

[64] S. E. Palmer. *Vision Science: Photons to Phenomenology*. The MIT Press, 1999.

[65] Stephen E Palmer. Perceptual organization in vision. *Stevens' handbook of experimental psychology*, 2002.

[66] Pierre Parent and Steven W Zucker. Trace inference, curvature consistency, and curve detection. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 11(8):823–839, 1989.

[67] V. Pătrăucean, R. Grompone von Gioi, and M. Ovsjanikov. Detection of mirror-symmetric image patches. In *CVPRW*, pages 211–216, 2013.

[68] Viorica Pătrăucean, Pierre Gurdjos, and Rafael Grompone von Gioi. A parameterless line segment and elliptical arc detector with enhanced ellipse fitting. In Andrew Fitzgibbon, Svetlana Lazebnik, Pietro Perona, Yoichi Sato, and Cordelia Schmid, editors, *Computer Vision - ECCV 2012*, Lecture Notes in Computer Science, pages 572–585. Springer Berlin Heidelberg, 2012.

[69] A. Pinar Saygin, I. Cicekli, and V. Akman. Turing test: 50 years later. *Minds and Machines*, 10(4):463–518, 2000.

[70] K. Preiss. *A Theoretical and Computational Investigation into Aspects of Human Visual Perception: Proximity and Transformations in Pattern Detection and Discrimination*. PhD thesis, University of Adelaide, June 2006.

[71] I. Rock. *The logic of perception*. MIT Press, Cambridge, MA, 1983.

[72] Sudeep Sarkar and Kim L. Boyer. Perceptual organization in computer vision: A review and a proposal for a classificatory structure. *IEEE Transactions on Systems, Man, and Cybernetics*, 23(2):382–399, 1993.

[73] M. Sassi, K. Vancleef, B. Machilsen, S. Panis, and J. Wagemans. Identification of everyday objects on the basis of gaborized outline versions. *i-Perception*, 1:121–142, 2010.

[74] Michaël Sassi, Maarten Demeyer, and Johan Wagemans. Peripheral contour grouping and saccade targeting: The role of mirror symmetry. *Symmetry*, 6(1):1–22, 2014.

[75] Roger N Shepard. *Perceptual organization*, chapter Psychophysical Complementarity, pages 279–341. Lawrence Erlbaum Associates, Inc., 1981.

[76] Kent A Stevens. *Surface perception from local analysis of texture and contour*. PhD thesis, Massachusetts Institue of Technology, 1980.

[77] Kent A Stevens. The visual interpretation of surface contours. *Artificial Intelligence*, 17(1):47–73, 1981.

[78] S.S Stevens. *Psychophysics*. Transaction Publishers, 1986.

[79] M. Tepper, P. Musé, and A. Almansa. Meaningful clustered forest: an automatic and robust clustering algorithm. *CoRR*, abs/1104.0651, 2011.

[80] S. P. Tripathy, A. J. Mussap, and H. B. Barlow. Detecting collinear dots in noise. *Vision Research*, 39(25):4161–4171, 1999.

[81] Alan M. Turing. Computing machinery and intelligence. *Mind*, 59:433–460, 1950.

[82] W. R. Uttal, Lynne M. Bunnel, and Stuart Corwin. On the detectability of straight lines in visual noise: An extension of French's paradigm into the millisecond domain. *Perception & Psychophysics*, 8(6):385–388, 1970.

[83] William R. Uttal. The effect of deviations from linearity on the detection of dotted line patterns. *Vision Res*, 13(11):2155–63, 1973.

[84] Kathleen Vancleef and Johan Wagemans. Component processes in contour integration: A direct comparison between snakes and ladders in a detection and a shape discrimination task. *Vision research*, 92:39–46, 2013.

[85] Maria Carolina Vanegas, Isabelle Bloch, and Jordi Inglada. Detection of aligned objects for high resolution image understanding. In *IGARSS*, pages 464–467, 2010.

[86] J. Wagemans, J. H. Elder, M. Kubovy, S. E. Palmer, M. A. Peterson, M. Singh, and R. von der Heydt. A century of Gestalt psychology in visual perception: I. Perceptual grouping and figure–ground organization. *Psychological bulletin*, 138(6):1172–1217, 2012.

[87] J. Wagemans, J. Feldman, S. Gepshtein, R. Kimchi, J. R. Pomerantz, P. A. van der Helm, and C. van Leeuwen. A century of Gestalt psychology in visual perception: II. Conceptual and theoretical foundations. *Psychological Bulletin*, 138(6):1218–1252, 2012.

[88] Johan Wagemans. Perceptual use of nonaccidental properties. *Canadian Journal of Psychology*, 46(2):236–279, 1992.

[89] Johan Wagemans. Skewed symmetry: A nonaccidental property used to perceive visual forms. *Journal of Experimental Psychology: Human Perception and Performance*, 19(2):364–380, 1993.

[90] Max Wertheimer. Untersuchungen zur Lehre von der Gestalt. II. *Psychologische Forschung*, 4(1):301–350, 1923. An abridged translation to English is included in [12].

[91] John Wilder, Jacob Feldman, and Manish Singh. Contour complexity and contour detection. *Journal of Vision*, 15(6):6, 2015.

[92] Andrew P Witkin. Intensity-based edge classification. In *AAAI*, volume 82, pages 36–41, 1982.

[93] Andrew P. Witkin. Scale-space filtering. In *Proceedings of the Eighth International Joint Conference on Artificial Intelligence - Volume 2*, IJCAI'83, pages 1019–1022, San Francisco, CA, USA, 1983. Morgan Kaufmann Publishers Inc.

[94] Andrew P Witkin. Scale-space filtering: A new approach to multi-scale description. In *Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP'84.*, volume 9, pages 150–153. IEEE, 1984.

[95] Andrew P. Witkin and Jay M. Tenenbaum. *Human and Machine Vision*, chapter On the role of structure in vision, pages 481–543. Academic Press, 1983.

[96] Andrew P. Witkin and Jay M. Tenenbaum. What is perceptual organization for? *IJCAI-83*, 2:1023–1026, 1983.

[97] Shih-Cheng Yen and Leif H. Finkel. Extraction of perceptually salient contours by striate cortical networks. *Vision Research*, 38(5):719 – 741, 1998.

[98] Song-Chun Zhu and David Mumford. A stochastic grammar of images. *Foundations and Trends in Computer Graphics and Vision*, 2(4):259–362, January 2006.

[99] Song-Chun Zhu and Alan Yuille. Region competition: Unifying snakes, region growing, and bayes/MDL for multiband image segmentation. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 18(9):884–900, 1996.

[100] Steven W. Zucker, Allan Dobbins, and Lee Iverson. Two stages of curve detection suggest two styles of visual computation. *Neural Comput.*, 1(1):68–81, March 1989.