On metacognition and the dynamics of selective attention

Samuel Recht

THÈSE DE DOCTORAT
DE L'UNIVERSITÉ PSL

Préparée à l'École normale supérieure

On metacognition and the dynamics of selective attention

Soutenue par
Samuel RECHT
Le 3 décembre 2019

Ecole doctorale n° 540
Lettres, Arts, Sciences humaines et sociales

Spécialité
Sciences cognitives, neurosciences, psychologie

Composition du jury :

Suliann, BEN HAMED
Directrice de recherche, CNRS – Université Lyon 1
Présidente

Thibault, GAJDOS
Directeur de recherche, CNRS – Aix-Marseille Université
Rapporteur

Yaffa, YESHURUN
Professeure, University of Haifa
Rapporteure

Stephen, FLEMING
Professeur, University College London
Examinateur

Claire, SERGENT
Maitresse de conférences, Université Paris-Descartes
Examinatrice

Pascal, MAMASSIAN
Directeur de recherche, CNRS – École normale supérieure
Directeur de thèse

Vincent, ROGER DE GARDELLE
Chargé de recherche, CNRS – Université Panthéon-Sorbonne
Directeur de thèse

On metacognition and the dynamics of selective attention

ABSTRACT

Adaptive decision-making requires precise monitoring of decision quality in light of both sensory uncertainty and the variability inherent in cognitive functions. This monitoring, or metacognitive reasoning, can be assessed by relating subjective confidence in a perceptual decision to objective accuracy. The very keys to cope with the variability of the environment may be selective attention, a known modulator of sensory processing, and reliable metacognitive access to attention. The present dissertation investigates the temporal construction of visual metacognition during and after the allocation of selective attention either to a point in time (temporal attention) or to a point in space (spatial attention).

In the General introduction, we begin, in the first section, by proposing an overview of perceptual decision making and confidence, zooming in on one influential framework: Signal Detection Theory. In the second section of the introduction, we set forth the most prominent results in the psychophysics of visual attention, and their potential role in shaping subjective judgments. Finally, in the last section, we review how the current state of the literature addresses the relationship between confidence and attention, and set the stage for the subsequent chapters of the dissertation.

The empirical work conducted during this PhD is presented across four chapters, and their supplementary materials. The order of these chapters was chosen to offer a dynamic view of the relationship between attention and confidence: each chapter leading to an increasingly noticeable attention-mediated divorce between confidence and performance.

This dissertation is then wrapped with a General discussion of those empirical findings, in which we propose an integrated account of the seemingly disparate results through the concept of attentional episodes. Finally, the interests and limits of the present work are also put in perspective via different potential follow-up questions.

Acknowledgements

I would like to thank my two amazing PhD supervisors, Pascal and Vincent. Before getting on that long – and oh so bumpy – road to a doctorate, we may ask ourselves if we picked the right subject, the right lab, and perhaps the right supervisor. I however did not hesitate a second. Together, Pascal and Vincent made the greatest tandem a student could dream to work with. Pascal, thank you for being always available when I was – randomly – dropping by your office for questions. And most of all, thank you for your wisdom: you have this incredible ability to 'feel' the questions that matters, and this *plus* to pure reasoning is foundational for a scientist in training. Thank you also for all these awesome VSS, ECVP and work/not-so-work moments. Vincent, thank you for all the fantastic discussions, writing sessions and lunches together: you have the gift of making all aspects of research a sort of game, where curiosity always win. You never disregarded any of my (sometimes wonky) ideas. Finally, I'm very grateful to both of you for letting me do whatever my heart was set to, but always keeping me in line with what would be best for me.

Thank you also to Laura and her lab who 'adopted' me: Garance, Damien, Camille, Laurie and Mehdi. Laura, thank you for your wonderful supervision, your advices, and your time. You are a wonderful scientist, and I really hope to continue working with you. Even if I was just a 'pièce rapportée', I always felt like a full waves-team member.

Thank you to all the past and present LSP crew, working along your side has made the days bright and fun all year long. Thanks Baptiste, Arthur, Hadrien, Thomas, Ljubica, Sabrina, Claudia, Shannon, Tarryn, Pierre and Nikos for these wonderful (vision) team moments. I would also like to thank Niko and his great lab members for hosting me in Münster.

Finally, I want to thank the ones without whom finishing this thesis would have never been possible. Monica, thank you for these incredible moments and discussions we always have about everything and anything, thank you for your patience and support, reading, and editing these endless pages. To my family, my parents and my sister, thank you for your support from the beginning, and thank you, Tara, for your kindness and your warm welcome during the writing of this dissertation.

TABLE OF CONTENTS

GENERAL INTRODUCTION

"While the mind is in suspense, it is swayed by a slight impulse one way or the other."
— Terence, Andria: The Fair Andrian

On a frigid December morning in 1633, Urbain Grandier, a catholic priest living in the French city of Loudun, is arrested at his home. The magistrate who has brought him the arrest warrant, the baron de Laubardemont, is the King's special envoy to investigate what will be later considered as one of the most notorious case of witchcraft in the history of France. The plot began a year earlier, in the Ursuline convent of the town, where a few of the seventeen nuns and the prioress Jeanne des Anges, started to feel unwell. The illness was quickly labelled as one affecting the mind, because many of the nuns were behaving oddly: they were strolling aimlessly around at night, crying frequently for no apparent reason, and expressing unusual desires. The prioress herself, Jeanne des Anges, said she was suffering from delusions and nightmares, and that her novices witnessed disturbing apparitions within the priory. The name of Urbain Grandier would be finally uttered, and the ecclesiastical jurisdiction alerted. The behaviour of the nuns and the 'unnatural' scenes the investigators would witness the following days inevitably led to the opening of a case against Urbain Grandier for allegation of witchcraft. After a long and procedural investigation (the file amassing a weight of not less than 4,000 pages), Grandier would be condemned to death by burning. The unfortunate was executed the 18th of August 1634.

In these times, France was slowly recovering from recent wounds: the Reformation profoundly marked the country, and the King, a catholic, sought to regain control over the cities still influenced by the Huguenots. First and foremost, the wound was one of trust, and the reconciliation with the populace would only be far ahead. The city of Loudun was the stage of tensions and divisions, to the point of being threatened with destructive retaliation from the King himself. For the historian and philosopher Michel de Certeau, the Loudun possession unfolded at a time of psychological and social mutations, where science and its logical method were progressively replacing bigotry and superstitious knowledge (Certeau, 2005). The metamorphosis is illustrated by the very presence in the investigation panel of the Grandier case, of physicians and apothecaries along with representatives of the church and the crown. There was now a need for objectivity.

In this context, how could a trial based on fictitious allegations and superstitious precepts have led to such a tragic conclusion? The first reason is political, Grandier had a powerful nemesis in the person of the Cardinal de Richelieu, to whom the Baron de Laubardemont was plainly devoted. There is a second, more intriguing reason: a significant part of the jury – physicians included – was convinced of the magical extent of the case. By all means, jurors certainly relied on second-hand accounts of some shady witnesses, but the *experts* in the panel, the experts themselves, *saw* the devil at work. Or did they, really? Rogier, Cosnier, Carré and Duclos, the medical doctors in charge of the case, wrote, after attending an interrogation of the nuns in April 1634: "*Nous avons jugé qu'il y a quelque chose qui dépasse la nature*" ("We have judged that there is something beyond nature"). However, there is something deeply equivocal in the report of the physicians, something that could expand to and even explain the whole case: it is the conflation of perception with judgment. In the words of one of the judges, Mgr. de La Rocheposay "I' didn't come here to *see* if the possession is genuine. I already *knew* it is the case" (Certeau, 2005).

The Loudun trial remains an interesting example of a witness's ability to bring his eyes to the point of lying. This apparent distortion between the percept and its cognitive processing might be considered as the basis of modern investigation in psychology and neuroscience: questioning the difference between believing and seeing, between what passes the retina and what sense one makes of it. Consequently, to understand how the mind makes sense of the world, we need to explore how it deals with uncertainty, what it cannot discern fully, what it needs to fill in the perceptive gaps with. More than two centuries after the Loudun possession, William James wrote, in his notorious *Principles of psychology*: "The brain is an instrument of *possibilities*, but of no *certainties*. But the consciousness, with its own ends present to it, and knowing also well which possibilities lead thereto and which away, will, if endowed with *causal efficacy*, reinforce the favourable possibilities and repress the unfavourable or indifferent ones" (p. 141, James, 1890). In James' words, our actions are the product of our ability to distil possibilities into certainties, a singularly appealing perspective in the era of the Bayesian brain.

The question of how our subjective sense of certainty fluctuates with our state of mind is the object of the present thesis. More precisely, we will zoom in on one aspect of the internal variability in perception an observer may encounter: selective attention. Orienting selective attention to a point in space – much like the physicians inspecting the unnatural gaze of the nuns in the Grandier case – is a fundamental resource to perceptual decision-making. But to what extent do the observers know their attentional state and its perceptual effects? In the first part of this general introduction, we will present the notion of perceptual decision-making, confidence and metacognition: broadly, the percept and the sense of certainty it may bring to the observer. We will then move to the concept of selective attention in vision, to give a brief overview of the topic, and how it can and does affect perception. The general introduction will be concluded by considering the current state of the literature on confidence when selective attention is manipulated, before finally diving into the heart of the dissertation *On metacognition and the dynamics of selective attention.*

1. Confidence in perceptual decision-making

"Every choice has its obverse, that is to say a renunciation, and so there is no difference between the act of choosing and the act of renouncing."
— Italo Calvino, The Castle of Crossed Destinies

Making a decision is a form of abandon, a renunciation in the words of Italian novelist Italo Calvino. This renunciation is sometimes a wrench, sometimes a relief, but the truth is that a world of possibilities is in the blink of an eye obliterated. The world of possibilities, the 'opportunity cost' in the economics literature, and the subjective conception we have of it, sways our everyday decisions and goes as far as forging our hopes and fears. Broadly, our perception of the environment is a never-ending flow of decisions, decisions that are perceptual in essence. A perceptual decision is a decision about what one has effectively perceived, and sometimes, when there is little consensus on the matter, the brain may have to make some renunciations or assumptions. In this view, perception is a question of decisions, and therefore of uncertainty reduction. There is an inherent reason for this uncertain ballet: the best choice never exists in the real word, because of the upper limit the mind has in tracing the probabilities of external events. To the physiologist Hermann von Helmholtz and many contemporary neuroscientists, "the human perceptual system [is] a statistical inference engine whose function is to infer the probable causes of sensory input" (Dayan, Hinton, Neal, & Zemel, 1995). One therefore has to choose, without certainty, to abandon possibilities, constantly.

We thus begin the first part of the introduction by focusing on two aspects of modern psychophysics in the context of the study of perceptual decision making: the how of perceptual decisions and the how much of subjective understanding of these decisions. To do so, we will put together two pieces of the puzzle: how the psychophysicist studies (a) the perceptual decision itself, also referred to as Type 1 (for 'first-order' decision) and (b) the related sense of confidence that goes with it, also known as Type 2 (for 'second-order decision'). The goal of this introduction is not to provide the reader with an exhaustive view of the field, but rather to *zoom in* to the most pertinent aspects of the question for the chapters to follow.

1.1 Signal detection theory and the study of perception

1.1.1 Perception as decision: the SDT framework

The perception of an object in a cluttered scene can be defined as the output of a function segregating signal from noise: for example, the brain has to determine the contours of the object, its belonging to a known family of objects, and its related semantic content, in order to finally infer its probable identity. The challenge for this inferential process is at the basis of what makes the brain such a fascinating and complex apparatus to study. The idea of a noise filtering function transforming uncertain input into response output had led, in the 60's, to the adaptation of detection theory to the field of psychology (Green & Swets, 1966). Signal Detection Theory, in psychophysics, posits that a perceptual decision results from the combination of some sensitivity (i.e., d') and some response bias (i.e., criterion), applied to a given input. The signal and noise probability distributions are assumed to be normal (i.e., Gaussian), and often of equal variance, providing a computationally tractable probability for each sensory evidence level in the decision space.

In the context of an experiment, a stimulus presented to the observer could, for example, be sampled from two possible categories: clockwise (stimulus A) versus counter-clockwise (stimulus B) oriented gratings. Each of these two categories is related to a given probability distribution of evidence (see fig. 1a). The likelihood functions of each of the two stimuli are often assumed to be of equal variance. The distance between the two distribution means (in Type 1 evidence units) corresponds to the internal sensitivity, to how distinguishable the two categories are from the observer's point of view. The greater the sensitivity, the better the discrimination. Presenting a stimulus to an observer will lead to some evidence accumulation: the point on the evidence axis representing a given sample is called the decision variable. There is, however, one last step before effectively converting Type 1 evidence into an actual decision. To respond, the observer has to choose the stimulus to be reported by placing a threshold, or criterion, along the Type 1 evidence axis: any value below this criterion will be classified as favouring stimulus A, and any value above the criterion as favouring stimulus B. SDT focuses precisely on this difference between the criterion – or bias - and the actual sensitivity of an observer.

Sensitivity, or d', can be estimated empirically using a 'hit rate' and a 'false alarm' rate. In our previous discrimination example, both A and B stimuli had the same variance; now, let us also assume that their means on the evidence axis are equal but of opposite signs. As such, it is possible to calculate the hit and false alarm rate using only one of the stimuli as reference. Taking stimulus A as the reference, we define hits as the number of times the observer responded 'A' when the stimulus A was presented (H_n); misses as the number of times the observer responded 'B' when the stimulus A was presented (M_n); correct rejections the number of times the observer responded 'B' when the stimulus B was presented (CR_n); and false-alarms as the number of times the observer responded 'A' when the stimulus B was presented (FA_n). The observer sensitivity is then defined as:

$$d' = \mathrm{Z}\left(\frac{H_n}{(H_n+M_n)}\right) - \mathrm{Z}\left(\frac{FA_n}{(FA_n+CR_n)}\right) \quad \text{(Eq. 1)}$$

Note the Z-transformation applied to hit and false alarm rates: it is the inverse of a normal distribution function, linking our empirical results to the underlying assumptions of the SDT model, two of which we mentioned earlier, the normal distribution and equal variance assumptions. The d' is expressed in terms of standard-deviation units. The criterion ('c'), or response bias, can be calculated as follows:

$$c = -0.5\left[\mathrm{Z}\left(\frac{H_n}{(H_n+M_n)}\right) + \mathrm{Z}\left(\frac{FA_n}{(FA_n+CR_n)}\right)\right] \quad \text{(Eq. 2)}$$

Contrary to the d', multiple measures of the criterion have been proposed in the literature (Macmillan & Creelman, 2005). Other calculations of the bias are the relative criterion ($c' = c/d'$) and the likelihood ratio (β).

For any d', there is an infinite set of hit rate (HR) and false alarm rate (FAR) combinations. We can plot a given d' as a curve in the (FAR; HR) space (fig. 1, b). The curve, relating FAR to HR for a given sensitivity level, has been called the

Receiver Operating Characteristic (ROC). The absence of sensitivity would correspond to a diagonal in the ROC space. When there is no sensitivity, HR and FAR are equal, and the area under the ROC curve (AUROC) is equal to 0.5. For a d' > 0, the AUROC becomes greater than 0.5. Both the AUROC and d' are therefore describing one and the same thing: the sensitivity of the observer. The criterion is embedded in the ROC space: for each point on a given ROC curve (or one given d'), there is a criterion. The criterion is simply the respective weights of hit rates and false-alarm rates along a stable sensitivity value. Figure 1b illustrates different ROC curves. The shape of those curves is the result of the SDT assumptions considered earlier (i.e., Gaussians with equal variance). Going back to our example, a full ROC curve can in theory be extrapolated from a single (FAR; HR) pair. However, this is only true if the assumptions of SDT hold: the calculation of a d' with this method (see Eq. 1) is therefore parametric, and cannot generalize to stimuli or experiments where, for example, stimulus evidence is not normally distributed. Importantly, the AUROC in itself is non-parametric: it does not assume a specific shape of the ROC. It is the SDT extrapolation of the full ROC curve from a single d' value that make the analysis parametric.

1.1.2 Perceptual and decision bias

From looking at figure 1a, the reader will notice that there are two ways the criterion relative value could be shifted as a function of the two evidence distributions: either with the two distributions fixed and the criterion moving, or with the two distributions jointly moving and the criterion fixed on the evidence axis. The evidence axis is concealed to the experimenter, it is an internal metric for which only relative measures (such as d') are available. Therefore, it is not possible, with SDT, to distinguish between a bias affecting perception (i.e., the two distributions) and a bias affecting decision (i.e., the criterion). This is an important point since it constrains the conclusion an experimenter can draw from the data, as we will see in the following sections of this introduction. As long as the experimenter stays agnostic about the exact source of the bias (perceptual or decisional), the interpretation of the criterion should not cause any specific problem.

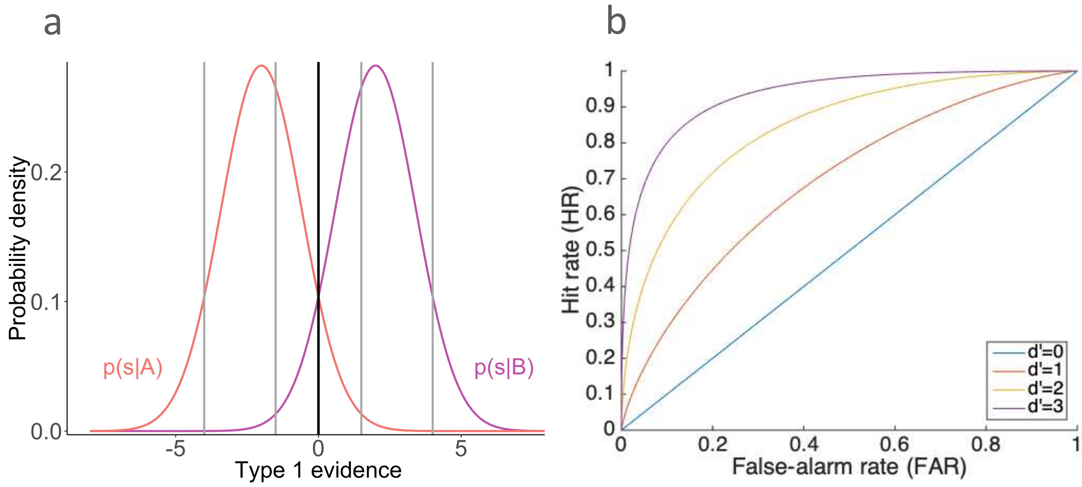

Figure 1. The Signal Detection Theory model. (a) The Type 1 evidence distribution ('s'), for each respective stimulus (A and B). The black vertical line represents an optimal criterion. Any sample of evidence to the left of the criterion will be considered as originating from stimulus A, and any sample to the right will be considered as originating from stimulus B. The vertical lines in grey represent different confidence criteria for three levels of confidence. (b) The figure plots different ROC curves pertaining to distinct sensitivity levels (d'). The diagonal line corresponds to chance level (d'=0).

1.1.3 When confidence was a path to the study of sensitivity

We saw in the previous section that for any d', there is an implied ROC curve determining the combination of hit and false-alarm rates giving rise to a metric of sensitivity (ROC is also known as an isosensitivity curve). The ROC curve allows for an infinity of possible hit/false alarm rate combinations representing an infinity of possible criteria, but all leading to the same sensitivity. This curve, however, is theoretical, and in the example given earlier, is based on parametric assumptions (i.e., equal-variance, Gaussian signals). Historically, the notion of perceptual confidence was brought to SDT with a very specific goal: calculating the empirical ROC or the true observable isosensitivity curve for a given observer (Green & Swets, 1966). The calculation of empirical ROC with this method allowed for observing of departures from predicted ROC when the criterion changes from context to context: this is often due to a violation of the unit-slope property, which assumes the equal weight of each stimulus in the calculation of d' (Macmillan & Creelman, 2005).

To calculate an empirical ROC, an experimenter needs to collect multiple (FAR; HR) pairs at a similar objective difficulty level. To do so, the experimenter can request participants to report the identity of the stimulus and to rate their relative confidence, for example, by adjusting a 4-point rating scale. Type 1 (stimulus discrimination) and confidence judgment responses can be given one at a time, or both at the same time. When both are given at the same time, a discrimination between two stimuli will have 8 (2 stimuli x 4 confidence levels) response options. Whether the second method is still differentiating Type 2 (confidence) from Type 1 responses, however, is a question to which we will return in greater detail later in the dissertation. In the context of SDT, each confidence level can be considered as representing a subjective criterion; it is, therefore, possible to calculate a distinct combination of hit and false-alarm rates resulting in the very same d', for each level of confidence. The value of confidence ratings in SDT has been, first and foremost, to provide quick estimates of sensitivity at various criteria. Confidence judgments became, in the SDT framework, a tool to enhance the reliability of Type 1 sensitivity measure.

We have thus far provided a glimpse at one of the great classical frameworks of psychophysics, SDT, along with an account of how confidence judgments can be used to distinguish sensitivity from bias. However, we have yet to tackle defining what perceptual confidence really is.

1.2 Confidence and metacognition

1.2.1 The many faces of perceptual confidence

Akin to many psychophysical concepts used in contemporary psychology, the notion of confidence judgment as a reliable descriptor of behaviour dates back to the 19th century. This being largely before the invention of SDT. In their *On small differences of sensation,* Peirce & Jastrow observed that confidence ratings directly correlated with actual performance (Peirce & Jastrow, 1884). They even went one step further, and proposed a formula which describes the degree of confidence (m) as logarithmically related to the probability of being correct (p) multiplied by a confidence constant (c):

$$m = c \log\left(\frac{p}{1-p}\right) \quad \text{(Eq. 3)}$$

The log mapping aside, the strong relation between accuracy in the first decision and its related confidence has been confirmed innumerable times since then. This tight relationship justifies the use of rating scales in estimating the ROC sensitivity curve (Macmillan & Creelman, 2005). However, using confidence ratings as a method to facilitate the calculation of perceptual sensitivity does not provide a clear representation of what confidence really is. At first glance, the use of the confidence rating scale could appear to precede the true understanding of its full implications, because confidence seems to be used in SDT often without much scrutiny. Yet, even reading one of the classic manuals of SDT in the original edition of 1966, a psychologist would have detailed implementation of the use of confidence ratings to build ROC curves, but go even farther as to providing references to the importance of confidence to enrich the quality of gathered information (Green & Swets, 1966). Around the same time, publications also point out the particularly rich nature of confidence among other forms of Type 2 judgments. For example, in a JASA paper from 1960, Frank R. Clarke wrote: “It was found that when the listeners were allowed a second-choice identification response, very little information was contained in these responses which was not already contained in the listeners' first identification response. When the second response was a confidence rating, a significant amount of information was added to that which was carried by the identification response” (Clarke, 1960). The notion of supplemental information carried by confidence judgments is not only intuitively appealing – after all, a decision made with great confidence does not equate a low-confidence decision –, it also highlights the possibility that confidence is a form of second-order decision worth studying in its own right.

We have now covered some terrain, and yet the most adequate definition of perceptual confidence remains elusive: is confidence fully reducible to Type 1 accuracy or does it bring some supplemental perceptual evidence to the observer? More specifically, is there a genuine interest in studying confidence apart from the first-order decision itself?

1.2.2 Relating confidence to the decision variable and Type 2 performance

To better picture how confidence can be expressed as a function of Type 1 responses, let us anchor ourselves in the empirical SDT structure we described earlier: confidence ratings are used to estimate a set of different criteria along the evidence axis, while the characteristics of the stimuli signals remained unaltered. A change of confidence is equivalent to a move of the criterion along the evidence axis (fig. 1a), and leads to a change in the distance between the criterion and the decision variable for a given sample. The absolute distance from the criterion to the decision variable can be considered as the confidence evidence or the magnitude of evidence beyond the decision criterion. The greater the magnitude, the greater the probability of the presented stimulus to be effectively sampled from the considered category. This principle can also be understood from another perspective: instead of moving the criterion, we can represent one criterion for each confidence level on the evidence axis (fig. 1a). This way, different magnitudes of confidence evidence will cross different criteria, for example from low to high confidence, allowing the continuous evidence value to be mapped onto a discrete confidence rating scale.

In the situation in which the Type 1 criterion (i.e., the criterion used for initial discrimination) is optimally placed at equal distance from the two signal means, the confidence criteria can be conveniently distributed at different distances from this Type 1 criterion. Confidence ratings will thus reflect the actual discriminability of the stimulus, a result coherent with the performance-confidence correlation found since the 19th century (Peirce & Jastrow, 1884). Here, if both discrimination and confidence are requested at the same time, the Type 1 criterion becomes simply one of the confidence criteria along the evidence dimension. How ever confidence is sampled the Type 1 criterion has up to now been taken to be optimal. Yet, one could also decide to relax the Type 1 optimal criterion prerequisite, building on the (far from unrealistic) assumption that observers in the real-world are not optimal observers (Rahnev & Denison, 2018). The effect of a suboptimal Type 1 criterion shift on confidence, given our initial assumptions, is easy to picture: confidence criteria will shift together with the decision criterion. An observer would thus be biased in both Type 1 and confidence judgments, and the difference between the criterion and decision variable would cease to solely reflect actual performance. If confidence reduces to Type 1 criterion shift, we could rely on SDT to provide an estimate of sensitivity, and confidence would simply represent monotonous differences in response biases. Yet, the situation becomes less straightforward if confidence does not neatly reduce to Type 1 evidence, that is, if confidence uses additional – or partially distinct – sources of evidence.

1.2.3 Pseudo Type 1 and pseudo Type 2

As noted earlier, there is no consensus in the use of Type 1 and Type 2 taxonomy. The nature of this ambiguity has evolved over the years, but the confusion remains.

Since the early days of SDT, we saw that confidence ratings had been used to determine the empirical ROC. This is achieved either (a) by requiring the participants to do a categorisation and confidence judgement at the same time or (b) by requiring the participants to first do a categorisation judgment and then, rate their confidence. Many studies and even manuals, including the well-known Macmillan & Creelman handbook of SDT, present the two methods as nearly equivalent (Macmillan & Creelman, 2005). However, this conflation is problematic if Type 2 is not solely based on Type 1 evidence signal. It is precisely for this reason that some authors refer to the second method (b) as 'pseudo Type 1', because it collapses two judgments into one evidence space and ignores their possible discrepancies (Galvin, Podd, Drga, & Whitmore, 2003).

On the other side of the spectrum lies what we may refer to as 'pseudo Type 2'. 'Pseudo Type 2' is the exact opposite of 'pseudo Type 1', but is still a potential source of misinterpretation. Pseudo Type 2 occurs when experimenters use an all-in-one kind of response, in which both Type 1 and confidence are reported at the same time, but treat confidence ratings as Type 2 responses in their subsequent analyses. The risk, in this case, depends on the definition an experimenter considers 'confidence' to have (Meyniel, Sigman, & Mainen, 2015; Pouget, Drugowitsch, & Kepecs, 2016). If 'confidence' refers to the simple 'readout' of Type 1 evidence, then pseudo Type 2 is not problematic. Yet, if 'confidence' refers to a meta-decision or a decision about a decision, then pseudo

Type 2 is inadequate, because it does not permit two decisions but only one, a Type 1-ish decision. A similar problem arises from 'opt-out' paradigms, where an observer is offered either to give a Type 1 response, or to withdraw from responding ('opt-out'). This method is widely used in confidence studies involving animals, for ease of comprehension (e.g., Kiani & Shadlen, 2009). Yet, it remains pseudo Type 2, in the sense that it partitions the Type 1 evidence space into three, rather than in two distinct options, but does not involve a second decision *per se*.

1.2.4 Confidence as a (real) Type 2 decision

Type 1 SDT does not appear to produce a clear idea of what confidence really means empirically. Without a clear definition, methods must give enough room for potential differences to arise between evidence used in Type 1 responses and evidence used in confidence judgments. To do so, confidence can be objectively related to its own form of accuracy. This accuracy would be as follows: having low confidence in a Type 1 response when this response is incorrect is not considered as erroneous as being highly confident about it. In a similar fashion as for Type 1, confidence has its own accuracy, and can thus be considered a true Type 2 decision, a decision about a decision, when one decision follows the other.

The Type 1/Type 2 dichotomy was first proposed by Clarke and colleagues to distinguish between two types of ROC curves (Clarke, Birdsall, & Tanner, 1959). At first glance, it may seem intuitive to compute a Type 2 d' using the Type 2 FAR and HR, like for Type 1, calculated as follows: hits as the number of times the observer responded 'High confidence' when the Type 1 response was correct ($H2_n$); misses as the number of times the observer responded 'High confidence' when the Type 1 response was incorrect ($M2_n$); correct rejections as the number of times the observer responded 'Low confidence' when the Type 1 response was incorrect ($CR2_n$); and false-alarms as the number of times the observer responded 'Low confidence' when the Type 1 response was correct ($FA2_n$). Then, replacing the respective values in Equation 1 by the aforementioned Type 2 values would lead to a Type 2, bias-free d' measurement of an observer's sensitivity in assessing self-performance. However, this would not be correct. The Z-transformation in Equation 1, which implements the parametric assumption of the SDT model (i.e., the Gaussian shape of the evidence signal) is the main stumbling block. Both mathematically and empirically, Type 2 distributions are far from being normally distributed, *especially* when Type 1 distributions are normally distributed (Fleming & Lau, 2014; Galvin et al., 2003). It is thus discouraged to use or to draw any conclusion on a Type 2 d' calculated in this way.

Yet, from the very same SDT literature, we know of a less restrictive method to estimate sensitivity: the empirical ROC. Given different levels of confidence ratings, it is possible to estimate the empirical Type 2 ROC (hereafter ROC2), but this approach is not without its limitations. There is a strong dependency between ROC2 and the underlying Type 1 ROC. As Galvin et al. pointed out in their paper considering the model assumptions for Type 1 and Type 2 analyses: "An important revelation of the theory is that Type 2 performance can be quite different from Type 1 performance and is highly dependent on the Type 1 criterion. The relationship between Type 1 and Type 2 discriminations depends on the performance measure chosen, the decision axes chosen for each of the two tasks, the Type 1 criterion used, the shape of the distributions underlying the Type 1 decision, and the prior probabilities of the Type 1 events" (Galvin et al., 2003).The psychophysicist thus faces a dilemma: to study Type 1 sensitivity apart from bias, it is possible to use confidence, but to draw a conclusion about Type 2 sensitivity, a similar approach would require the underlying Type 1 sensitivity and criterion to remain unchanged across conditions and experimental manipulations. This type of scenario is quite problematic for an experimenter, as it does not allow for an exhaustive study of confidence and Type 2 sensitivity across variable difficulty levels.

1.3 How to measure confidence and metacognition?

In this dissertation, as in a large part of the literature, the terms Type 2 decision and metacognitive decision, or metacognition, will be used interchangeably. Usually, both metacognitive sensitivity (i.e., Type 2 sensitivity) and metacognitive bias (i.e., Type 2 criterion) contribute to what is defined as 'confidence'. Previously, we presented the need for an extensive study of Type 2 decision and its relative evidence signal. There are two distinct paths to solve this question: the quest for a model that accurately captures the behaviour of Type 2 decisions or the development of a descriptive approach with less assumptions. In the following section, we will discuss both the model-based and descriptive approaches to the problem of Type 2 decision.

1.3.1 Model-based approaches

The Type 2 d', as noted earlier, violates the normality assumption of the underlying evidence distributions, and is thus not a viable candidate for a Type 2 model. An alternative approach named the 'meta-d'' has been developed in the metacognitive literature (Maniscalco & Lau, 2012). The method builds on the main assumption that in the case of an ideal observer SDT model, Type 1 and Type 2 responses are mathematically related. The main assumption relies in taking for granted that the empirical ROC2 has the behaviour of a theoretical ROC2. Thus, with this model, confidence ratings allow for an estimate of what would be the maximum Type 1 d', given the empirical Type 2 ROC (see section 1.1.1). The d' predicted using this 'inverted model' (predicting theoretical Type 1 by way of empirical Type 2) is called the meta-d'. The meta-d' can be compared to the observed empirical d' of the observer, giving a measure of metacognitive efficiency (meta-d'/d'). In this view, a metacognitive ratio of 1 would indicate perfect metacognitive access given the model. To be effective, this model, however, requires two assumptions. In order to invert the generative model giving rise to a ROC2, the meta-d' approach assumes the equal-variance of Type 1 evidence distributions. The meta-d', at least as implemented by Maniscalco & Lau in their seminal paper of 2012, also posits the Type 1 criterion to be fixed. The Type 1 criterion calculated from the empirical data is then injected in the inverted model as a constant (Maniscalco and Lau, 2012). The only parameters estimated through the fitting procedure are the meta-d' and the Type 2 criteria. Despite these rigid assumptions, simulations show this method to be quite robust to change in Type 1 criterion, and importantly, robust to changes in the variance of the evidence used in ROC2 (Barrett, Dienes, & Seth, 2013).

The grounding in Signal Detection Theory of this disseration was done because it is the most common method in the litterature review the reader will find. However, there are other influential frameworks to model perceptual decision-making and confidence. Notably, the SDT model lacks a significant dimension: time. The accumulation of evidence that, for instance, a visual stimulus might be subject to is considered in SDT as occuring *before* the modelisation stage (or not occuring at all): the model remains agnostic as to the exact nature and structure of the accumulation process. One alternative framework models directly the accumulation stage, presenting it as a vital aspect of decision-making. This framework has led to the development of the 'accumulation of evidence' family of models. For example, in a discrimination task involving two stimuli, two accumulators can gather evidence for each respective alternative (Raab, 1962) and the first accumulator to reach a pre-defined bound would trigger a response. Using this approach, Vickers and colleagues defined confidence as a readout of the balance of evidence between the two alternatives (Vickers & Packer, 1982). Of course, the two accumulators do not need to be fully segregated, they could be more or less correlated, or even anticorrelated. When fully anticorellated, they would reduce to one single accumulator in which the evidence sign conventionnaly codes for stimulus category. This anticorrelation model has gain much popularity over the years, termed the Drift Diffusion Model (Bogacz, Brown, Moehlis, Holmes, & Cohen, 2006). All of these models integrate both accuracy and, unlike the SDT model, response times in a common framework, allowing most notably for a study of speed-accuracy trade-off. The 'accumulation of evidence' approach is therefore well-suited to paradigms in which the stimulus lifeclycle is long, as in a random dot motion task (RDM or RDK). The RDM task presents the observer with a cloud of moving dots, with variable degrees of moving coherence. The experimenter can, with this method, manipulate the duration of evidence accumulation, and for example, study how confidence adjusts to such variations (Kiani et al., 2014; Kiani & Shadlen, 2009). Yet, for static stimuli with short presentation time, SDT may be more appropriate, because of the low variability in evidence accumulation across trials the paradigm provides. As such, in the studies we conduct, we will stick to the SDT.

1.3.2 Descriptive analyses and 'model-free' approaches

There is an intrinsic advantage in using a model-free approach to tackle a problem for which the boundaries are still unclear, one in which, as we have touched upon, there is no final consensus on whether confidence evidence is pure Type 1 evidence or not. Dissociations between Type 1 and Type 2 evidence have been found in the literature, suggesting at least partially distinct processing (Fleming & Daw, 2017; Mamassian, 2016). The intrinsic difficulty with a model-free approach however is the capacity to disambiguate metacognitive bias from sensitivity, which model-based do very well. For instance, less conservative techniques have been proposed in the literature, such as simple correlation analyses between accuracy and confidence (Nelson, 1984). The problem of correlations is the conflation between bias and sensitivity: an overconfident observer will not elicit the same correlation coefficient than another observer, despite potentially similar metacognitive sensitivity (Masson & Rotello, 2009).

Sometimes, the solution lies in changing the paradigm, rather than the analyses. In the previous sections, we mainly described cases where confidence was collected using rating scales. By definition, a scale cannot prevent observers from picking their own range of values, a behaviour which can inflate metacognitive bias. An alternative methodology combines the principle of a two-alternative forced choice (2AFC) with a confidence judgment to tap directly into metacognitive sensitivity. In a confidence 2AFC, observers first perform two Type 1 judgments, for two distinct stimuli or trials, and then they have to select which of these two responses they are the most confident about (Barthelmé & Mamassian, 2009; Barthelme et al., 2010; de Gardelle, Le Corre, & Mamassian, 2016; de Gardelle & Mamassian, 2014). This approach has the notable advantage of reducing differences in scale perception and interpretation and thus bias. The data can then be analysed via both model-based or model-free analyses. For example, it is possible to group all trials labelled as 'low confidence' and the remaining trials or 'high confidence' trials, and calculate a d' for each of these two groups, as both of these groups having the same number of samples. Another technique used to reduce bias is to use a different form of rating that capitalizes on incentives. One example is post-decision wagering: using a carefully-designed payoff matrix, it is possible to engage participants in betting on their perceptual decision (Persaud, McLeod, & Cowey, 2007). A form of no-loss gambling, 'Matching Probability', has also been shown to improve confidence estimates (Massoni, Gajdos, & Vergnaud, 2014). There are thus certain methods available for evaluating metacognitive sensitivity without assuming a rigid model.

1.4 Dissociations and confidence models

The multiple methods, such as meta-d' and confidence 2AFC, which have been used to circumvent the problem of Type 2 bias, have uncovered a certain number of situations in which Type 1 and Type 2 decisions dissociate. The notable enthusiasm in the search for dissociations is rooted in the importance of dissociations to the construction of a proper confidence model. A dissociation between first and second order decisions would indicate a difference between the evidence used for Type 1 decisions and the evidence used for Type 2 decisions, making it an interesting topic of study in its own right. Here, we will highlight some of the arguments in favour of a – at least partial – distinction between the two kinds of evidence. Of course, these dissociations and potential differences are for the most part exceptional cases, and the overall tendency is of a close coupling between confidence and perceptual decision. A first very robust finding is the simple possibility of error detection: observers can first respond to some particular perceptual task, and immediately after the Type 1 response, declare they are certain to be wrong. However, the phenomena of error detection and confidence judgments have often been considered as two distinct questions, though empirically and conceptually related (Yeung & Summerfield, 2012). A second set of findings comes from clinical studies as well as animal studies, which have revealed the existence of differences between the two decision types, for example following lesions (e.g., Fleming, Ryu, Golfinos, & Blackmon, 2014; Komura, Nikkuni, Hirashima, Uetake, & Miyamoto, 2013). A third set of studies have identified dissociations by investigating the potential effect of the Type 1 action itself on Type 2 evidence. For example, confidence has been shown to be affected by covert motor preparation, despite stable Type 1 accuracy (Fleming et al., 2015; Gajdos, Fleming, Saez Garcia, Weindel, & Davranche, 2019). Finally, there are a whole slew of other studies that have discovered dissociations when minutely manipulating factors such as attention, context or stimuli (e.g., Baldassi, Megna,

& Burr, 2006; Graziano & Sigman, 2009; Rahnev et al., 2011), or when accounting for developmental and individual differences (Barttfeld et al., 2013; Stephen M Fleming, Weil, Nagy, Dolan, & Rees, 2010; Weil et al., 2013). Rather than bore the reader with a list of all the dissociations, we invite the reader to refer to a very comprehensive review of both the observed dissociations and models of confidence, by Fleming & Daw (2017).

In their review, Fleming & Daw (2017) step aside from an exhaustive analysis of dissociations to makes sense of their implications for our conception of confidence, by proposing a taxonomy of three distinct – albeit related – families of models for Type 1/Type 2 decision-making: (a) the first-order account, in which the same evidence signal contributes to both decisions; (b) the post-decisional account, in which the evidence signal for Type 1 continues to be accumulated after the response is made and therefore affects Type 2 differently; (c) and the second-order account, in which Type 1 and Type 2 evidence signals are segregated from the start, with some potential covariance between them. In their work, the authors argue that the second-order account (c), is the most robust to the multiplicity of sometimes contradictory results in the literature.

In this first section, we have provided an overview of the interests of studying confidence as a form of second-order perceptual decision. We first outlined Signal Detection Theory and its implications to our understanding of perceptual decision-making, and introduced the notion of perceptual confidence as one of fundamental value. We then developed this idea that confidence is a phenomenon worth studying in itself, and we presented some of the methods and tools the literature has been using to better understand confidence, notably through study of dissociations. Though mentioned only briefly, one of the factors known to elicit dissociations between Type 1 and Type 2 decision is selective attention, the factor that we will focus on in our studies. In the next section, we will motive this choice by presenting an overview of the principle of selective attention, and diving into the importance of the study of attention for our understanding of confidence.

2. Visual attention: a psychophysical perspective

"How can one ask the eyes of the body, or those of the mind, to see more than they see? Our attention can increase precision, clarify and intensify; it cannot bring forth in the field of perception what was not there in the first place."

— Henri Bergson, The Creative Mind: An Introduction to Metaphysics

The concept of selective attention, as a product of the mind, has developed progressively. Long before the 19th century's structural psychology of Titchener and Wundt, philosophers offered systematic accounts of the process by which the mind selects and manipulates percepts. Christian Wolff, a contemporary of Leibniz, presented in very descriptive terms the capacity of an observer to divide attention between multiple tasks. He notably cites the example of Julius Caesar, who allegedly dictated four letters while writing a fifth (Wolff, 1732; James, 1890). The definitive emancipation of psychology from philosophy in the early 19th century marked the dawn of experimental psychology: thenceforth, psychology would capitalize on recently developed instruments and machineries to study the mind. The psychologist Wilhelm Wundt thereby extensively used – amongst other things - a modern chronometer borrowed from his mentor, the physiologist Hermann von Helmholtz, to measure the speed of thoughts. Structuralism, which can be considered as the first "school" of modern psychology, pictured the mind as a structure which can be divided into various independent subparts. The goals of this movement, led by Wundt and his pupil, Edward D. Titchener, was to develop a psychophysical understanding of the mind using a blend of precise objective behavioural measures and pure introspection. This last ingredient, which put a significant burden in the eye of the beholder, would be later criticized as incompatible with the principles of the scientific method (Leahey, 1981). In particular, introspection would be accused of lacking any objective ground on which results can be evaluated and compared. These accusations had the unfortunate consequence of portraying psychophysics as an objective method investigating solely objective mechanisms, where introspection shall have, if anything, a relatively minor place. Yet, as we saw in the previous chapters, the objective study of introspection is possible, and has more recently paved the way to fruitful discoveries on how the perceptual system operates.

The question of attention, on the other hand, became a central topic in psychological science at the beginning of the 20th century, along what will be for long deemed as its identical twin: consciousness. As William James put it, “Millions of items of the outward order are present to my senses which never properly enter into my experience. Why? Because they have no *interest* for me. *My experience is what I agree to attend to*” (p. 402, James, 1890). Paradoxically, the field of psychology and recently of cognitive neuroscience often lack an unequivocal definition of attention, and circumvent this dilemma by focusing on what it does, rather than what it is (Anderson, 2011). The notoriously famous quote by James - “Everyone knows what attention is” - remains a perfect diagnosis of the pathology: as human beings, the everyday practice of our attention shall suffice to its essential understanding. James however had offered a definition of attention: “It is the taking possession by the mind, in clear and vivid form, of one out of what seem several simultaneously possible objects or trains of thought. (…) It implies withdrawal from some things in order to deal effectively with others (…)” (p. 405, James, 1890). Of course, this definition would not satisfy a purist of the scientific method, even less so a philosopher, who will inevitably question the meaning of 'possession', 'clear', 'objects' and 'thought'. A contemporary psychologist too may argue that attention might take possession of more than one object simultaneously. Nevertheless, this definition has not necessary changed much since the publication of *The principles of psychology* more than a century ago. One interesting aspect of attention is often overlooked in James’s definition: the notion of withdrawal. This idea of attention as a mechanism which select a stimulus while being detrimental to the processing of other stimuli is still the cornerstone of the definition of attention to this day. To quote the psychologist and psychophysicist Marisa Carrasco, “it is the mechanism that turns looking into seeing. (…) Attention allows us to selectively process the vast amount of information with which we are confronted, prioritizing some aspects of information while ignoring others by focusing on a certain location or aspect of the visual scene” (Carrasco, 2011). The selection, and prioritization of a stimulus deemed relevant is what attention is about: the limited resources that any living organism has at its disposal *de facto* requires some sort of filtering mechanism. In humans and many animals, this selection process can be adjusted dynamically in space and time to privilege the very information needed to maintain homeostasis.

In order to distinguish between the nature of the different objects of selective attention, a rich – sometimes redundant - taxonomy has emerged over the years. For the sake of simplicity, we will focus here on two aspects of attention in the domain of vision, namely spatial and temporal attention. Spatial attention relates to the prioritization and enhancement of a stimulus processing at a particular location (Carrasco, 2011). Temporal attention, on the other hand, refers to the prioritization and enhancement of a stimulus at a particular point in time (Coull & Nobre, 1998; Nobre & van Ede, 2017). As the reader will have noticed, the definition of the visual stimulus however remains largely unspecified here. It is likewise plausible to select a stimulus not by its spatial or temporal aspects, but based on more intrinsic features (such as colour, or shape). This third type of selectivity in visual attention has been coined feature-based attention, and is the subject of a significant body of literature (see Maunsell & Treue, 2006; Carrasco, 2011 for reviews). A last, hybrid version of attentional selection has been called object-based attention (Scholl, 2001). In object-based attention, the selection process is governed by the structure of the object itself (i.e., focusing attention on a rectangular shape facilitates the processing of the stimuli held within it). We will not cover feature-based and object-based attention in the present thesis.

In the visual domain, attention emerged as an important factor of sensory processing. Why single out a mechanism like attention when the transition from fine-grained to coarse resolution - between the fovea and the periphery - seems already to play the role of a focal filter? Would eye movements themselves not be the attentional spotlight? Yet, there is a specific interest in studying perception when eyes are kept still, gazing towards a fixed location. In this situation, even in the absence of significant movement of the eyes, that is, with no alteration of the retinal image, attention can nonetheless affect sensory processing. Orienting attention to an object while gazing at it has been coined overt attention; attending to an object while keeping the eyes focused somewhere else has been defined as covert attention. However, the boundary between covert and overt orienting is not always clear, as it relies on precise experimental monitoring of the eyes. Even with eye-tracking, the role of overt oculomotor activity has been shown to affect seemingly covert attentional process. Notably, micro-saccades (high velocity, involuntary saccades with a maximum amplitude of 1° of visual angle) have been suggested to interact with temporal attention (Denison, Yuval-Greenberg, & Carrasco, 2019) and to explain some of the patterns assumed to be the footprint of covert spatial attention (e.g., Engbert & Kliegl, 2003; Hafed & Ignashchenkova, 2013; Tian, Yoshida, & Hafed, 2016; Yuval-Greenberg, Merriam, & Heeger, 2014). Therefore, the definition of so-called covert attention cannot fully escape the overt orienting spectrum, despite being at the far extremity of it.

In the present chapter, we propose an overview of the psychophysical approach to the question of visual attention. The subject is well too broad for an exhaustive account of the question, and we will mostly concentrate on the behavioural effects of visual attention, with limited, albeit necessary, references to the underlying neural bases at stake. In a first part, we will consider spatial attention, first by presenting the seminal works that led to the 'gold standard' of cueing paradigms, and then by discussing the spatiotemporal characteristics of spatial attention and its effects on perception. In a second part, a similar approach will be dedicated to temporal attention, its relation to spatial attention, and its effects on perception. Finally, we will present the interesting and somewhat ambiguous relationship attention entertains with subjective perception. We thus hope to set the stage for the last sections of this introduction, which will propose an overview of the research on attention and confidence.

2.1 Spatial attention

2.1.1 Seminal work and paradigms

Behind the notion of spatial attention lies a process of selection: attention ought to pull apart one subset of the visual field for further processing. Selecting a particular location with greater acuity should allow an observer to 'see better' at this location, and therefore to report its content more accurately compared to unattended loci. A difference in perceptual performance between an attended versus unattended location is at the basis of the psychophysical study of attention. Initially, attention had been described as a spotlight (Posner, 1980), but the darkness supposed to bathe the remaining part of the visual scene is often relative:

there may be room for seeing outside the focus of attention, as we will see later. In other studies, attention has been compared to a zoom lens, sometimes with coarser granularity in the periphery of its focus (Eriksen & St. James, 1986; Eriksen & Yeh, 1985). The latent principle behind these metaphoric takes is the seriality of the attentional process. By definition there is a theoretical upper limit to the span of attention, it cannot encompass the whole visual scene. This observation has its origin in empirical results from tasks involving conjunction search: when a target is embedded amongst distractors, the time it takes to identify the target is proportional to the number of distractors, suggesting that attention explores the visual scene in a discrete, serial pattern (Treisman & Gelade, 1980). From this perspective, the spotlight metaphor has the advantage of evoking a unique but adjustable resource. However, there is a caveat: while the literature suggests that the size of the attentional window is somewhat flexible, increasing the scope of selection often came at a cost in terms of performance (Eriksen & Murphy, 1987; Eriksen & St. James, 1986). The principle of adjusting an attentional lens in both its location and size with varying degrees of resolution makes the zoom lens an interesting candidate to illustrate the concept of attention selection. However, attention may not simply magnify, but may also trade-off visual acuity between locations. In other words, attending one location in a visual field leads to greater resolution at that location, but decreases resolution elsewhere compared to baseline (Herrmann, Montaser-Kouhsari, Carrasco, & Heeger, 2010; Pestilli & Carrasco, 2005).

In the laboratory, manipulating covert visual attention is usually achieved using pre-cues. The general principle is to present a salient stimulus not long before the onset of the stimulus of interest, to attract attention towards the pre-cued location and facilitate discrimination. This classical paradigm is often referred to as the 'Posner cueing paradigm', from the name of Michael Posner who operationalized the approach in a landmark study in the 80's (Posner, 1980). A typical Posner cueing experiment involves two distinct placeholders, on each sides of a fixation cross displayed at the centre of the screen. On each trial, the participant is presented with a central pre-cue, indicating two possible scenarios: either the pre-cue is neutral, in which case the target is equally likely to appear on both placeholders, or the pre-cue is indicating one location, predicting with ~80% chance where the target will appear. From the experimenter's point of view, there are three conditions: valid, invalid and neutral. It is then possible to compare the response time of the participant in the valid condition or in the invalid condition, to the neutral condition. Compared to the neutral condition which serves as a baseline, participants are typically faster in the valid trials and slower in invalid trials (Posner, 1980). Importantly, this experiment has been conducted on both overt and covert attention, eliciting similar patterns. Over the years, this pattern of results regarding response times has been shown to be very robust, being replicated with different types of cues and targets, and forms of report (fig. 2).

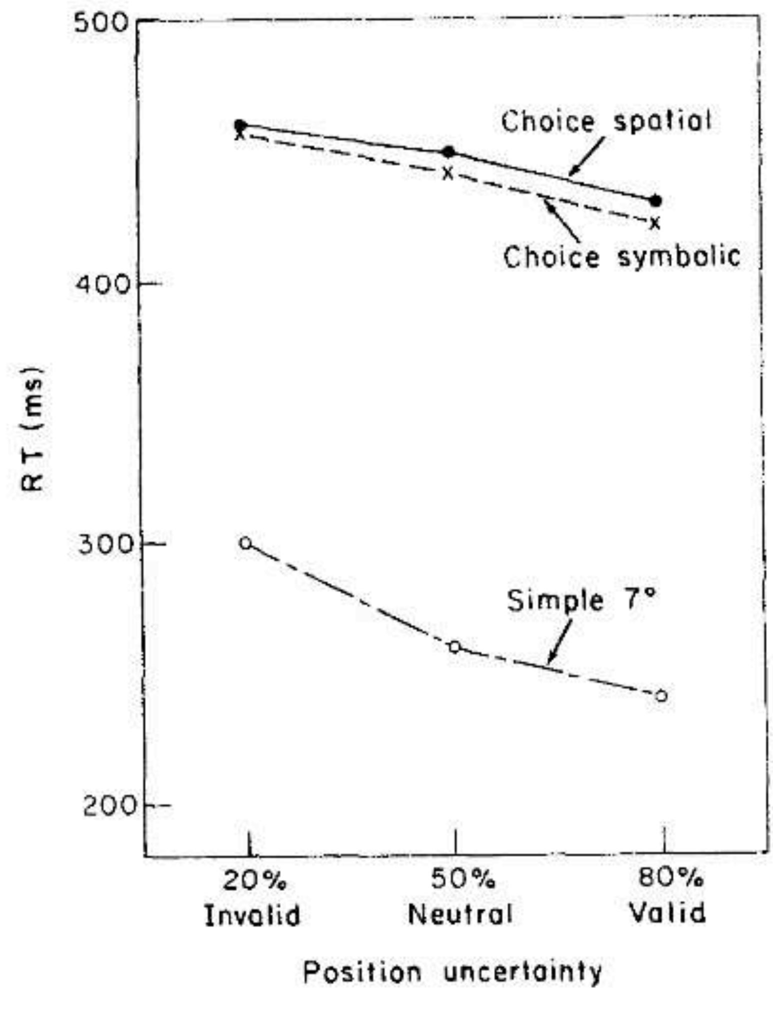

Figure 2. Typical results of a spatial cueing experiment. Mean reaction times are represented as a function of condition (invalid, neutral, valid). The neutral condition occurred 50% of the time, and the valid/invalid conditions the remaining 50% of the time, with an imbalance favouring valid condition (80%). 'Choice spatial' is a task in which participants have to discriminate the location of the target compared to a reference; 'Choice symbolic' is a task in which the

participant has to identify the target as letter or digit; and 'Simple 7°' is a simple detection of a luminance increment task. The RT cost is greater for invalid than neutral trials, and the same is true from neutral versus valid trials, across tasks.

The description of the Posner task given so far, however, is lacking one essential factor: the nature of the cue. In his original publication, Posner used two types of cues: a central and a peripheral one. The rational is that attention can be either oriented voluntarily or attracted involuntary towards a location by a salient event. To manipulate these two seemingly distinct forms of orienting, Posner used either a central cue, such as an arrow pointing towards the location of interest (endogenous cue), or a peripheral cue, like a salient change in contrast appearing nearby the cued location (exogenous cue).

2.1.2 Voluntary and involuntary spatial attention

The principle that attention may be subject to the will, or may genuinely escape from it, is presented by James in these terms: "Attention may be either passive, reflex, non-voluntary, effortless; or active and voluntary" (p. 417, James, 1890). The difference between voluntary orienting of the attention locus and its involuntary capture has been defined later on through the endogenous/exogenous dichotomy. Endogenous orienting of attention refers to a voluntary, often sustained allocation of attention, while exogenous orienting is automatic but short-lasting (i.e., transient). Such a distinction is built on two primary pillars: the role of the will and the role of time.

The characteristics of endogenous attention are a voluntary orienting which can be sustained over time. As we saw earlier, using a peripheral cue - that is, a cue indicating a given but distant location - requires a voluntary component to shift attention. Following an allocation process taking roughly 300ms, voluntary attention has been shown to increase performance in detecting or discriminating the attended stimulus over sustained periods. Importantly, the voluntary nature of the shift makes it receptive to cue reliability, that is, to the actual relevance of cue information (Giordano, McElree, & Carrasco, 2009; Sperling & Melchner, 1978).

The characteristics of exogenous attention are in essence the opposite of endogenous attention: its initiation is involuntary, and its duration is brief. Exogenous attention hinges on salient events and can be allocated in approximately 100ms. Such a rapid deployment has a downside: unlike voluntary attention, exogenous attention is short-lived. Furthermore, there is evidence for exogenous attention being automatic and largely irrepressible: an observer cannot but attend to the transient's location, even if it is detrimental or irrelevant to the task at hand (e.g., Giordano et al., 2009; Yeshurun & Carrasco, 1998; Yeshurun & Rashal, 2010).

2.1.3 How spatial attention affects behaviour and sensory processing

A myriad of psychophysical studies has considered the behavioural effects of attention in the last 20 years. For example, spatial attention has been shown to affect spatial resolution (or acuity), texture segmentation, contrast sensitivity, temporal resolution, and visual search (e.g., Carrasco, Williams, & Yeshurun, 2002; Golla, Ignashchenkova, Haarmeier, & Thier, 2004; Montagna, Pestilli, & Carrasco, 2009; Nakayama & Martini, 2011; Yeshurun & Carrasco, 1998; Yeshurun & Levy, 2003). Three distinct mechanisms are thought to improve performance at attended location: signal enhancement, (external) noise exclusion, and distractor suppression. They all share a common behavioural signature, namely, greater accuracy, but their distinct influence on the later can be estimated using a model-based analysis, and/or electrophysiology. Signal enhancement is the increase in strength – or gain – of the neuronal activity coding for the attended stimulus along the cortical hierarchy, and is thought to be a fundamental mechanism of spatial selective attention (e.g., Carrasco, Penpeci-Talgar, & Eckstein, 2000; Carrasco et al., 2002; Lu & Dosher, 1998; Morrone, Denti, & Spinelli, 2002). In contrast, noise exclusion is the suppression of irrelevant input, a mechanism which thus indirectly facilitates the processing of the relevant input. Finally, distractor suppression may be understood as a form of noise exclusion where the nature of the noise has a distinctive flavour, as we will see in the next paragraph.

In a landmark study, Lu & Dosher (1998) presented participants with a stable signal consisting in an oriented sinusoidal grating, but varied the level of noise in which the grating was embedded. Combining this paradigm with an observer model allowed the authors to quantify the effect of attention on both the signal, and the external/internal noise. They found that attention was, for the most part, acting through signal enhancement and external noise exclusion. This notion of 'exclusion' is to be understood as a filtering process affecting the sensory system early on: only the stimuli within the attentional locus *and* tuned to the filters are passed along the visual system for further processing (Lu & Dosher, 1998; Lu, Lesmes, & Dosher, 2002). On the other hand, internal noise, which pertains to the inherently noisy functioning of the perceptual system, did not appear to be affected by attention.

The distractor suppression account of attention expands the filtering principle to both external and internal noise arising from the spatiotemporal uncertainty about the characteristics of a stimulus among multiple distractors (e.g., Morgan, Ward, & Castet, 1998; Shiu & Pashler, 1994; Verghese, 2001). Here, filtering is not thought as a simple bottleneck hindering the unattended when it goes by, but as an active suppression mechanism that decreases activity of the neuronal population coding for more or less complex distractors. Whether this suppressive activity acts through a decisional or perceptual mechanism, however, is often difficult to tell. To quote Carrasco: "In general, decision-based explanations of attention argue that selection allows observers to 'listen' to useful filters and base choices upon those filters. In this sense, distractor suppression can be seen as an external noise reduction mechanism that operates via a decision template that is moulded around the target attributes" (Carrasco, 2011). However, to isolate criterion shift from actual change in sensitivity, a careful experimental design which – amongst other things – prevents the confounding of cue predictivity and accuracy measure is paramount.

There is an overall consensus on the effect of endogenous attention on both reaction time and accuracy. Here, however, we shall differentiate these two dimensions, often conflated into the single, albeit vague, 'performance' qualifier. This distinction is essential when considering with more scrutiny the way exogenous attention is thought to affect visual processing, since it has sometimes been shown to affect response times to a greater extent than accuracy (Prinzmetal, McCool, & Park, 2005). The consensus on the effect of Posner cueing on sensitivity leaves very little doubt that flashing a transient in the vicinity of an ongoing target promotes contrast sensitivity and spatial resolution. However, the definition of exogenous attention as a selective process has been questioned. More precisely, the effects of peripheral cues on accuracy have been shown to extend to virtually any number of cues and targets, at no significant cost (Solomon & Morgan, 2018; Solomon, 2004). The apparent unselective nature of the process in certain paradigms brought some authors to suggest that the effects linked to exogenous attention might result from a pre-attentive mechanism with a largely unlimited capacity (Eckstein, Thomas, Palmer, & Shimozaki, 2000; Solomon & Morgan, 2018; Solomon, 2004; Spekreijse, 2000). This view, however, remains difficult to reconcile with the literature presented earlier, which proposes that attention trades-off acuity across the visual field.

A just as substantive body of literature relates to the neural and functional bases of spatial attention. From the study of the primate brain, electrophysiologists have postulated the existence of a large-scale network that is involved in the control of selective attention. These regions are the occipital, temporal, parietal and frontal cortices (fig. 3), supplemented by subcortical activity from the thalamus and mesencephalon (Buschman & Kastner, 2015; Corbetta & Shulman, 2002; Kastner & Ungerleider, 2000). The involvement of such a large-scale system in attention processing is supported by multiple clinical studies showing that unilateral damage to the human analogues of these cortical and subcortical sites elicits attentional deficiencies such as visual neglect (Damasio, Damasio, & Chui, 1980; Heilman & Valenstein, 1979; Karnath, Ferber, & Himmelbach, 2001; Mort et al., 2003). In the non-clinical human population, allocating attention to a given point in space has likewise been shown to increase activity in multiple regions including the superior parietal lobule (SPL), the intraparietal sulcus (IPS), and the supplementary and frontal eye fields (SEF and FEF, respectively, see Buschman & Kastner, 2015 for a review).

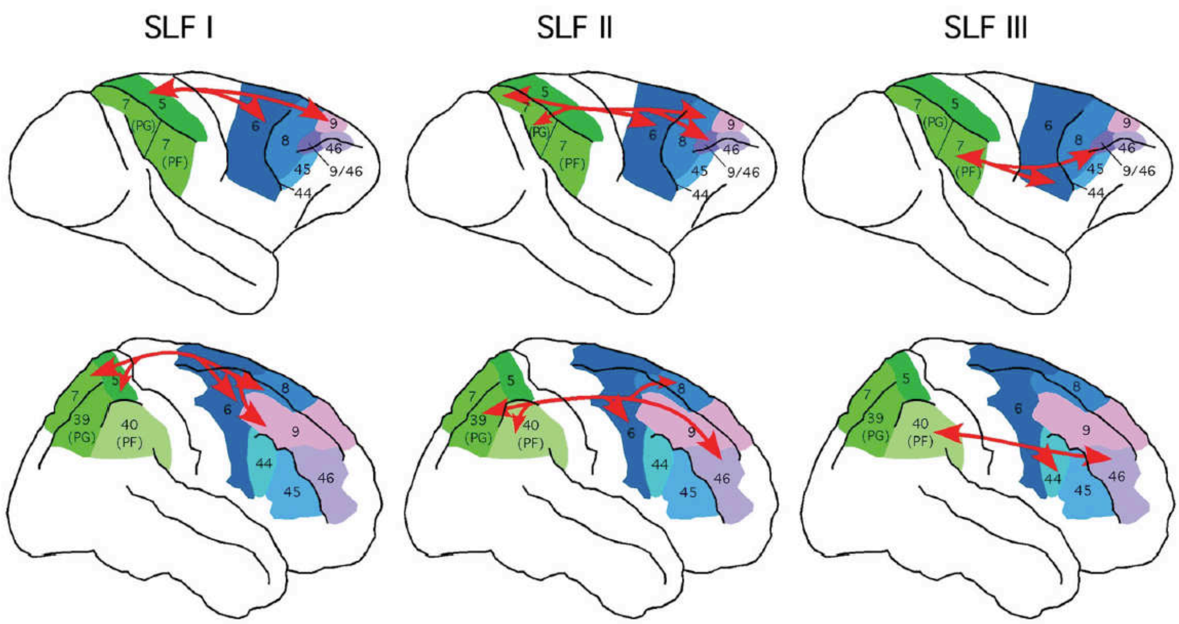

Figure 3. The frontoparietal regions constituting the attention network of the monkey (upper row) and human (lower row) brain. Brodmann areas are identified by their respective number. Each of the three different subdivisions of the superior longitudinal fasciculus (SLF, an associative fibre tract) are represented with the respective projections of the attentional network (red arrows), from dorsal (SFL I) to ventral (SFL III). © The figure is borrowed from Ptak, 2012.

How does this large-scale network reflect the endogenous and exogenous attention dichotomy? In an influential review, Corbetta and Shulman (2002) proposed that two mostly segregated networks for orienting and reorienting cognitive resources existed in humans: the dorsal frontoparietal network, involved in (mostly) voluntary, top-down orienting of attention to features and locations (fig 3, SFL I); and the ventral network (fig. 3, SFL II), dedicated to the detection of unattended, unexpected salient stimuli and involving, amongst other areas, the temporoparietal junction (TPJ) and the ventral-frontal cortex (VFC). This dorsal/ventral taxonomy backed by a large body of literature may appear to be unequivocal support that there are two networks for endogenous and exogenous orienting systems; however, it has likewise been proposed that both systems are one and the same (Peelen, Heslenfeld, & Theeuwes, 2004). Most of the literature, nevertheless, is less hard-lined, suggesting only a partial overlap, that is, a hybrid recruitment of these networks in endogenous and exogenous mechanisms of spatial attention. The primary distinction is often considered to be the directionality of the attentional modulation: while exogenous attention is supposed to be 'stimulus-driven' and to favour a 'bottom-up' activity, endogenous or 'goal-driven' attention should arise from 'top-down' control (Buschman & Miller, 2007). It is also tempting to propose, given its automatic nature, that exogenous orienting recruits subcortical structures to a greater extent than endogenous orienting. Yet, subcortical regions such as the superior colliculus (SC) have been shown to be involved in both exogenous and endogenous attention (Lovejoy & Krauzlis, 2010; Zackon, Casson, Zafar, Stelmach, & Racette, 1999). Despite no clear-cut differential trend on the functional networks of exogenous and endogenous attention, studies considering the time course of brain signal have pointed out differences between the two, with exogenous attention affecting the network earlier than endogenous attention (Busse, Katzner, & Treue, 2008; Hopfinger & West, 2006).

The existence of such a long-range attentional network permits, through interareal synchrony, to regulate the firing rate of the neuronal population coding for an attended location (Buschman & Kastner, 2015; Fries, 2015). At the neuron level, this attention has a direct impact on the properties of receptive fields (RF), that is, on the sensitivity range of the neuron with respect to the stimulus characteristics. Mechanistically, the implementation of such 'upregulation' is, however, far from straightforward: many findings in the literature appear to conflict. Some studies propose that attention acts through a fixed (response) gain factor (e.g., McAdams & Maunsell, 1999), but others show evidence for a shift in contrast gain (e.g., Reynolds, Pasternak, & Desimone, 2000) or find a mixture of response and contrast gain (Williford & Maunsell, 2006). These disparate findings have led to the development of multiple theories. Some have proposed that attention shrinks RF around the attended stimulus (Moran & Desimone, 1985) or that attention biases the competition between neurons in favour of the attended signal (Desimone & Duncan, 1995). Others have suggested that attention scales neuronal activity through a response (McAdams & Maunsell, 1999) or contrast gain (Reynolds et al., 2000), or that neuronal tuning curves are sharpened when attending a given stimulus (Spitzer, Desimone, & Moran, 1988). A relatively recent theory attempted to reconciliate these seemingly contradictory results through the use of the normalization principle, considered to be canonical neural computation (Reynolds & Heeger, 2009). In this model, recently corroborated by empirical evidence (Herrmann et al., 2010), the effects of attention on the RF depend on the size of the stimulus relative to the 'attentional field', the 'attentional field' defining the spatial spread of attention. When the stimulus is large enough to surpass the spatial boundaries of the attentional field, attention would act predominantly through response gain, but when the stimulus is small and the attentional field is large, attention would lead mostly to contrast gain (fig. 4).

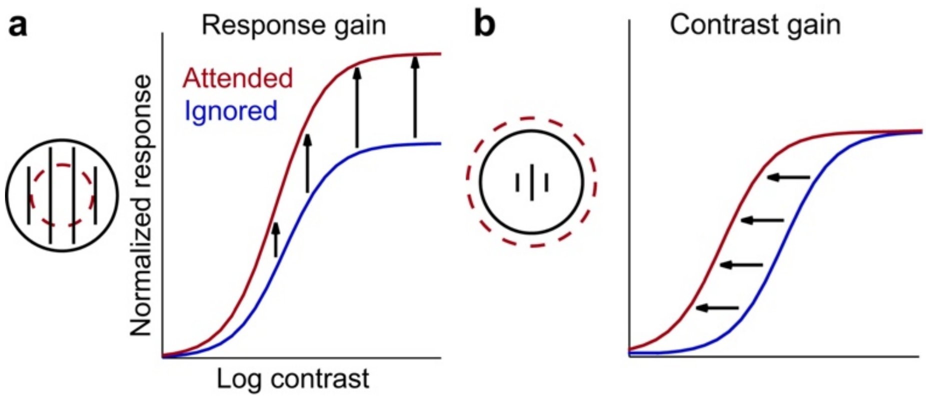

Figure 4. The normalisation model of attention. The two panels illustrate a contrast-response function for a hypothetical neuron. Attending a stimulus (in red) can lead to a shift relative to the RF normalized response (y-axis) or relative to the contrast level (x-axis) depending on the respective size of the stimulus and the attentional field. (a) depicts the situation where the stimulus is larger than the attentional field, leading to response gain. (b) shows the situation where the stimulus is smaller than the attentional field, leading to contrast gain. The ignored stimulus response (in blue) is the contrast-response function of the considered RF when attention is oriented in the opposite hemifield.

© Figure adapted from Herrmann et al., 2010.

The effects of spatial attention on the temporal characteristics of a stimulus are sometimes overlooked, with most of the work in the field being preoccupied by spatial conundrums. Yet, the temporal dimension of spatial attention is precisely one of the two pillars that distinguishes exogenous from endogenous spatial attention. Perhaps the most well-known effect of time on exogenous spatial attention is the "inhibition of return" phenomenon (Posner, Rafal, Choate, & Vaughan, 1985). The inhibition of return refers to the reversing in average response time between the valid and invalid conditions occurring approximately 400ms after cue onset. After this delay, an observer will be faster for targets appearing at the uncued location. This seemingly odd pattern has sparked work on the temporal structure of spatial attention, with theories attempting to posit behavioural relevance (e.g., Chica & Lupiáñez, 2009; Lupiáñez, Klein, & Bartolomeo, 2006; Posner et al., 1985; Wang & Klein, 2010). One theory (Klein & MacInnes, 1999) claims that such a mechanism may be beneficial for an individual, by allowing her to forage across multiple items in the visual scene.

2.2 Temporal attention

2.2.1 Initial paradigm and assumptions

We decided to distinguish spatial from temporal attention, even if the role of both space and time in the orienting process might require a global, unsegregated mechanism. Indeed, the literature has privileged a distinct apparatus for spatial and temporal attention. There are, of course, valid reasons to privilege distinct mechanisms: intuitively, orienting to a point in space and to a point in time are quite different conceptually. This is especially true because the orienting process occurs in one of two different dimensions, one may stay constant, while the other varies, and as such the underlying mechanism may too be distinct. Therefore, we will make the distinction as follows: when attention is oriented to a point in time, but remains at a single point in space, we will use the term 'temporal attention'. When the orienting is both to a point in time and to a distinct point in space, we will refer to it as the temporal aspect of spatial attention.

Perhaps the most well know temporal attention phenomenon is the 'Attentional Blink' (Broadbent & Broadbent, 1987; Raymond, Shapiro, & Arnell, 1992). When presented with a rapid stream of visual stimuli (or "RSVP", see fig. 5a), the selection of a first target (T1) within the stream may prevent an upcoming target (T2) from being accurately reported (fig. 5b). The suppression of the second target, however, is observed uniquely within certain time intervals (between 100ms and 600ms after the first target). Notably, the Attentional Blink has been used to

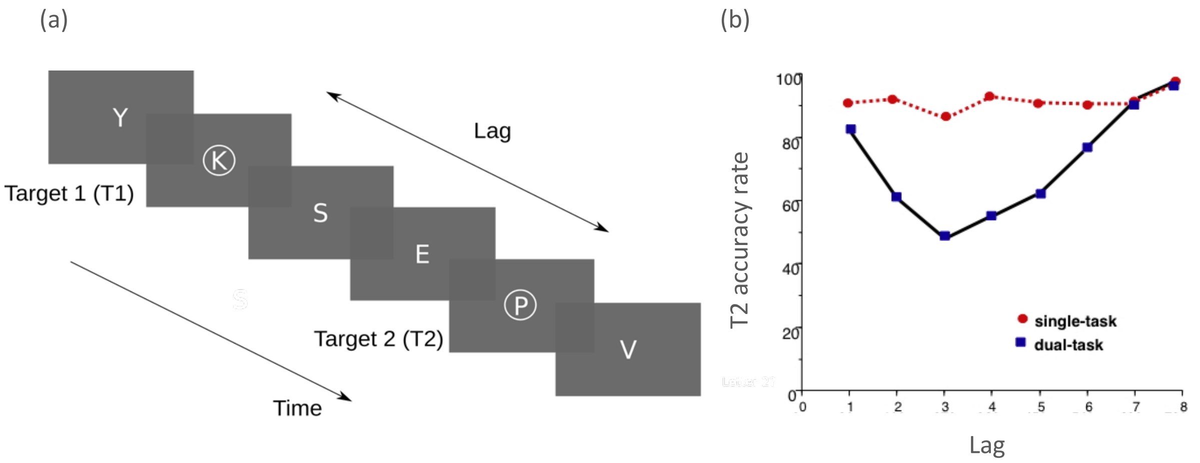

Figure 5. A typical Attentional Blink experiment. (a) Two cued targets (circled letters: Target 1 and 2) are embedded within a rapid visual stream of letters. The number of distractor letters placed in between the two targets is varied (that is, the 'Lag'; with lag-3 depicted here). Each letter appears for 33 ms, and is followed by a 50 ms interstimulus interval. (b) The probability of reporting the second target correctly, when only it is cued for report (in red) and when both targets are cued (in blue). The x-axis represents the lag, that is the relative serial position from the first target. The Attentional Blink refers to the strong drop in performance occurring between lags 2 and 6. Interestingly, however, when the second target is presented immediately after the first, both are accurately reported.

motivate the theory for a limited-capacity, serial processing system (e.g., Chun & Potter, 1995; Duncan, Ward, & Shapiro, 1994). A second interesting phenomenon observed in Attention Blink studies is the initial sparing of accuracy occurring when the two targets are successive (at lag-1, see fig. 5b). This phenomenon has been labelled 'lag-1 sparing' (e.g., Hommel & Akyürek, 2005). To determine the nature of the limitation, a myriad of studies have manipulated factors such as type, distribution and number of targets and delved into fine-grained analyses of the neural mechanisms involved (for in-depth reviews, see Dux & Marois, 2009; Martens & Wyble, 2010). These studies have provided us with a time-sensitive measure of attentional selection capacity, while also identifying limitations of the system and putting forth the notion of 'perceptual episodes' in perception (Snir & Yeshurun, 2017; Wyble Brad, Potter, Bowman, & Nieuwenstein, 2011).

2.2.2 Perceptual episodes and temporal attention

Though it may seem evident that this Attentional Blink is an aspect of temporal attention, it may rather be an illustration of the limitations of the sensory stream (Nobre & van Ede, 2017). This is a particularly salient confound in paradigms in which the exact distribution of report errors in the RSVP remains unknown to the experimenter, for example in protocols involving target selection on the basis of intrinsic differences (such as letter within digits RSVP). Making hypothesis about a potential episodic attention-dependent process in the Attentional Blink paradigm requires extracting the temporal structure of the whole selection episode, something which has been implemented using an RSVP where targets differ from distractors on the basis of incidental features (Chun, 1997; Goodbourn et al., 2016; Martini, 2012; Vul, Hanus, & Kanwisher, 2008, 2009; Vul, Nieuwenstein, & Kanwisher, 2008). Incidental features (such as cued letter within a non-cued letters RSVP), compared to more intrinsic differences (such as letter within a digits RSVP), allow some uncertainty about target identity, and could lead to erroneous selection in challenging situations. In these paradigms, collecting the serial position of each reported error provides supplementary details as to the nature of the selection episode, such as the centre of mass and the spread of the response distribution. The distribution of errors in the case of a single target is usually of Gaussian-like shape centred on the correct letter/item (fig. 6a). However, when considering the distribution of responses for the second target, the shape is dependent on the lag: for lag-1, the distribution is similar to the one of T1; for lag-2 and lag-3, the distribution is heavily distorted, suggesting strong suppression (fig 6, b and c); and for longer lags, the distribution is Gaussian again, but shifted towards later items, suggesting delays in response selection (Goodbourn et al., 2016; Vul et al., 2009; Vul, Hanus, et al., 2008; Vul, Nieuwenstein, et al., 2008). These experimental designs thus allows to probe the frequency of report not only per lag, but also per item positions: changes in the distribution of report around targets can be understood as a footprint of temporal selection during orienting of attention (Chun, 1997; Goodbourn et al., 2016; Reeves & Sperling, 1986; Vul, Hanus, et al., 2008).

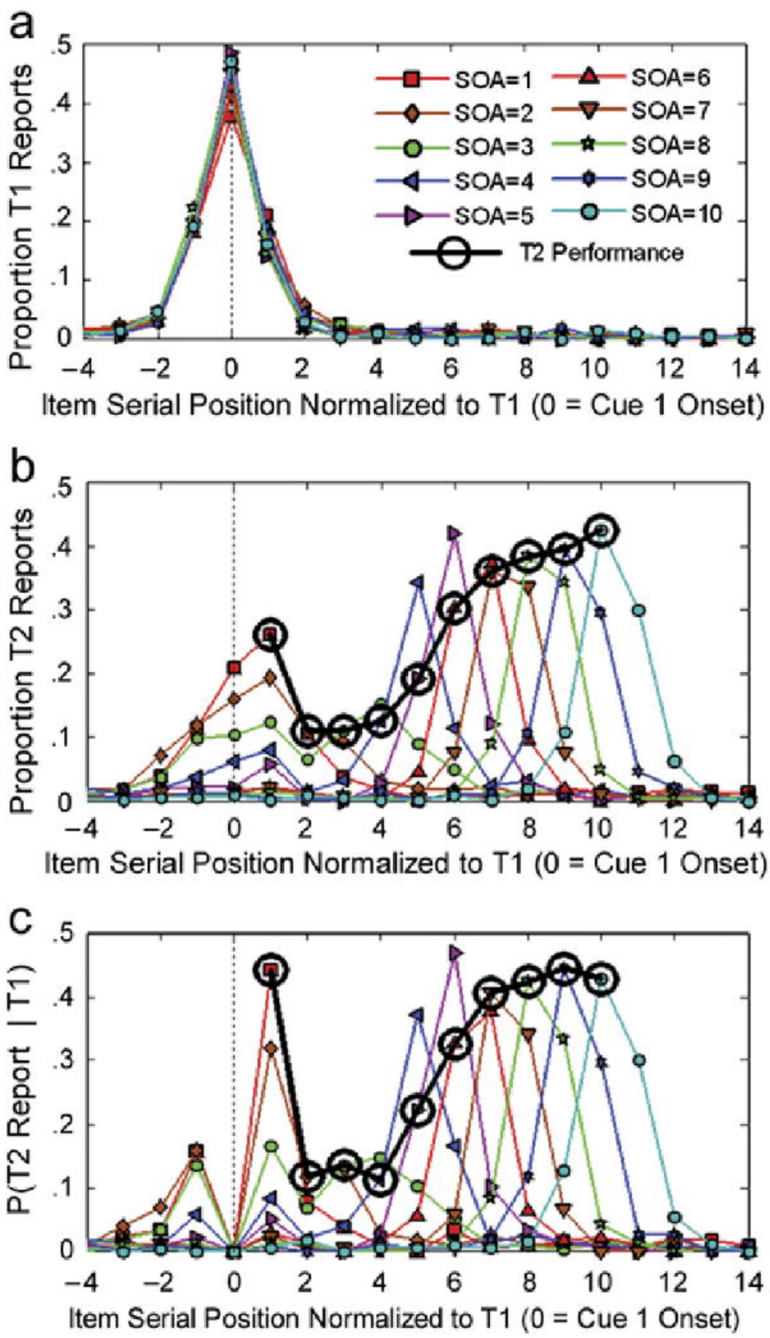

Figure 6. Probing selection episodes in the Attentional Blink. Each panel depicts the probability of reporting a letter for a given position in the RSVP, for T1 (a), T2 (b) and T2 when T1 was reported correctly (c). The SOAs (coloured lines) correspond to each lag, and the black line represents the probability of report for the correct position (that is, when the target is reported). In (b) and (c), we can see a clear delay in selection from lag 3 to lags 8-9.

© Figure reproduced from Vul, Nieuwenstein, et al. (2008).

A wide-spread interpretation of the Attentional Blink findings is that attention must be deployed to a particular time window to be effective, and that it fails to reorient when a second episode is too close to the previous episode (e.g., Raymond et al., 1992; Wyble, Bowman, & Nieuwenstein, 2009; Wyble et al., 2011). This episodic account can also explain the lag-1 sparing effect, where both targets would fall within the same selection episode (Akyürek et al., 2012; Dell'Acqua, Doro, Dux, Losier, & Jolicoeur, 2016; Goodbourn et al., 2016; Hommel & Akyürek, 2005). This proposition is reinforced by the existence of order-reversal trials, where the two targets are accurately selected but reported in the wrong order, as well as binding errors, where two distinct stimuli are fused together (Akyürek et al., 2012; Simione, Akyürek, Vastola, Raffone, & Bowman, 2017). An open question, however, remains: the exact nature of the mechanistic involvement of attention in the shaping of the selection episode. One prominent model of temporal attention posits that these selection episodes are in fact attentional episodes: attention would therefore determine the size, spread and location of the selection distribution according to task demand and available resources (e.g., Wyble et al., 2011). A second line of thought suggests that these episodes are purely perceptual, each with its own fixed characteristics (Martini, 2012), with attention selecting and 'freezing' one of these episodes for later report (Snir & Yeshurun, 2017).

2.2.3 Behavioural and physiological signatures of temporal attention

The study of the Attentional Blink has provided valuable details as to the nature of selection during orienting of attention in time. Attentional selection in this phenomenon, however, captures the 'on-line' process, where the observer only knows which item has to be attended at target onset. As such, the voluntary selection process might be still ongoing after target offset, possibly through the consolidation of iconic or working memory traces (e.g., Akyürek & Hommel, 2005; Craston, Wyble, Chennu, & Bowman, 2009).

However, like spatial attention, temporal orienting could likewise require more or less proactive, and top-down, control. The literature points to the existence of such an exogenous/endogenous dichotomy in temporal attention.

Much like exogenous attention does in space, temporal exogenous attention briefly improves performance in discriminating a stimulus despite its irrelevance for the task at hand (Coull, Frith, Büchel, & Nobre, 2000; Lawrence & Klein, 2013; McCormick, Redden, Lawrence, & Klein, 2018; Rohenkohl, Coull, & Nobre, 2011). Akin to spatial attention, temporal attention has also been recently shown to trade-off performance across points in time (Denison, Heeger, & Carrasco, 2017). For example, when performance is increased at a given time point, it is degraded at other time points compared to baseline.

Behavioural studies have elucidated the various natures of temporal attention, but functional imaging and electrophysiology has allowed us to get a more in-depth understanding of how temporal attention may function on a neural level. (for a review, see Nobre & van Ede, 2017). However, deciphering how temporal attention is able to accurately adjust the sensory system to time contingencies has been hindered by the limited understanding of how the brain computes interval-time range (Muller & Nobre, 2014). Using temporal cueing affected the activity in the left parietal cortex and the left ventral premotor PFC (Coull & Nobre, 1998; Nobre & Rohenkohl, 2014; Wiener, Turkeltaub, & Coslett, 2010). It has been postulated that these sensorimotor regions observed during the temporal orienting of attention might play a role analogous to the one of the oculomotor areas in the control of spatial attention (Morillon, Schroeder, & Wyart, 2014; Nobre & van Ede, 2017; Schubotz, 2007). Yet, to maintain timing precision over a sustained period and across multiple events, the brain needs some sort of internal clock mechanism. One interesting candidate for this complex estimation process are the brain oscillations themselves, proposed to have a critical role in the sampling of visual information and the orienting of spatial attention (e.g., Dugué et al., 2016; Fries, 2015; Landau & Fries, 2012; Landau et al., 2015; VanRullen, 2016).

****

Selective attention appears to be a critical element of visual perception. Therefore, the idea that attention is the gate to conscious access is appealing: after all, selection in time and space is what intuitively defines consciousness. However, the exact definition of what it is to be consciously perceived is harder to discern. This apparent bound between attention and the subjective experience of the world has led many psychologists to consider them as one and the same. However, recent works, returning to the study of introspection initiated by Wundt and Titchener, suggest a more complex relationship. Attention may very well be the key to understand how humans cope with uncertainty: through it, the brain grasps elements of the scene deemed meaningful, and allows them to be consciously perceived.

2.3 Attention and subjective perception

2.3.1 Attention and subjectivity: from early binding to dissociation

How does attention affect our subjective experience in time and space? The idea that our eyes see more than what we subjectively perceive can be found in early modern philosophy. Gottfried W. Leibniz, for example, differentiates between perception, which relates to the continuous sensory flow passing the retina, and apperception, which defines the limited part of the sensory stream that is actually perceived consciously (Leibnitz, 1765). Later, James insists on the fundamental role of (in)attention in removing the background noise from our everyday life: “We do not notice the ticking of the clock, the noise of the city streets, or the roaring of the brook near the house; and even the din of a foundry or factory will not mingle with the thoughts of its workers, if they have been there long enough. (…) Various entoptic images, muscæ volitantes, etc., although constantly present, are hardly ever known. The pressure of our clothes and shoes, the beating of our hearts and arteries, our breathing, certain steadfast bodily pains, habitual odors, tastes in the mouth, etc., are examples from other senses, of the same lapse into unconsciousness of any too unchanging content” (p.455, James, 1890). Unchanging perceptual content, James argues, is what make us inattentive, that is, utterly unconscious of the underlying percepts. But is this really so?

Certain experimental results from the literature show that, indeed, inattention leads to significant 'blindness'. For example, the phenomenon of ‘change blindness’ is one in which makes people are often unaware of significant changes made to a seemingly static image, when the view of the image is disrupted for a brief period of time when change occurs. Change blindness has been observed with a variety of disruptive events, including eyeblinks (O’Regan, Deubel, Clark, & Rensink, 2000), flickers (Rensink, O’Regan, & Clark, 1997) and saccades

(McConkie & Currie, 1996). This effect has even been found with no concealment of the image at all, but by simply superimposing randomly some small, high contrast shapes on the image of interest (O'Regan, Rensink, & Clark, 1999). Another phenomenon is 'inattentional blindness'. In these paradigms, the change can happen in plain view, but remains unnoticed (Simons & Chabris, 1999). The usual explanation for this effect is inattention: when an observer orients her attention endogenously toward a given location or stimulus of interest, the remaining non-attended stimuli are unlikely to reach awareness (Newby & Rock, 1998; Rensink, O'Regan, & Clark, 2000).

These two phenomena point to a fundamental role of attention in shaping the access to perceptual content: only what is in the focus of attention is effectively perceived by the viewer (Rensink et al., 2000). This same would hold for a changing stimulus, as was first proposed a century ago by Helmholtz as the 'law of inattention' (p.455-457, James, 1890): attention should be oriented – or attracted – to the changing stimulus for a change to be noticed.

However, the principle of a tight bond between attention and subjective perception, supported by change blindness and inattentional blindness studies, is questioned by the discovery of dissociations between attention and subjective perception. Early findings in the clinical population, for example, showed the existence of a condition in which a patient, who is clinically blind in a focal part of the visual field, still shows above-chance discrimination performance for stimulus appearing within the impaired region (see Weiskrantz, 1996 for a review). This observation was later confirmed using artificially-induced blindsight with transcranial magnetic stimulation (Boyer, Harrison, & Ro, 2005) and by observations from functional neuroimaging (Leh, Johansen-Berg, & Ptito, 2006). In the normal population, too, dissociations have been found, even in the most foundational of paradigms (e.g., Brascamp, Van Boxtel, Knapen, & Blake, 2010; Crick & Koch, 2003; Kanai, Tsuchiya, & Verstraten, 2006; Mulckhuyse & Theeuwes, 2010; Wyart, Dehaene, & Tallon-Baudry, 2012; Wyart & Tallon-Baudry, 2008).

For example, in priming paradigms, the processing of a stimulus is facilitated by the exposure to a previous, sometimes subliminal, one, indicating a potential dissociation between processed sensory information and the subjective experience of this information. This mechanism has been proposed to necessitate temporal attention, even in case of unconscious primes (Naccache, Blandin, & Dehaene, 2002). Likewise in masking paradigms, a stimulus is blocked from reaching awareness but is still processed by the sensory cortices in cases when a second stimulus is presented approximately 30ms after the first (Breitmeyer & Ogmen, 2010). For example, a form of masking, called metacontrast masking, has been used to induce artificial (or relative) blindsight in normal observers (Lau & Passingham, 2006). However, the role of such paradigm in eliciting a response bias, rather than a change in pure subjective experience has been discussed (Balsdon & Azzopardi, 2015; Jannati & Di Lollo, 2012). Both priming and masking paradigms elicit above chance performance, but with altered subjective experience. However, the notion of 'altered subjective experience' has to be taken with caution here: it is not possible to distinguish a response bias from a genuine perceptual bias in the current context (see Section 1.1.2).

The principle of distinct mechanisms for visual attention and subjective perception relies from evidence derived from situations in which either attention or subjective perception is considered to be absent. Yet, it is difficult to prove the complete absence – or genuine presence - of a process like attention or subjective access. Furthermore, there can been an addition confounding factor, the capacity of the paradigm to accurately capture the participant's subjective percept or of the analysis to accurately interpret the objective and subjective report mechanisms (Balsdon & Azzopardi, 2015; Balsdon & Clifford, 2018; Jannati & Di Lollo, 2012).

As such, a different way to approach the problem of attention and subjective perception is to study how the perceptual content of subjective experience is modulated by the attentional locus. In other terms, rather than trying to remove one element – attention or subjectivity – from the table, it is possible to manipulate one, and track the pattern elicited in the other.

2.3.2 Attention and perceptual appearance

In the previous section, we observed the intricate, and at times complicated, relationship between attention and subjective perception. The difficulty in distinguishing the two processes lies in the conceptual, sometimes blurred, boundaries between the two principles, and the lack of conviction that the two may indeed be separated. To top all that off, the very definition of attention in awareness or consciousness studies takes many forms: sustained, transient, temporal or spatial. The sheer multitude of forms carries the risk of inflating the number of one-shot cases. An alternative approach was considered in a landmark paper by Carrasco, Ling, & Read, 2004. In their study, the authors used an exogenous cueing paradigm to investigate how spatial attention affects a stimulus' appearance (see fig. 7 and its legend for more details on the paradigm). The results revealed that participants judged an attended stimulus as having higher contrast than the unattended one, despite both having the same objective contrast level: attention would alter appearance. This pattern has been shown to be robust over a large range of contrasts, and has been replicated with endogenous attention (Liu, Abrams, & Carrasco, 2009). These findings have given rise to a vast literature that considered attention to modulate subjective perception in multiple dimensions such as spatial frequency and gap size (Gobell & Carrasco, 2005), flickering (Montagna & Carrasco, 2006) or speed perception (Turatto, Vescovi, & Valsecchi, 2007). These results point to the live possibility that attention can alter appearance, and lead to a subjective, not objective, change in appearance at the attended location (for a recent review, see Carrasco & Barbot, 2019). Attention would thus induce a dissociation between subjective and objective experience.

Nevertheless, this paradigm, both original and subsequent versions, too is not without its critics. An equally consistent interpretation of the results is a shift in the decision criterion, or the presence of a 'decision bias' (see Section 1.1.2): the participant is not experiencing an altered percept, but rather changes her decision criterion for non-perceptual reasons. A study that replaced the comparative judgment ("Which target has higher contrast"), like the one used in Carrasco et al. (2004), by an equality judgment ("Are the two targets equal in contrast?") found no effect of attention on appearance (Schneider & Komlos, 2008). A like observation was made against the claim that attention modulates subjective motion speed (Valsecchi, Vescovi, & Turatto, 2010). Yet, the methodology behind these opposing findings too has been debated, with subsequent studies

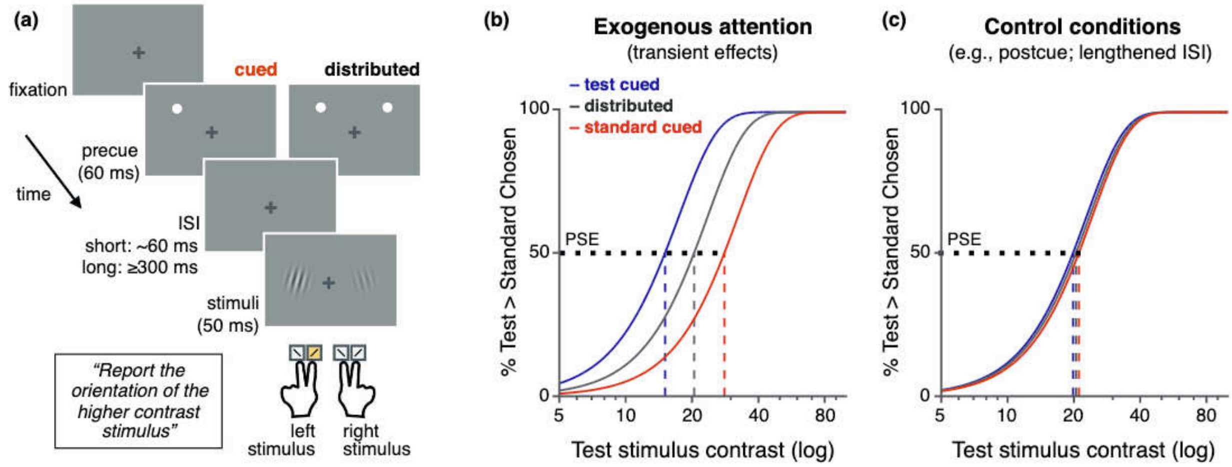

Figure 7. Effect of exogenous spatial attention on apparent contrast. Each trial contains a fixed contrast grating (standard stimulus) and a variable contrast grating (test stimulus, with contrast level sampled from a range of contrasts). (a) On each trial, a salient pre-cue is displayed on one or both sides of a fixation cross. After a variable delay, two low contrast gratings are presented during 50ms. The participant has to report the orientation of the higher contrast stimulus. However, the real variable of interest is the stimulus the participant considers to have greater contrast. (b) The proportion of time when the test stimulus is preferred to the standard as a function of stimulus log contrast. We observe a contrast gain shift for the stimulus preceded by a unique cue at same location (test cued). This demonstrates that the test contrast appeared greater (see the point of subjective equality, PSE) and the standard contrast appeared lower compared to baseline ('distributed', grey line). Notably, discrimination accuracy was also greater for cued location, confirming a manipulation of attention in the task. The red psychometric function represents the proportion of responses when the fixed contrast grating (standard) is cued. (c) Increasing the delay between pre-cue and stimuli, or replacing pre-cues by post-cues does not elicit such a shift.

© Figure reproduced from the recent Carrasco & Barbot (2019) review on the subject.

showing alteration of appearance while controlling for decision bias (Anton-Erxleben, Abrams, & Carrasco, 2010, 2011; Carrasco & Barbot, 2019; Schneider, 2011; Zhou, Buetti, Lu, & Cai, 2018).

However, these criticisms can be somewhat countered with two commonly used controls for the aforementioned appearance experiments: post-cueing and lengthening the interstimulus interval (fig. 7c). Both are conditions catered to making the effect of exogenous attention vanish: in post-cueing, the orienting ought to occur before target onset since the cue follows the target; in lengthened ISI, the target is displayed too late, and exogenous attention has time to disengage. However, post-cueing too has been shown to sometimes induce a similar facilitation effect to pre-cues (e.g., Griffin & Nobre, 2003), comparable to classical iconic memory experiments (Sperling, 1960). Broadly then, cueing attention to a particular location after the stimulus is gone might still enhance accuracy of stimulus report. Accordingly, one may not want to rule out a potential effect of post-cues on subjective experience.

Recent work has identified an effect coined 'retro-perception', where cueing a location after a stimulus is gone facilitates discrimination and increases subjective visibility, up to ~400ms after stimulus offset (Sergent et al., 2013; Thibault, Van Den Berg, Cavanagh, & Sergent, 2016). Interestingly, these results demonstrate that subjective experience too can be influenced ‘offline’ by attentional orienting, likely by retrospectively recovering the very sensory traces that otherwise would have been ignored. In that sense, both the notion of attention and ‘inattention’ ought to be considered in light of an internal perceptual timeline, where the moment of internal access is not necessary in phase with the moment of external presentation (Sergent, 2018).

2.3.3 Conclusion

In this chapter, our goal was to draw up a concise but varied overview of the research on attention. First, we outlined the objective behavioural signatures of visual (covert) attention in both space and time and observed how a selective process is fundamental to behaviour. Second, we discussed the intricate and somewhat complicated relationship that attention and subjective experience have, highlighting the difficulties in detecting and measuring the very distinction between the two.

In the early days of psychology, attention had been studied through introspection, that is, through the subjective impression of focus attention brings to the mind (e.g., James, 1890; Leibnitz, 1765). The boundaries, if any, between subjective and objective experience during attentional orienting however remain unclear. Perhaps strengthening the link between subjective and objective metrics could help: the value of introspection may precisely lie in the potential for an individual to evaluate the result of her actions, to get a step closer to objective experience. Rather than asking for subjective visibility judgments, the experimenter might consider using confidence to tap into the objective measure of introspective ability when attention fluctuates.

3. Confidence and attention

"Subjective sensations are of interest chiefly for scientific investigations only. If they happen to be noticed in the ordinary activity of the senses, they merely distract the attention."

— Hermann von Helmholtz, Treatise on physiological optics: The perceptions of vision (1825).

In Helmholtz' words, the experimenter's curiosity may often overtake what perception is really about. Today, through perceptual confidence the experimenter can study *subjective* impressions in *objective* terms. Attention thus became a viable candidate to probe the limits of introspective ability. Here, we begin by looking at the state of the literature on confidence during attention orienting. As the reader will quickly notice, the field is far from offering an integrated and agreed upon view of the relationship between confidence and attention. We will discuss the role of metacognition on attention and the way in which it can be studied and interpreted theoretically. Finally, we will present a methodological account of the challenges the joint investigation of confidence and attention faces, and set forth concrete solutions, which we will then implement in the following chapters.

3.1 What we know (or do not know)

3.1.1 Confidence is not affected by attention

The first paper to look at the effect of spatial endogenous attention on confidence using a classical cueing paradigm found that confidence remains oblivious to the increase in accuracy induced by endogenous attention (Wilimzig, Tsuchiya, Fahle, Einhäuser, & Koch, 2008). In this study, participants were requested to report the orientation of a sinusoidal grating followed by a mask (see fig. 8 for further protocol details). In order to induce an endogenous orienting of attention, a pre-cue indicated target location in 75% of the trials, and the wrong location in the remaining trials. At the end of each trial, participants had to report the orientation of the target (clockwise versus counter clockwise) and, importantly, to rate their confidence on a 6-point scale. Furthermore, participants were asked to make a speeded response. The authors found no significant difference in average confidence between attended and unattended trials, unlike the significant difference in performance. However, the speeded response has been criticized, because it may not have allowed for a measure of the true confidence level of the participants. (Kurtz, Shapcott, Kaiser, Schmiedt, & Schmid, 2017; Zizlsperger, Sauvigny, & Haarmeier, 2012). To quote Zizlsperger et al. (2012): "the speeded response design may have biased the certainty report, conceivably by forcing decision makers to answer as soon as a minimum level of certainty had been attained or as early as they were ready to give an assessment of their confidence in a decision at all".

Speeded response or not, further studies have found null effects between attention and confidence. One study suggested that the effects of exogenous attention may not at all be incorporated into confidence judgments (Kurtz et al., 2017). In their study, the authors combined a Posner cueing paradigm with a reproduction task to probe both endogenous and exogenous orienting of attention and their respective effects on confidence. Participants were presented with a sinusoidal grating target placed at one of 8 possible locations around a fixation point. A pre-cue, either peripheral or central, was presented before target onset to induce endogenous or exogenous orienting of attention. At the end of the trial, participants had to reproduce the orientation of the target, and report their confidence on a near-continuous coloured rating scale. Their analyses showed no

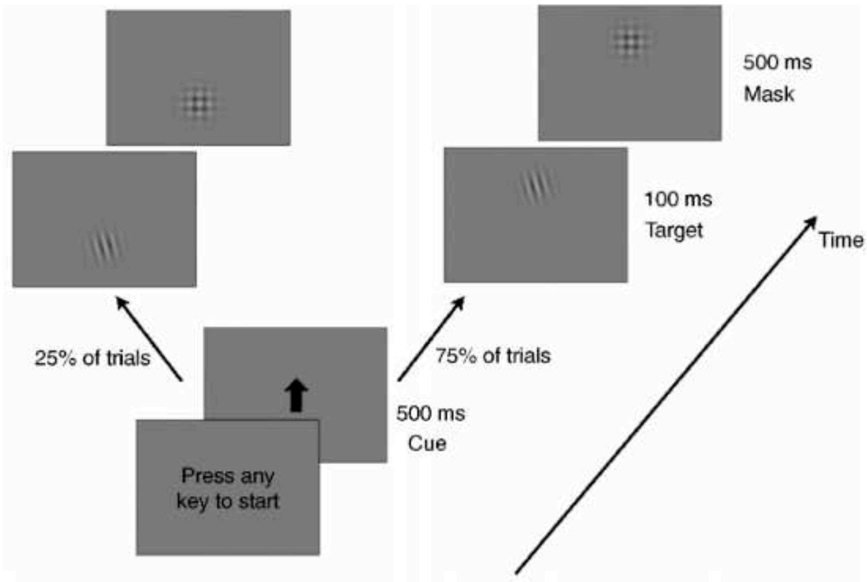

Figure 8. Experimental protocol of the first experiment in the Wilimzig et al. study. Participants were asked to report the orientation of a sinusoidal grating, which had 6 distinct but fixed difficulty levels, followed by a mask. Once the tilt (clockwise or counter clockwise) was reported, participants had to rate their confidence on a 6-point scale. A pre-cue indicated the target location in 75% of trials, inducing endogenous orienting of attention.

© Reproduced from Wilimzig et al. (2008).

effect of exogenous attention on confidence, but an effect of endogenous attention on confidence. However, the paradigm itself may have been the source of this null effect. Their implementation of the cueing method might have precluded a true exogenous orienting of attention for two reasons. The first reason pertains to task-relevance: exogenous, automatic orienting of attention is considered to be irrepressible, and to be activated by unpredictive cues. The present study used 100% predictive pre-cues, when the control to ensure the exogenous nature of an attentional manipulation is to use fully unpredictive pre-cues (Carrasco, 2011). The second reason is relying solely on time course to validate exogenous cueing. As the authors themselves acknowledge: "One limitation […] is that there is no possibility to validate post-hoc that the experimental manipulations really resulted in the deployment of different forms of attention. This claim relies on the well documented differences in the time courses between exogenous attention and endogenous attention" (Kurtz et al., 2017). However, this lack of possibility is limited to this paradigm. A possible solution to this problem would be to use multiple cue-to-target intervals, and to look for a decay in exogenous effects at longer time frames. Thus far it seems that confidence may not be affected by attention. However, the potential confounds make it difficult to extract anything more than a hypothesis.

3.1.2 Confidence decreases with attention

How could an observer be less confident in that to which she attends? As counterintuitive as it sounds, confidence has been showed to decrease with selective attention in certain paradigms.

One study observed a negative relationship between attention and confidence when using visual clutters (Baldassi et al., 2006). The authors adapted a classical visual search task where a target was either presented alone, or amongst multiple similar distractors. After stimuli offset, participants were asked either (1) to indicate the orientation of the target on a discrete, finite scale, (2) to reproduce the orientation (by adjusting a probe to match the target orientation), or (3) to estimate their confidence in the direction of the tilt on a discrete scale. The authors found that when the target was presented alone, errors were usually of low magnitude, and confidence was low for these erroneous responses. Yet, when the target was embedded within distractors, errors were mostly of high magnitude, but led to greater confidence. However, in this study attention had a particular definition: more distractors, less attention. While adding distractors ought to increase location uncertainty and give rise to attentional foraging patterns, the very state of spatial attention in the cluttered environment is difficult to predict. Nonetheless, this result does demonstrate that lower spatial uncertainty, probably mediated by selective attention, could lead to lower confidence and errors.

Another study examined the effect of selective attention on confidence using an Eriksen flanker task (Schoenherr, Leth-Steensen, & Petrusic, 2010). The Eriksen flanker task consists in a target stimulus surrounded by flanking stimuli (Eriksen & Eriksen, 1974). In their paradigm, Schoenherr and colleagues presented the target and distractors for only one frame (~17ms) followed by a short blank and a mask, to increase the probability of errors. In half of the trials, participants were requested to rate their confidence on a 50% to 100% rating scale. In Eriksen flanker tasks, the observer has to categorise the target according, for example to two predefined sets (‘F’ and ‘H’ with one key, ‘P’ and ‘N’ with the other key). In this example, an 'F' letter can be surrounded by either 'H's (same category, congruent), 'P's (opposite category, incongruent), or '#' (neutral). When the target and flankers were congruent (i.e., belonging to the same category), the participant’s response were on average more accurate and faster than in the neutral condition (i.e., ‘F’ within ‘#’s). However, when target and flankers are incongruent (i.e., ‘F’ within ‘P’s), response times and errors were higher than in the neutral and congruent conditions. Confidence however followed a different pattern. The authors found that for incongruent trials, participants were significantly overconfident, that is, showed greater confidence but lower accuracy compared to congruent trials. Notably, the experimenters also matched accuracy levels between congruent and incongruent, and still observer overconfidence. The authors thus present this result as an illustration of introspective failure when selective attention is absent.

When looking more specifically at a narrowly defined type of attention orienting, Rahnev and colleagues provided evidence that conservative subjective bias is induced by endogenous attention (Rahnev et al., 2011). The study capitalized on the possibility of mitigating the effect of attention on accuracy by equating sensitivity between attended and unattended location. In the first experiment, the authors showed that for a simple detection task, the participants’ decision criterion was higher for cued compared to uncued location, suggesting that participants were less prone to report stimulus presence at attended location (fig. 9a). A second experiment was conducted to control for the difference in physical contrast due to the sensitivity matching procedure. While sensitivity matching was thus not possible anymore in this second experiment, they still observed a conservative shift in the criterion for the attended location (fig. 9b). Given the attention-induced conservatory shift to the detection criterion, the authors further tested the effect of attention on Type-2 visibility judgment. They replaced the detection task with a discrimination task, and equated sensitivity between conditions via contrast adjustment. A visibility rating scale was added at the end of the trial. They found that attention lowered subjective visibility, mirroring the effects found on the criterion (fig. 9c). Finally, as in the previous experiment, they verified that equating contrast levels between cued and uncued

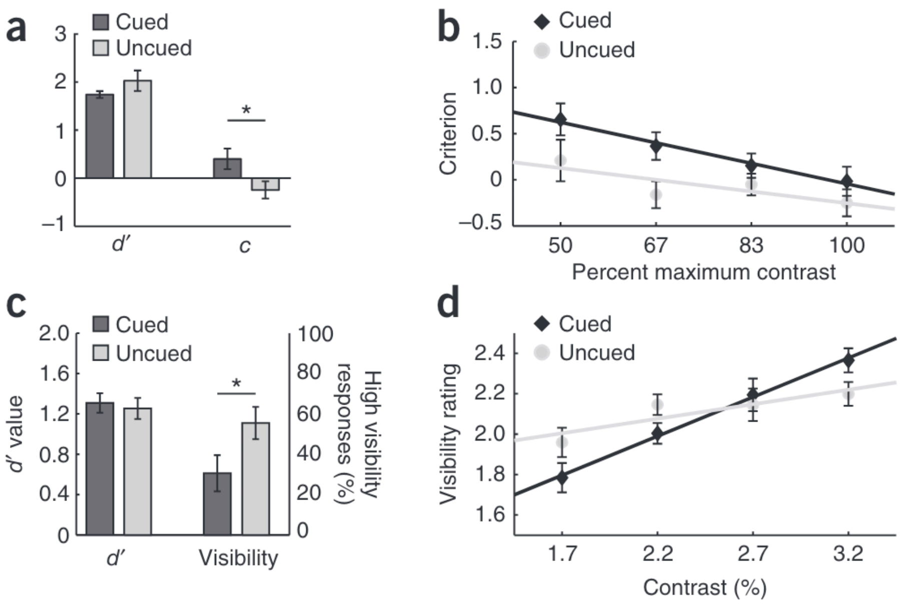

Figure 9. Attention induces conservative biases. (a) In experiment 1, sensitivity (d') is roughly matched between cued and uncued condition, but the criterion is lower for the uncued condition. (b) In a control for experiment 1, the contrast (not the sensitivity) is fixed between conditions, but the difference in criterion remains similar across all contrast levels. (c) In experiment 2, the task was changed to a discrimination task (left versus right 45° orientation), and sensitivity was equated between the two conditions. Visibility ratings were collected and the results show that visibility was lower for the cued location. (d) A control for experiment 2 presented variable sensitivity between locations. At low contrasts, attention elicited lower visibility, but the opposite was true for high contrasts.

locations didn't alter the pattern observed for subjective visibility. Interestingly, when considering equal contrast levels between cued and uncued locations, they found that the conservative effect of attention on visibility was level-dependent: when contrasts were low at both locations, visibility was greater at uncued location, but the pattern reversed for high contrasts (fig. 9d). Taken together, these results suggest that at low contrasts, attention causes a conservative bias in subjective perception.

At first glance, this last study appears to probe visibility judgments and have little to do with confidence. However, a conservative criterion shift in a detection task has been interpreted as a decrease in confidence (Macmillan & Creelman, 2005), and a follow up study by Rahnev et al. verified the effect on confidence directly.

To directly test the role global fluctuations in attention may have on confidence, the authors looked for a potential negative relationship between the spontaneous activity in the dorsal attentional network (see the Section 2.1.3 of this introduction for more details on the architecture of this network) and confidence (Rahnev, Bahdo, de Lange, & Lau, 2012). Increased activity in the dorsal network correlates with a state of focused attention, while decreased activity the opposite. The results revealed that when the attentional network was in a phase of increased functional activity prior to stimulus onset, confidence in stimulus discrimination (that is, giving the direction of a random dot motion cloud) was lower on average. This difference is not reducible to a simple change in accuracy, because pre-stimulus activity was not predictive of performance. This second, physiological, result provides supplemental evidence in favour of a negative relationship between attention and confidence. Nevertheless, the definition of attention between the psychophysical (Rahnev et al., 2011) and the neuroimaging studies (Rahnev, Bahdo, de Lange, & Lau, 2012) can vary drastically: in the former, attention was defined as the endogenous spatial orienting of attention, and in the latter, as a much broader, possibly diffused, state of overall vigilance. In sum, it seems that attention can actually decrease confidence, but care needs to be taken when interpreting results because of varying interpretations and definitions of attention.

3.1.3 Confidence increases with attention

Finally, we turn to the hypothesis that rings the most intuitive: confidence increases with attention. Such a positive relation between attention and confidence stands in manifest contrast with the work presented earlier, but is not devoid of empirical evidence.

Zizlsperger and colleagues found a positive relation between attention and confidence when investigating the relation between spatial or feature-based attention and confidence using a random dot motion (RDM) task (Zizlsperger et al., 2012; Zizlsperger, Sauvigny, Händel, & Haarmeier, 2014). The details of the experimental protocol are given in the legend of fig. 10. Here, we will focus on the studies regarding spatial attention, but the authors found the results to hold for feature-based attention. They showed that attention induced a shift of accuracy rate and confidence in the same *direction* – both were positively correlated – and that the increase in accuracy and confidence was not of the same *magnitude*. The second aspect, the authors argue, is embodied in the observed dissociation between the size of adjustment for confidence versus the one for accuracy. While their finding about the direction of the effect seems in line with the known usual correlation between accuracy and confidence, the dissociation in magnitude should be considered with more caution. For the magnitude comparison, the authors used a z-score transformation for confidence ratings and accuracy, presuming a one-to-one relationship between the two. However, it is difficult to defend a strict one-to-one mapping between the metacognitive – or confidence - space and the decision space, a question defined as the confidence calibration problem, or bias (see the Section 1.2 of this introduction). Yet, they checked that the confidence ratings were relatively equally distributed, with no significant aggregate around some values or metacognitive bias, a control which might alleviate the risk of false positive.

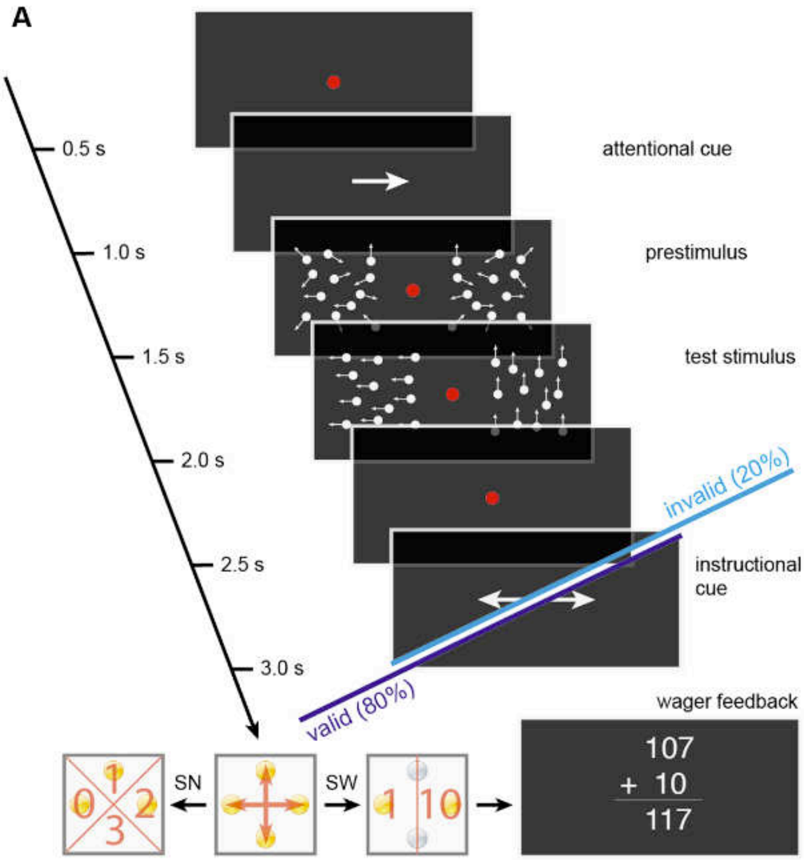

Fig. 10. Design of the Zizlsperger et al. experiment. The participants were presented with two RDM on each side of the fixation point. On each trial, a central cue indicated with 80% validity the RDM which would have to be reported. First, there was a Type 1 response, then a Type 2 response. RDMs were presented with random motion coherence, followed by 2s of different coherence levels (a variable difficulty level sampled around fixed ranges). After stimuli offset, a prompt marked the side of the RDM which have to be discriminated. There were four different possible directions (up, down, right, left). After their Type 1 response, the participants were grouped such that half had to estimate their confidence on a 4-point confidence scale, and the remaining half had to use a post-decisional wagering procedure by placing a wager of 1 or 10 points. In the second paper, only the first confidence rating method was used.

© Reproduced from Zizlsperger et al. (2012).

In the same vein as Zizlsperger et al., a recent study by Denison and colleagues strengthen the evidence for a positive relationship between confidence and attention even further (Denison, Adler, Carrasco, & Ma, 2018). In the study, the authors combined an endogenous cueing task allowing for a flexible decision rule with a model-based analysis to probe whether confidence takes uncertainty into account. While the overall experimental design was fairly comparable to classical cueing tasks, the nature of their Type 1 decision was different: instead of asking the participants to report the orientation, direction or presence of the stimulus, they used a categorization task in which the stimulus belonged to one or two, partially overlapping statistical categories. On each trial and following a valid, neutral or invalid cue, they presented the participant with four drifting gratings around the fixation point. The grating orientations were sampled from two possible Gaussian distributions, with distinct variance: category 1 which had low variance (3°) and category 2 which had high variance (12°). importantly, given that both categories were centred on the horizontal meridian, they both shared a partial overlap, leading to a progressive shift of the optimal choice boundaries as a function of uncertainty. Confidence, too, was collected together with the Type 1 decision (4 levels for each category). In comparison to neutral cues, valid cues induced a greater categorisation accuracy and invalid cues lower categorisation. Accuracy was considered as a function of stimulus orientation, and the authors observed a 'w' shape response profile, reflecting lower accuracy at the category boundaries (fig. 11b, first row), in function of attention. Notably, confidence judgements elicited a similar pattern, with lower confidence at the boundaries, an effect amplified in the valid condition compared to neutral and invalid (fig. 11, second row). More precisely, attention was sharpening the edges of the decision space boundaries, a modulation reflected in confidence judgments.

To summarize, these papers provide evidence for a positive relationship between attention and confidence, with a strong increase in confidence at the attended location, a finding seemingly at odds with the results presented in the previous parts. It is important however to keep in mind the definition of attention in each study, as we have tried to highlight throughout this dissertation: in the Kurtz et al. study presented earlier in this work (see the part entitled "Confidence is not affected by attention"), confidence was found insensitive to exogenous attention, but was positively correlated with endogenous attention, a result compatible with Zizlsperger et al. and Denison et al. studies.

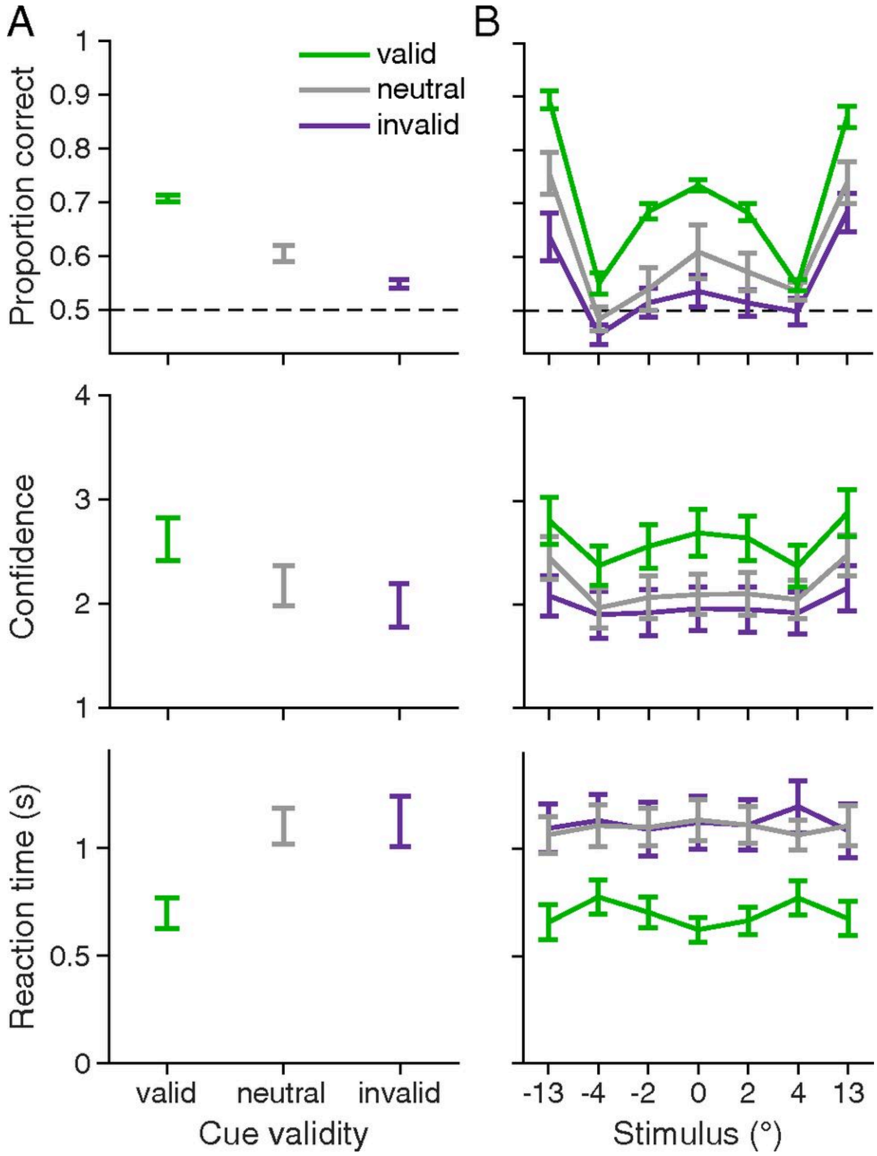

Figure 11. Incorporation of attention-dependent uncertainty into perceptual decisions and confidence. (a) The average proportion correct, confidence, and reaction times for the valid, neutral and invalid cues. We see that confidence reflects attention. (b) The same metrics, this time as a function of stimulus orientation. Notice the drop of performance and confidence at the edges of category boundaries.

© Reproduced from Denison et al. (2018).

3.2 Criterion, models and limitations

3.2.1 Proposed mechanisms for attention and confidence

Though ostensibly counterintuitive, the hypothesis of a negative relationship between attention and confidence has the most empirical support. Out of the nine studies reported here, four of them are suggesting such a relationship, and most of these studies have offered a similar mechanistic account, sometimes in different terms. For example, Baldassi et al. (2006) proposed an SDT-based framework where the candidate orientation of a target grating is selected as the one with maximum evidence from a group of noisy evidence accumulators coding for each presented stimulus. This way, increasing the number of stimuli, and thus likely decreasing attentional focus, would also increase the probability that one of the distractor accumulators has gathered stronger evidence than the target. This increase would lead a unimodal distribution of the internal representation of orientation to become bimodal (see fig. 12 and its legend for details on the model). Importantly, this internal bimodal representation of evidence has a direct effect of on confidence: as illustrated on fig. 12B, the errors part of the distribution is shifted toward a non-zero orientation, which means errors of stronger magnitude coupled with an increased sense of confidence (when we consider confidence to relate to the distance from zero in the internal evidence space).

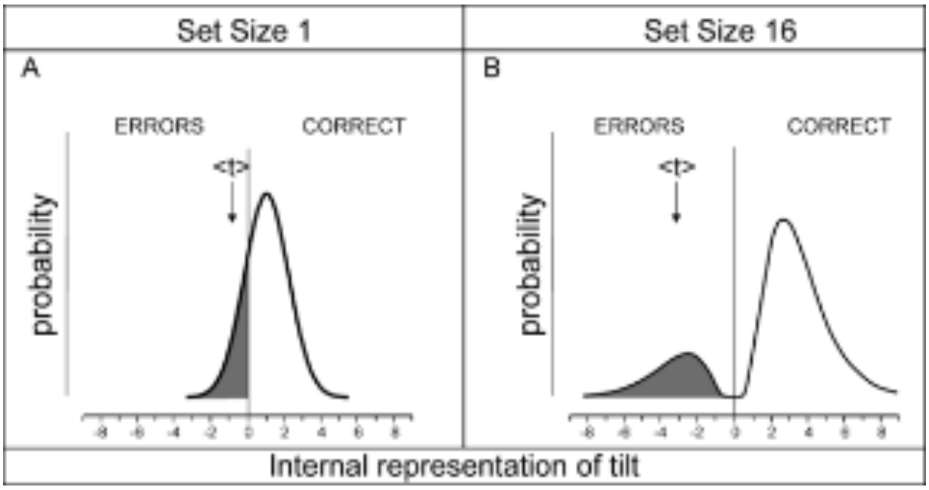

Figure 12. Observer model proposed by Baldassi et al (2006). The authors state, "This model assumes that each stimulus will be analyzed locally by detectors perturbed by uncorrelated neural noise. When the target is presented in isolation, the internal representation of tilt can be described by a probability density function (pdf) well approximated by a Gaussian distribution centered at the physical angle of tilt with a standard deviation equal to the presumed neural noise (A). When the angle of tilt is equal to the standard deviation of the noise, responses will be 76% correct, the usual definition of threshold (detectability index d′ = 1). When distractors are introduced, the situation becomes more complex as observers do not know a priori which stimulus to monitor. Each stimulus should generate a noisy neural representation that can be described by pdfs like that of figure A, but centered at vertical for the distractors. If we assume that the visual system chooses the most tilted of these noisy signals ("signed max rule") then the internal representation of tilt at each trial will be sampled from the bimodal pdf of maxima described in (B)." (Baldassi et al., 2006). © Figure reproduced from Baldassi et al. (2006).

In a similar perspective, but with a clearer definition on how attention affected evidence, Rahnev and colleagues proposed a generalisable account of under-confidence within the attentional locus. They showed that increasing the noise or variance of a given stimulus representation would inevitably lead to a larger part of the signal falling beyond the (unchanging) criterion boundary for high confidence (Bang, Shekhar, & Rahnev, 2019; Rahnev et al., 2012; Rahnev et al., 2011; Rahnev, Maniscalco, Luber, Lau, & Lisanby, 2012). Figure 13 illustrates how such a mechanism would function. This principle can easily be used to explain the negative effect of attention on confidence: given that attention has been shown to increase the signal-to-noise ratio (see Section 2.1.3), reduction in noise at attended location would lead to lower confidence. However, to be effective, this mechanism has two important prerequisites. First, the criteria for both Type 1 (e.g., discrimination) and Type 2 (i.e., confidence) must be fixed. Moreover, the same criteria should be applied to both the attended and unattended location of the visual field, as shown on figure 13 (Gorea & Sagi, 2000). Second, the mean evidence for the signal must remain unaffected by attention. This model however remains speculative, because the literature has not yet reached a consensus on the stability of these two parameters. These assumptions would have to be tested through further modelling and experimental work.

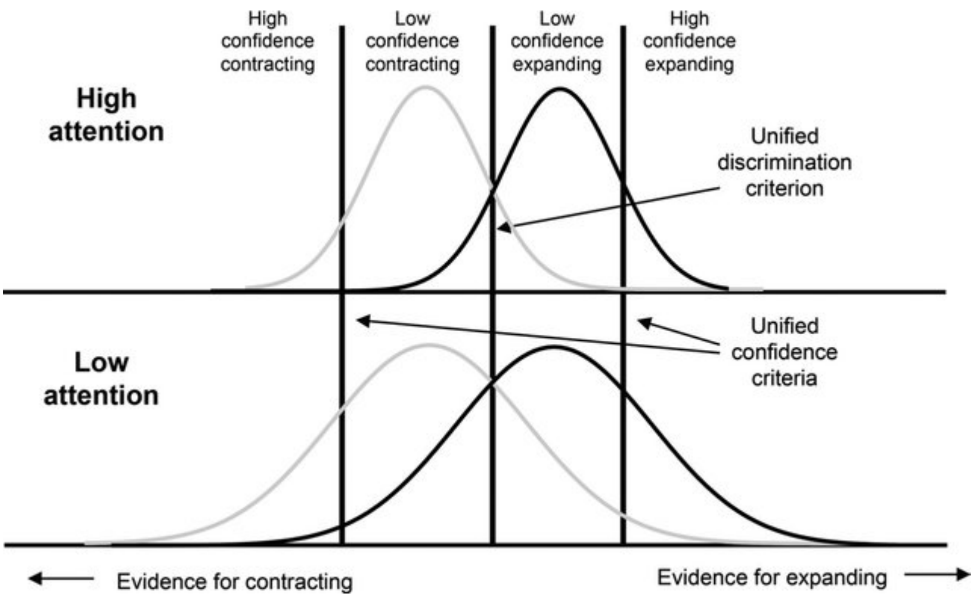

Figure 13. Variance-Reduction model of attention and confidence. A classical SDT approach to a discrimination task (here, the task was to discriminate between a contracting and an expanding RDM); the two stimuli elicit two Gaussian-shaped evidence signals (grey and black curves). The observer applies a discrimination criterion (central vertical line) for Type 1 judgment (contracting versus expanding), and a confidence criterion for each stimulus category (high and low confidence,). When assuming fixed criteria for High attention (attended) and Low

attention (unattended) locations, the increase in variance for the Low attention condition inflates the probability of high confidence. Note that attention does not affect the mean of the signal strength in this model.

© Figure reproduced from Rahnev et al. (2012).

The criterion, thus far, has been considered to be stable; however, it could very well be dynamic. Denison and colleagues use a non-SDT model-based analysis, with a formal model comparison approach unlike much other work in the domain to capture the potential dynamic nature of the criterion. When one thinks model-based one often thinks of less flexible in terms of prior assumption: there approach permits a flexibility the previous SDT-based modelling couldn't afford, a dynamic criterion. The rational concerns the incapacity of SDT to distinguish between a fixed and flexible criterion if the characteristics (mean and variance) of the internal evidence distribution are changing. As we saw earlier (Section 3.1.3), Denison and colleagues have had experimental results at the other end of the spectrum: confidence positively reflects the increase in accuracy induced by attention. Their nested categorization task, as we saw previously, allowed them to observe disruption of accuracy and confidence at the category boundaries, where uncertainty reaches maximum (fig. 11). Furthermore, this effect was modulated by attention, and reflected in confidence judgments: such a modulation is the statistical signature of a flexible decision rule, a dynamic criterion. These results highlight the role of attention in criterion adjustments when confronted with variable uncertainty level, and the ability of confidence to reflect such adjustments. They also point out the necessity to remain careful in assuming a unified criterion in attention manipulations.

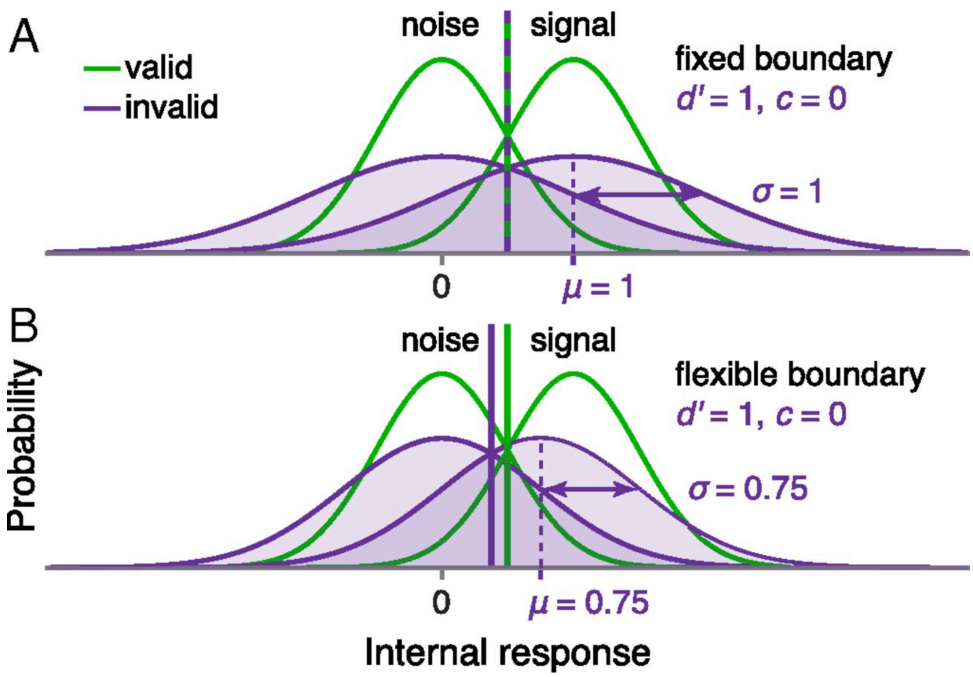

Fig. 14. SDT cannot distinguish the unified from flexible criterion when mean and variance of the signal are dynamic. The green curves represent the noise and signal for the valid (cued, attended) stimulus. The purple curves represent the noise and signal for the invalid (uncued, unattended) stimulus. Vertical lines represent the criterion for the considered condition (valid/invalid). In (A), the criterion is the same for both valid and invalid condition. In (B), the criterion between the valid and invalid condition is different, because of a change in mean and variance of the noise and signal of the valid condition. Note that the position of the criterion remains in both conditions optimal relative to the considered signal and noise (distance from the curves' maxima).

3.2.2 A limiting factor

As different and contradictory as these results may look, there are distinct patterns. We saw in Section 1.1 the importance of distinguishing sensitivity from bias in perceptual decision-making. The very definition of confidence – a decision about the quality of a preceding decision – required the same distinction at the Type 2 level. Therefore, there is an inherent risk in collapsing both metacognitive bias and sensitivity within one and the same the notion of 'average confidence'.

Especially when looking for the sources of evidence used in confidence judgments, metacognitive bias is often unsolicited noise that - if not carefully isolated - may well distort experimenter's conclusion. The way metacognitive sensitivity and bias are defined or implemented in studies is thus critical (Fleming & Lau, 2014), but the studies on joint investigation of both confidence and attention presented here often lack semantic consensus.

The debate on the actual nature of confidence has led to a theoretical distinction between first-order, post-decisional and second-order accounts of confidence (see the Section 1.4 for a non-exhaustive summary of the question). The essential parameter in this taxonomy is time: post-decisional and second-order accounts of confidence involve a subsequent (or parallel but distinct) accumulation of evidence after the Type 1 decision is made. This assumption is not compatible with the SDT-based account of attention and confidence described in the previous section (e.g., Baldassi et al. 2006, Rahnev et al. 2011): confidence evidence is limited to the same evidence signal as the first-order decision (fig. 13). In the study by Denison and colleagues, despite the flexible nature of the decision rules and the possibility for the mean and variance of the perceptual signal to vary, confidence evidence is limited to Type 1 evidence. The coupling of both category discrimination and confidence judgment in the same keypress prevented any potential post-decisional evidence accumulation or action-related influence (see Section 1.2.3). To quote the authors: "Using a single button press for choice and confidence prevented postchoice influences on the confidence judgment and emphasized that confidence should reflect the observer's perception rather than a preceding motor response." (Denison et al., 2018). Their model-based work may thus have asserted a pure first-order origin of confidence, in contrast to the increasing body of literature favouring a post-decisional, or second-order account for confidence (for reviews, see: (Fleming & Daw, 2017; Mamassian, 2016; Pouget, Drugowitsch, & Kepecs, 2016). A first-order account of confidence, however, may or may not capture how confidence is calculated in everyday situations: is it instantaneous with the decision or does it come afterward, or perhaps only if necessary?

There is thus an absence of a real temporal account of confidence in these studies (e.g. Pleskac & Busemeyer, 2010). However, there is also a lack of consideration of the temporal structure of their attentional manipulations. The cueing paradigms used in most of these studies, for example, always involve a fixed cue-to-target interval for all trials. This static approach to the question of attention is built on the known time course of endogenous (~300ms) and exogenous (~100ms) orienting processes, but does not permit any conclusion on the possibly distinct time course of confidence and attention in perception (e.g., Rahnev, Koizumi, McCurdy, D'Esposito, & Lau, 2015). Therefore, the literature currently lacks both an empirical and a mechanistic account on the temporal structure of attention and the corresponding confidence judgments.

3.3 Time, the great absentee

In the last part of this introduction, we will present the challenges faced by the joint study of attention and confidence through the dimension of time. First, we will detail the prerequisites for a model-free, bias-minimising study of confidence, which are necessary to preclude any *a priori* assumption on the first-order, post-decisional or second-order nature of confidence and to set the correct foundations for our knowledge of the attention-confidence time course. Second, we will exploit the rich base of paradigms available in the attention literature to incorporate time, not only as a signature of endogenous versus exogenous orienting, but also as a tool to manipulate and divert the 'boundaries' of an attentional episode, and observe how confidence responds to these distortions.

3.3.1 Prerequisites for the study of confidence and attention through time

To avoid biasing our study of the effect of time on confidence with a priori assumptions, we aim to meet three strict conditions.

The first is to not cut off confidence early in the evidence gathering process. The paradigm must require from the participant not one, but two responses ordered in time, a principle already widely applied in the literature

(Kurtz et al., 2017; Schoenherr et al., 2010; Wilimzig et al., 2008; Zizlsperger et al., 2012, 2014). For example, instead of asking a category judgment for which the keys represent both the category and the degree of confidence in it (Denison et al., 2018), the experimenter could ask two distinct keypresses: one for the category, and one for the degree of confidence.

The second is to clearly distinguish bias and sensitivity in metacognitive judgments. However, this aspect proves to be more challenging, since differentiating metacognitive sensitivity from bias is often precisely based on model assumptions, like in the meta-d' approach (see first part of this introduction). In the following chapters, we will mostly use one approach to circumvent this problem: confidence 2AFC, which has been proven as a reliable, model-free method to isolate metacognitive ability from bias (Barthelmé & Mamassian, 2009; Barthelme et al., 2010; de Gardelle et al., 2016; de Gardelle & Mamassian, 2014). Some work also uses more classical, per-confidence performance aggregates, which can be considered as a rough proxy for metacognitive ability, and will hopefully allow for direct comparison with like experimental approaches. The reader is invited to refer to the Section 1.3.2 of this introduction for more details on the confidence 2AFC technique.

The third, is that we cannot rely anymore on a single generative model to estimate the 'evidence' contributing to confidence, because it would increase the number of assumptions on confidence sources. Therefore, here we will use an alternative method which enhances the information provided per response, at the trial-level. This will be achieved using reproduction tasks instead of detection or discrimination tasks, or use discrimination tasks with a large number of alternatives, which is comparable to reproduction tasks. Reproduction tasks have the advantage of conveying much more information on the magnitude of errors in each trial, such error magnitude being an objective metric on which confidence judgments can be analysed on a per-trial basis.

In respecting these three conditions, we hope to build up a bias-minimising approach to the relationship between confidence and attention, seen through time.

3.3.2 General experimental approach

Instead of waddling through both endogenous and exogenous attention straight off the bat, we zoom first into exogenous attention. Exogenous attention is a particularly viable candidate to probe the relationship between attention and confidence, because of its automaticity. Only one paper considered exogenous attention and confidence, and the paradigm left open questions as to the effectiveness of the attentional manipulation (Kurtz et al., 2017; see the Section 3.1.1). Thus, in the **first chapter** of this dissertation, the results of a simple cueing paradigm, with variable cue-to-target intervals, examining both attention and confidence, will be presented.

In a second chapter, we move to the temporal aspects of endogenous spatial attention. The study of the time course of endogenous spatial attention and confidence cannot be achieved in the same way as for exogenous attention: since the orienting is voluntary, it can be sustained, therefore changing the cue-to-target interval won't provide much variability in accuracy and thus (potentially) confidence. In the context of endogenous orienting of attention, it would be more important to consider the 'end' of the attentional episode, the moment attention *disengages* from one location in order to reorient elsewhere. The paradigm, in this case, shall induce an initial orienting of attention, a disengagement from this initial orienting, and track the signature of this disengagement on both accuracy (Type 1) and confidence (Type 2) measures. In the **second chapter,** we will present an implementation of such a paradigm, where both attention and confidence were tracked with high temporal resolution during their disengagement from a point in space.

In the third chapter, we will be interested in a more 'real-life scenario', where temporal attention is challenged to cope with the speed of change of the visual scene. The questions the experimenter asks, the design of the paradigm, and the independent variables that are manipulated are the framework through which the scientific question should be considered. The risk is, sometimes, to forget the very nature of real-world situations: that is, the fundamental uncertainty of their unfolding. The study of confidence is precisely interesting for the bridge it permits between the objective uncertainty of the world and the subjective goals of the observer. As we saw earlier in this work, one of the main tools for decreasing perceptual variability that observers have at their disposal is attention, but this tool is not devoid of limits. While studies do interest themselves in dissociations, they do not often look at limitations, especially in the temporal structure of attention. Do observers really know when their attention failed them? In the **third chapter** of this thesis, we will present an experiment where we push temporal attention to its very limits, by disrupting or delaying attentional selection, and observe how confidence adapts to these phenomena.

Finally, there is one aspect of the time course of spatial attention that has been often overlooked in behavioural studies: the orienting process itself. In the three first chapters, we considered the results of the orienting of attention in space and time. However, much less is known about the orienting process itself. Most studies used fixed cue-to-target delays (e.g., 300ms), and the very duration of attentional orienting – that is, the time it takes for spatial attention to be allocated – is therefore not investigated at all. Yet, there is an intrinsic variability in sensory processing, and attention does not escape this principle. The 300 milliseconds needed for "attention to be effective" are, after all, an average: sometimes attention is allocated earlier, sometimes later. In the **fourth and last chapter of this thesis**, we will present an experiment where we measure the trial-by-trial variability in the timing of endogenous and exogenous spatial attention. In this final study, we look at the orienting process of attention itself, and its effect on metacognition and confidence judgments.

We thus aim at setting the foundations of a time-based approach to the relationship between confidence and attention.

****

CHAPTER 1 | VISUAL CONFIDENCE AND EXOGENOUS CUES

In the first chapter of this dissertation, we use a canonical implementation of a cueing paradigm to study the relationship between attention and confidence. Thus far, the literature about metacognition and attention did not investigate the effect of exogenous manipulation of attention on confidence judgments. Despite the famous, and importantly highly reproduced, exogenous "Posner paradigm" being successfully applied to study many aspects of spatial attention on both the sensory and cognitive levels, not a single study, to our knowledge, directly combined it with confidence judgments (without potential confounds, see Section 3.1.1 of the General introduction). Yet, the role of exogenous attention on conscious perception and visibility judgments has been studied using many different attentional paradigms, including spatial and temporal cueing: for example, exogenous cues have been proposed to alter subjective visibility (see Section 2.3.2 of the General introduction). In Chapter 1, we thus investigate the effect of exogenous spatial attention on confidence judgments, via a highly reproduced paradigm in the attentional literature.

Visual confidence and exogenous spatial cues

Samuel Recht (1), Vincent de Gardelle (2)+, Pascal Mamassian (1)+

Affiliations:

(1) Laboratoire des systèmes perceptifs, Département d'études cognitives, École normale supérieure, PSL University, CNRS, 75005 Paris, France

(2) CNRS and Paris School of Economics, Paris, France

+ these authors share senior authorship

Abstract

Perceptual sensitivity can be increased shortly after a brief exogenous cue. In an experimental design with a completely uninformative exogenous cue, we asked whether human observers were able to monitor the change of performance induced by the cue. We found that an increase of perceptual sensitivity in the first 150 ms after cue onset was accompanied by an increase in confidence that the perceptual decision was correct. These simultaneous changes in sensitivity and confidence resulted in metacognitive sensitivity that was stable across all delays after cue onset. These results suggest that in spite of exogenous attention being sometimes seen as occurring very early in visual processing, human observers are able to track the changes in performance that follows these attentional effects.

Introduction

Sitting in your favourite spot, you are gazing at the numerous trees of the park, deep into the music coming from your headphones. Then, suddenly, a buzzing object passes through your point of vision: it is too close not to react. Your reaction is spontaneous and quick; you did not have to think about it. When the "bee" finally turns out to be a fly, you go back to your music and soon forgot about it.

Sometimes, the saliency of an event alters perception in a way that enables a quick disengagement from the ongoing task. However, this automatic capture of attention at a particular location in the visual field can be both beneficial and detrimental, depending on the context. Selective spatial attention has been defined as the prioritization and enhancement of a stimulus at a particular location (Carrasco, 2011; Posner, 1980). This selective process can be either exogenous or endogenous, that is either involuntary or voluntary. While exogenous attention has been described as a rapid (~100ms), but short-lasting, bottom-up and automatic enhancement, top-down endogenous attention has a slower deployment rate (~300ms) but can be sustained in time.

Endogenous attention enhances and prioritizes information that is deemed relevant for the observer. Exogenous attention is supposed to be much less task-specific, but comparatively faster: it enables an organism to react quickly to a potential threat. However, this automatic response comes at a cost: the non-specificity of a response grounded mainly on saliency could lead attention to be captured by irrelevant events. In psychophysical experiments, cues made of sharp contrast transients in the vicinity of a following target are often used to trigger exogenous orienting (Carrasco, 2011): a target appearing nearby shortly after cue onset will be on average more quickly (Posner, 1980) and more accurately reported (Carrasco, 2011).

Because spatial attention affects sensory information, being able to introspect on whether attention was deployed is a good indication of the quality of one's own sensory information. This knowledge is particularly useful, as the observer may decide to look longer at an object when the quality of the sensory information is too low (i.e., a nearby moving object could be a bee or a fly). The subjective estimation of the decision's accuracy about a visual stimulus has been coined visual confidence (Mamassian, 2016). Visual confidence has been proposed to play a role in numerous decisional processes, including adaptive learning (Guggenmos, Wilbertz, Hebart, & Sterzer, 2016; Hainguerlot, Vergnaud, & De Gardelle, 2018; Zylberberg, Wolpert, & Shadlen, 2018), information seeking (Desender, Boldt, & Yeung, 2018) and the integration of multiple decision stages (van den Berg, Zylberberg, Kiani, Shadlen, & Wolpert, 2016). Confidence can therefore be considered as an integrative metric for pre-decisional, decisional and post-decisional processing, and it has also been regarded as a form of common, supramodal currency for the perceptual system (de Gardelle & Mamassian, 2014; Faivre, Filevich, Solovey, Kühn, & Blanke, 2018).

While the effect of attention on confidence has also been considered in the literature, the findings are mixed: some studies showed dissociations between accuracy and confidence during manipulation of spatial attention (Rahnev et al., 2011; Wilimzig, Tsuchiya, Fahle, Einhäuser, & Koch, 2008) or temporal attention (Recht, Mamassian, & de Gardelle, 2019). Other studies suggested that spatial attention induces an increase in both sensitivity and confidence (Denison, Adler, Carrasco, & Ma, 2018; Zizlsperger, Sauvigny, & Haarmeier, 2012; Zizlsperger, Sauvigny, Händel, & Haarmeier, 2014). Most of these studies, however, consider endogenous orienting of attention. To our knowledge, only one study investigated the link between spatial exogenous cueing and confidence, finding no evidence supporting the integration of exogenous effects into confidence judgments (Kurtz, Shapcott, Kaiser, Schmiedt, & Schmid, 2017). In this study, participants' reproduction of an oriented stimulus was more accurate when the stimulus was preceded by a peripheral pre-cue, in comparison to a condition where no pre-cue was used, but this increase in performance was not accompanied by an increase in confidence. This protocol however, as is acknowledged by the authors, cannot guarantee a exogenous orientation, strictly speaking, because the authors presented predictive cues, guarantee the exogenous nature of their manipulation.

Observers have been shown not only to monitor cognitive states such as confidence, but also complex cognitive processes, such as attention dynamics during visual search (Reyes & Sackur, 2014, 2017). However, the very nature of exogenous cueing makes it a relevant candidate for testing the limits of introspective access to the state of the attentional system, A particularly interesting case is when the cue is experimentally made task-irrelevant. In this case, there is no a priori reason to favour valid over invalid cues in confidence estimates, given that the cue predictability of the target location is at chance level. Nonetheless, valid cueing triggers a benefit in sensitivity at short cue-to-target latencies, and a good metacognitive observer should in principle be more confident when sensitivity is greater.

Here, we used a canonical exogenous cueing paradigm in which participants had to report the orientation of a low contrast Gabor patch presented at one of two locations (e.g., Pestilli & Carrasco, 2005), followed by a confidence judgment. The target stimulus was preceded by a peripheral cue that was not predictive of the target's location, and the cue-to-target-onset-asynchrony (hereafter CTOA) was varied. Using different intervals allowed us to analyse the temporal dynamics of both sensitivity and confidence. We also investigated whether confidence could accurately track sensitivity dynamics following cue's onset, an ability also known as metacognitive efficiency (Maniscalco & Lau, 2012). in line with previous studies, our prediction was that sensitivity would only be affected by exogenous cues at short (100-150ms) but not long (>150ms) cue-to-target onset asynchronies (CTOAs), and that response times will be faster for valid cues at short (100-150ms) but not long (>150ms) CTOAs (Carrasco, 2011). As for confidence, we hypothesized that, given the close expected relationship between sensitivity and confidence judgments, we should observe an effect of valid exogenous cues on confidence only at short CTOAs. However, given the involuntary nature of exogenous cueing effect, and the unpredictability of such cues, participants' confidence could also not be affected by the cue, leading to a cue-mediated dissociation between sensitivity and confidence at short, but not long CTOAs. We found evidence that confidence tracks the initial gain in sensitivity induced by exogenous pre-cueing and that metacognitive ability remains stable during and after this initial boost of performance.

Material & methods

Participants

Ten right-handed participants were recruited in the French RISC pool of participants. They all provided informed written consent prior to the experiment and received 30 euros for their time. The experiment was divided into three sessions of one hour each, over three different days. The experimental procedure received approval from the Paris School of Economics (PSE) ethics review board.

Stimuli

Target and distractor consisted in two 2° Gabor patches (spatial frequency: 5 cpd; fixed 12 % contrast) with Gaussian envelope. They were displayed at 5-degrees eccentricity from the centre of the screen, on the horizontal midline. A 0.4-degrees fixation dot was presented at the centre of the screen. Target and distractor were always presented ipsilaterally. The pre-cue consisted in a 2° black line displayed 1.5° above the target/distractor centre. Stimuli were generated using Python programming language and the PsychoPy toolbox (Peirce, 2007) on a computer running Linux Ubuntu.

Procedure

Participants sat in a dark room during the experiment, 57 cm from the screen (CRT monitor, 1920 ×1080 pixels, 100 Hz refresh rate), with their head maintained using a chinrest. After a 200ms inter-stimulus interval (ITI), each trial started with the fixation dot being displayed on a grey background for a variable time period sampled from an exponential decay (scale: 500ms, bounded within the [300,1000] ms interval). This was done to maximize temporal uncertainty about stimuli onset. At the end of this delay, a cue was flashed during 60ms. After a variable cue-to-target onset asynchrony (5 different CTOA conditions: 100, 150, 250, 450 and 850ms), both target and distractor were displayed on either side of the fixation dot for 30ms. The target was oriented either clockwise or counter-clockwise relative to vertical, and the distractor was always horizontal. Participants were informed that the target was always the non-horizontal Gabor. Participants were requested to categorize the target as clockwise versus counter-clockwise (Type 1 decision) and press the corresponding key on the keyboard (left arrow for counter clockwise, right arrow for clockwise). In 50% of the trials, the target appeared at the same location as the cue ("valid" condition), and for the remaining trials at the opposite location ("invalid" condition). The cue was therefore fully unpredictive, and participants had no further incentives to orient their attention voluntarily towards the cued location. After their response, participants were prompted to report their confidence in the correctness of their response using the up/down arrow keys (Type 2 decision): is your confidence for this trial higher or lower than average? Participants started with 10 practice trials with feedback prior to the calibration (see below), which was then followed by the main experiment. Participants were provided with a 10 second break every 60 trials. The design was fully factorial with 5 CTOAs conditions X valid/invalid condition, with pseudo randomization per virtual blocks of 20 trials.

Participants were instructed to fixate the centre of the screen during the whole trial period, given that target location was unpredictable. The purpose of the task was not to probe covert attention specifically, but rather to estimate the effect of exogenous cueing in a more ecological setting. As such, no eye-tracking monitoring was used in the present study. Participants completed 3 sessions of 1 hour each, consisting in 560 trials per session (1680 trials in total).

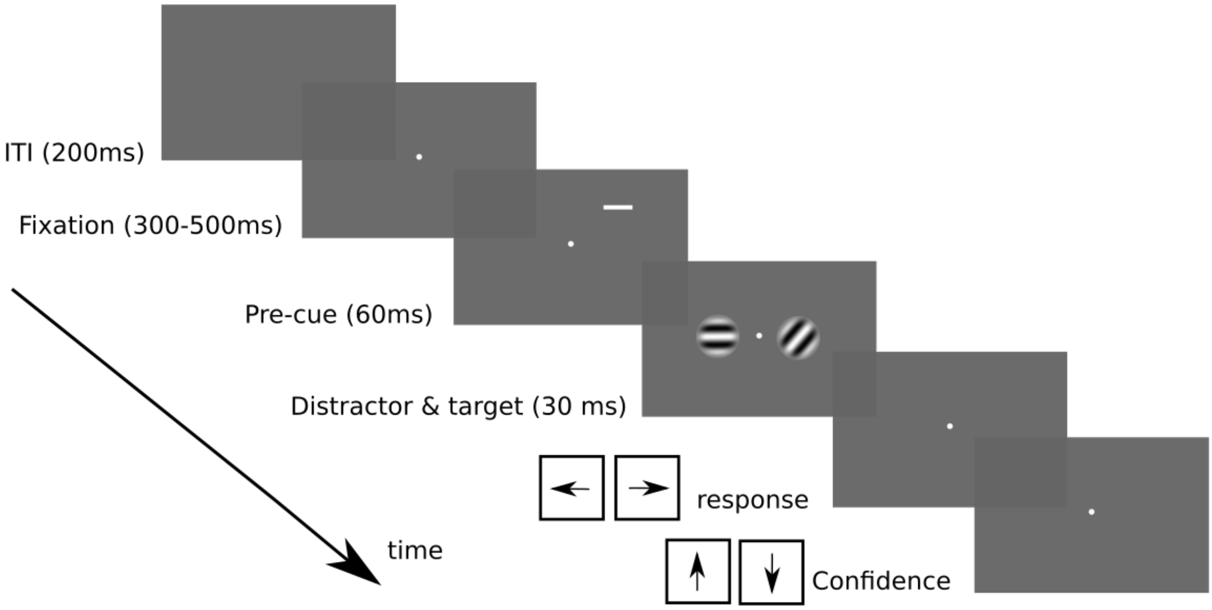

Figure 1. Experimental protocol: On each trial, after a random delay, a cue is briefly presented on one side of the fixation cross. After a variable cue-to-target onset asynchrony (CTOA), a target and a distractor are presented. The target can appear slightly below the cue (valid condition, as illustrated here) or on the opposite side (invalid condition). The target is an oriented Gabor patch (either clockwise or counter-clockwise) and the distractor is a horizontal Gabor patch. After target and distractor offset, the participant has to report target orientation and then to rate her confidence in the response on a two-point scale (more or less confident than average).

Calibration

The psychometric function was estimated prior to the beginning of the experiment for each participant in order to aim for a 75% average perceptual accuracy. This function plots the proportion of "counter clockwise" responses against the difference (in degrees) between the two possible visual orientations of the target. From the participant perspective, the task during this calibration part looked similar to the one in the main experiment, but the orientation of the target was varied from trial to trial using an Accelerated Stochastic Approximation (ASA) staircase procedure (Kesten, 1958). In the calibration part, the cue was systematically displayed on both the target and distractor side, to provide participants with only temporal - but no spatial - information about target onset. Cue-to-target interval during calibration was fixed at 100ms. Confidence estimates during calibration were not evaluated. At the end of the calibration, the psychometric curve was fit to the data using Maximum Likelihood Estimation

(MLE), to extract angle values (for clockwise and counter clockwise targets) leading to 75% accuracy. These values were then kept constant for the whole session, in particular to avoid trial-by-trial fluctuation of objective and subjective difficulty level independently of condition manipulation. Keeping stable difficulty levels has been shown to reduce the risk of inflating metacognitive ability estimates (Rahnev & Fleming, 2019).

MEASURES

We are interested in estimating both perceptual (Type 1) and meta-perceptual (Type 2) sensitivities. We thus used Signal Detection Theory (SDT) was used to estimate Type 1 sensitivity (d') which provided us with a bias-free measure of accuracy. Trials were grouped using the clockwise-oriented category as signal, leading to four categories of trials: (a) hits, where a CW target was correctly reported as CW; (b) misses, in which a CW target was reported as CCW; (c) false alarms, where a CCW target was reported as CW; (d) correct rejections, where CCW was reported as CCW. This grouping was conducted for each participant and each condition separately, and sensitivity (d') was calculated as the difference in z-scores between the hit rate and the false alarm rate.

As a reliable proxy for Type 2 sensitivity (that is, how well confidence ratings relate to objective accuracy), we used Meta-d’, as it is less prone than other measures to shifts in Type 1 sensitivity or response bias. It corresponds to the Type 1 sensitivity that would produce the collected Type 2 (or confidence) responses, if the observers were optimal at the metacognitive level (Maniscalco & Lau, 2012). This value, the meta-d’, can then be compared to the actual sensitivity (d') objectively measured for each participant. In particular, the meta-d' is equal to the d' when the participant has optimal metacognitive access to Type 1 decision information. The ratio meta-d'/d', or “m-ratio” is referred to as metacognitive efficiency. To investigate the effect of cueing on metacognitive efficiency, we thus considered the m-ratio, after estimating d’ and meta-d’ using Maximum Likelihood methods. This procedure was applied for each participant, CTOA and pre-cue validity separately.

For clarity, and because we were interested in within – not between – participant variability, the errors bars in the following figures are based on the 95% CI of the within-participant variability. These CI were calculated using the Cousineau-Morey intervals (Baguley, 2012; Cousineau, 2005; Morey, 2008).

When necessary, ANOVAs were corrected using the Greenhouse-Geisser adjustment and t-tests were corrected using the Welch-Satterthwaite adjustment. We report Student t-test with a lowercase t when Shapiro-Wilk normality test did not fail, and Wilcoxon signed ranked test using uppercase T otherwise.

RESULTS

EXOGENOUS PRE-CUES AFFECT PERFORMANCE AND CONFIDENCE AT SHORT CTOA

We first evaluated how performance and confidence were affected by exogenous pre-cues, with separate ANOVAs for sensitivity, response times (RTs) and average confidence as successive dependent variables, and with pre-cue validity and cue-to-target onset asynchrony (CTOA) as independent variables.

Perceptual sensitivity was affected by the interaction between CTOA and validity ($F(3.2,28.8)=4.25$, $MSE=0.06$, $p=0.012$), with no main effect of CTOA ($F(2.9,26.3)=0.3$, $MSE=0.1$, $p=0.334$) or validity ($F(1,9)=3.7$, $MSE=0.08$, $p=0.087$). Paired t-tests between the valid and invalid conditions at each CTOA confirmed a significant gain in sensitivity for the valid condition at 100 ms CTOA ($T(9) = 52$, $p=0.0098$) and 150ms ($T(9)=50$, $p=0.020$), but not for other CTOAs ($p > 0.3$). These results thus confirmed that our cueing procedure successfully affected perceptual performance at short lags (fig. 2A), consistent with the automatic capture of attention.

To ensure that the effect on sensitivity was not simply the result of a speed-accuracy trade-off, we looked at response times (fig. 2B). We found that RTs exhibited the same pattern as sensitivity did. The repeated-measures ANOVA showed an effect of CTOA ($F(2.1, 19.3) = 6.3$, $MSE=0.004$, $p=0.007$), no effect of validity ($F(1,9) = 4.01$, $MSE = 0.003$, $p=0.076$), but an interaction ($F(3.14, 28.3) = 4.8$, $MSE = 0.001$, $p=0.007$). Paired t-tests between the valid and invalid

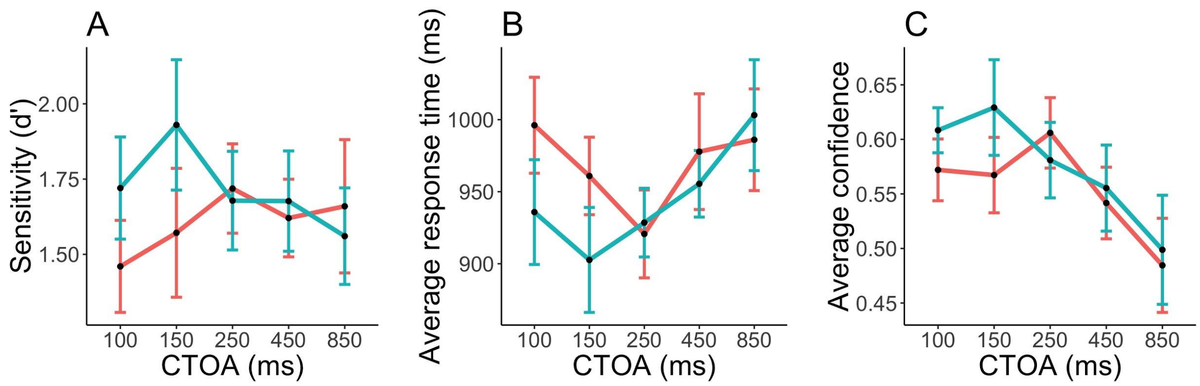

Figure 2. Cueing dynamics: (A) Average sensitivity (d') as a function of cue-to-target onset asynchrony (CTOA) for valid (blue) and invalid (red) cues. Sensitivity is greater at valid location for short CTOAs. (B) Average response time as a function of CTOA and cue validity, with lower response times for valid location, short CTOAs. (C) Average confidence as a function of CTOA and cue validity. The 100ms CTOA showed greater confidence for valid cues. Error bars represent within-subjects, 95% CI.

conditions at each CTOA showed significantly lower response times in the valid condition for the 100ms ($t(9)=-2.56$, $p=0.031$) and 150ms ($t(9)=-2.77$, $p=0.022$) CTOAs, but not for other CTOAs (all $p>0.07$). These results demonstrate that the gain in sensitivity was accompanied by a decrease in response times, and could thus not be the result of a speed-accuracy trade-off.

Confidence was affected in a like way as perception (Fig. 2C). The ANOVA showed a main effect of CTOA ($F(2.1,18.6)=10.11$, $MSE=0.008$, $p=0.001$), no effect of validity ($F(1,9)=3.9$, $MSE=0.003$, $p=0.079$), but an interaction between CTOA and validity ($F(2,18.1)=4.07$, $MSE=0.002$, $p=0.034$). Paired t-tests between the valid and invalid conditions at each CTOA confirmed a significantly higher confidence for the valid condition at 100 ms CTOA ($T(9) = 48$, $p=0.037$), but not for other CTOAs ($p > 0.08$). In other words, confidence and performance increased in similar ways: for valid trials at short CTOAs.

It is interesting to note that there was a main effect of CTOA on confidence despite sensitivity being stable overall across CTOAs. It is also unlikely to reflect temporal expectations, given the flat hazard rate used in our design. It might however reflect response times, which show a slight increase at longer CTOAs.

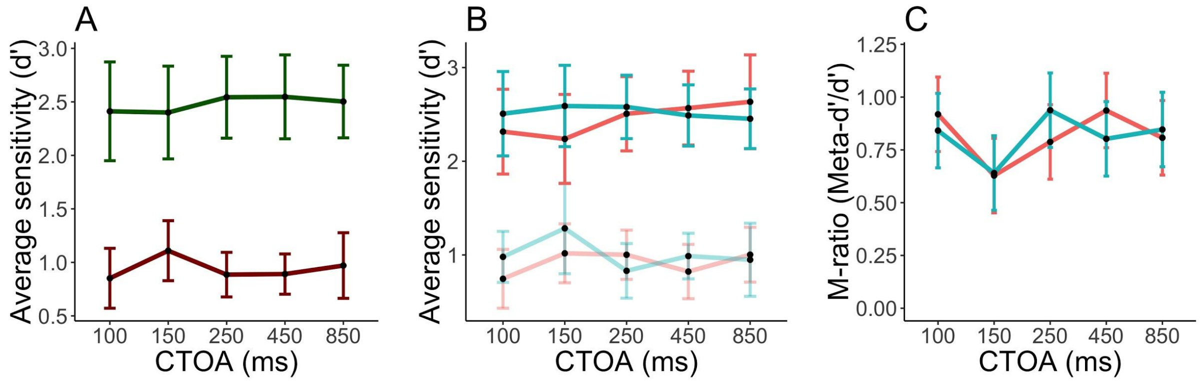

Figure 3. Cueing effects and metacognitive efficiency: (A) The average sensitivity for High (green) and Low (red) confidence trials, per CTOAs, a first measure of metacognition. Metacognition was stable across time. (B) The average sensitivity for High (high alpha) and Low (low alpha) confidence trials as a function of CTOA, this time grouped for valid (blue) and invalid (red) condition. (C) The metacognitive efficiency (or "m-ratio"), which is the ratio of meta-d'/d', as a function of CTOAs and validity. Cueing does not significantly affect metacognitive sensitivity or efficiency, a result coherent with the observed relationship between confidence and sensitivity. Error bars present the within-subject 95 % CI.

To confirm the similarity between the cueing effects on sensitivity and confidence, we calculated the cueing effect (valid minus invalid) for each CTOA, for confidence and sensitivity separately, and evaluated Pearson's correlation across the 5 CTOAs for each participant. In line with our expectations, these correlations were globally positive ($T(9) = 47$, $p=0.048$), although this result was statistically modest.

METACOGNITION IS STABLE ACROSS CONDITIONS

A first analysis was conducted to check the presence of overall metacognitive insight by comparing high and low confidence trials. Specifically, we conducted a repeated-measures ANOVA with sensitivity as the dependant variable and CTOA, validity and confidence as independent variables and we found only a significant effect of confidence on sensitivity ($F(1,9)=85.33$, $MSE=1.36$, $p<0.001$), with no other main effects or interactions (all $p>0.09$). Therefore, we found that when participants expressed higher confidence, their sensitivity was indeed higher, which indicates some metacognitive ability (Fig. 3A and 3B).

To further quantify metacognitive efficiency (fig. 3C), we estimated the ratio of meta-d' over d' for each CTOA and cue validity condition, for each participant. We find no effect of CTOA or validity on metacognitive sensitivity or efficiency (Fig. 4). An ANOVA with the m-ratio (meta-d'/d') as dependent variable and CTOA and validity as independent variables showed no significant effect of CTOA ($F(2,18.1) = 0.99$, $MSE = 0.27$, $p = 0.4$), or validity ($F(1,9) = 0.9$, $MSE = 0.09$, $p = 0.4$), and no interaction ($F(3.3, 26.5) = 0.6$, $MSE = 0.1$, $p=0.6$). This result is consistent with the interpretation that validity or CTOA affected both meta-d' and d' in a similar way, leading to a stable metacognitive efficiency despite a fluctuation of metacognitive sensitivity.

Discussion

We found that exogenous cues only affect sensitivity and response times at short cue-to-target intervals (fig. 2, A and B), replicating numerous previous studies (for a review, see Carrasco, 2011). Confidence judgments reflected this initial boost in sensitivity, leading to greater confidence for the valid than for the invalid pre-cues (fig. 2C). This effect was short-lived, and disappeared together with the difference in sensitivity for longer cue-to-target intervals. This suggests that even with spatial uninformative transients, the computations underlying confidence still have access to the early gain in accuracy induced by valid pre-cueing. Furthermore, temporal proximity between cue and target boosted confidence, independently of the cue validity and overall sensitivity and response times trend (fig. 2C).

A recent paper that investigated whether exogenous cueing could influence confidence found no effect (Kurtz et al., 2017). However, the cues used in this paper were predictive and only one CTOA was used, thus making it difficult to rule out non-exogenous effects. In the present study, using non-predictive cues and several CTOAs, we could control for these aspects. We found that an exogenous cue increased both sensitivity and confidence at short CTOAs. These effects were observed despite participants being clearly informed that cue location was randomly drawn, and that there was no reason to expect that the target would appear at the same location. It appears that participants understood these instructions, since for longer cue-to-target intervals sensitivity was similar between the valid and invalid locations (suggesting that participants did not reallocate their attention voluntarily where the cue appeared). However, they nevertheless adjusted their confidence level to the initial increase in sensitivity. In addition, we also used a standard measure to assess metacognition, and we did not find any effect of cueing on metacognitive sensitivity. Despite the size of our cueing effect on confidence being small, a true absence of any confidence-accuracy correlation at the shortest CTOA should have resulted in metacognitive efficiency fluctuating across CTOAs, something we did not observe in our data.

One important aspect of the present paradigm is the complete non-predictability of the spatial cue. To our knowledge, no studies to date have considered the role of exogenous non-predictive cues in the building of confidence estimates. Most of the previous works used semi or fully predictive cues, leading to a possible effect of both spatial expectation and attention, and the possibility that attentional effects were in fact driven by expectations. This might be a problem given the claims for a functional difference between expectation and attention (Summerfield & Egner, 2009). Separating attention from expectation effects on metacognition is particularly critical, given recent findings that expectations can enhance metacognitive abilities (Sherman, Seth, Barrett, & Kanai, 2015; Sherman, Seth, & Kanai, 2016). Spatial expectations and exogenous attention are typically considered as two independent processes, with the effects of exogenous attention on sensitivity (Giordano, McElree, & Carrasco, 2009) or response times (Meijs, Klaassen, Bokeria, Van Gaal, & De Lange, 2018) immune to changes in spatial expectations. However, whether this independence between attention and expectation effects also holds for confidence judgments remains an open empirical question for future research.

Regarding the temporal profile of confidence, we found an unexpected over-confidence for short cue-target intervals, irrespective of cue validity (fig. 2C). This bias might be due to the distribution of CTOAs, as our paradigm included more cue-targets intervals below 300ms. In our design, the choice of favouring shorter over longer cue-target intervals was meant to flatten the hazard rate, in order to ensure that vigilance remains stable across CTOAs. The fact that overall sensitivity was relatively stable over time suggests that this manipulation was successful. Nonetheless, it is still possible that with this distribution of CTOAs, participants felt more familiar with shorter CTOAs and were influenced by this familiarity during their confidence judgments.

While exogenous cues are often considered as a trigger for selective attention, it has also been proposed that the sensitivity boost following sharp contrast transients might result from low-level sensory effects (Solomon & Morgan, 2018; Solomon, 2004). In these studies, exogenous pre-cueing has been shown to uniformly boost sensitivity at cued locations, even when more than one location was cued. One approach to tackle the selectivity of the cueing process and its effect on confidence might be to use a certain type of neutral cues, where both locations are cued simultaneously. The observed early boost in sensitivity and confidence might actually come from a facilitation at the valid location, a suppression at the invalid location, or a mixture of both, as suggested in the literature on exogenous selective attention (Carrasco, 2011). Whether confidence is equally sensitive to suppression and facilitation induced by cueing is a question for further work to address.

Conclusion

Confidence judgments were able to adjust to the initial boost in sensitivity induced by transient cues. Importantly, participants knew that the cues were fully unpredictive, and had no reason to favour valid over invalid locations. The early increase in both sensitivity and confidence was equally short-lived, and disappeared for longer cue-target intervals. Metacognitive ability, however, remained stable across different cue-target periods. These results suggest that visual confidence is able to track the perceptual effects of unpredictive exogenous cues.

Data Availability

The data for the experiment is freely available as part of the Confidence database, via Open Science Framework: https://osf.io/s46pr/.

Acknowledgements

This research was supported by the Agence Nationale de la Recherche (ANR-16-CE28-0002 and ANR-16-ASTR-0014 to VdG, ANR-17-EURE-0017 and ANR-18-CE28-0015 to PM) and by PSL University (doctoral scholarship to SR).

Contributions

SR, VdG and PM designed the experiment. SR conducted the experiment. SR, VdG and PM analyzed the data and wrote the manuscript.

Competing interests

The authors declare no competing interests.

References

Baguley, T. (2012). Calculating and graphing within-subject confidence intervals for ANOVA. Behavior Research Methods. https://doi.org/10.3758/s13428-011-0123-7

Carrasco, M. (2011). Visual attention: The past 25 years. Vision Research. https://doi.org/10.1016/j.visres.2011.04.012

Carrasco, M., Ling, S., & Read, S. (2004). Attention alters appearance. Nature Neuroscience, 7(3), 308–313. https://doi.org/10.1038/nn1194

Cousineau, D. (2005). Confidence intervals in within-subject designs: A simpler solution to Loftus and Masson's method. Tutorials in Quantitative Methods for Psychology. https://doi.org/10.20982/tqmp.01.1.p042

de Gardelle, V., & Mamassian, P. (2014). Does Confidence Use a Common Currency Across Two Visual Tasks? Psychological Science, 25(6), 1286–1288. https://doi.org/10.1177/0956797614528956

Denison, R. N., Adler, W. T., Carrasco, M., & Ma, W. J. (2018). Humans incorporate attention-dependent uncertainty into perceptual decisions and confidence. Proceedings of the National Academy of Sciences, 201717720. https://doi.org/10.1073/pnas.1717720115

Desender, K., Boldt, A., & Yeung, N. (2018). Subjective Confidence Predicts Information Seeking in Decision Making. Psychological Science. https://doi.org/10.1177/0956797617744771

Failing, M. F., & Theeuwes, J. (2014). Exogenous visual orienting by reward. Journal of Vision. https://doi.org/10.1167/14.5.6

Faivre, N., Filevich, E., Solovey, G., Kühn, S., & Blanke, O. (2018). Behavioral, modeling, and electrophysiological evidence for supramodality in human metacognition. Journal of Neuroscience. https://doi.org/10.1523/JNEUROSCI.0322-17.2017

Giordano, A. M., McElree, B., & Carrasco, M. (2009). On the automaticity and flexibility of covert attention: A speed-accuracy trade-off analysis. Journal of Vision. https://doi.org/10.1167/9.3.30

Guggenmos, M., Wilbertz, G., Hebart, M. N., & Sterzer, P. (2016). Mesolimbic confidence signals guide perceptual learning in the absence of external feedback. ELife. https://doi.org/10.7554/eLife.13388

Hainguerlot, M., Vergnaud, J. C., & De Gardelle, V. (2018). Metacognitive ability predicts learning cue-stimulus associations in the absence of external feedback. Scientific Reports. https://doi.org/10.1038/s41598-018-23936-9

Kesten, H. (1958). Accelerated Stochastic Approximation. The Annals of Mathematical Statistics, 29(1), 41–59. https://doi.org/10.1214/aoms/1177706705

Kurtz, P., Shapcott, K. A., Kaiser, J., Schmiedt, J. T., & Schmid, M. C. (2017). The Influence of Endogenous and Exogenous Spatial Attention on Decision Confidence. Scientific Reports, 7(1). https://doi.org/10.1038/s41598-017-06715-w

Mamassian, P. (2016). Visual Confidence. Annual Review of Vision Science, 2(1), annurev-vision-111815-114630. https://doi.org/10.1146/annurev-vision-111815-114630

Maniscalco, B., & Lau, H. (2012). A signal detection theoretic approach for estimating metacognitive sensitivity from confidence ratings. Consciousness and Cognition, 21(1), 422–430. https://doi.org/10.1016/j.concog.2011.09.021

Meijs, E. L., Klaassen, F. H., Bokeria, L., Van Gaal, S., & De Lange, F. P. (2018). Cue predictability does not modulate bottom-up attentional capture. Royal Society Open Science. https://doi.org/10.1098/rsos.180524

Morey, R. D. (2008). Confidence Intervals from Normalized Data: A correction to Cousineau (2005). Tutorials in Quantitative Methods for Psychology. https://doi.org/10.20982/tqmp.04.2.p061

Munneke, J., Hoppenbrouwers, S. S., & Theeuwes, J. (2015). Reward can modulate attentional capture, independent of top-down set. Attention, Perception, and Psychophysics. https://doi.org/10.3758/s13414-015-0958-6

Peirce, J. W. (2007). PsychoPy-Psychophysics software in Python. Journal of Neuroscience Methods. https://doi.org/10.1016/j.jneumeth.2006.11.017

Pestilli, F., & Carrasco, M. (2005). Attention enhances contrast sensitivity at cued and impairs it at uncued locations. Vision Research, 45(14), 1867–1875. https://doi.org/10.1016/j.visres.2005.01.019

Posner, M. I. (1980). Orienting of attention. Quarterly Journal of Experimental Psychology. https://doi.org/10.1080/00335558008248231

Rahnev, D., Maniscalco, B., Graves, T., Huang, E., de Lange, F. P., & Lau, H. (2011). Attention induces conservative subjective biases in visual perception. Nature Neuroscience, 14(12), 1513–1515. https://doi.org/10.1038/nn.2948

Rahnev, Dobromir, & Fleming, S. M. (2019). How experimental procedures influence estimates of metacognitive ability. Neuroscience of Consciousness. https://doi.org/10.1093/nc/niz009

Recht, S., Mamassian, P., & de Gardelle, V. (2019). Temporal attention causes systematic biases in visual confidence. Scientific Reports, 9(1), 11622. https://doi.org/10.1038/s41598-019-48063-x

Reyes, G., & Sackur, J. (2014). Introspection during visual search. Consciousness and Cognition. https://doi.org/10.1016/j.concog.2014.08.009

Reyes, G., & Sackur, J. (2017). Introspective access to implicit shifts of attention. Consciousness and Cognition, 48, 11–20. https://doi.org/10.1016/j.concog.2016.10.003

Rosenthal, D. (2018). Consciousness and confidence. Neuropsychologia. https://doi.org/10.1016/j.neuropsychologia.2018.01.018

Schneider, K. A., & Komlos, M. (2008). Attention biases decisions but does not alter appearance. Journal of Vision. https://doi.org/10.1167/8.15.3

Sergent, C., Wyart, V., Babo-Rebelo, M., Cohen, L., Naccache, L., & Tallon-Baudry, C. (2013). Cueing attention after the stimulus is gone can retrospectively trigger conscious perception. Current Biology, 23(2), 150-155.

Sherman, M. T., Seth, A. K., Barrett, A. B., & Kanai, R. (2015). Prior expectations facilitate metacognition for perceptual decision. Consciousness and Cognition. https://doi.org/10.1016/j.concog.2015.04.015

Sherman, Maxine T., Seth, A. K., & Kanai, R. (2016). Predictions shape confidence in right inferior frontal gyrus. Journal of Neuroscience. https://doi.org/10.1523/JNEUROSCI.1092-16.2016

Solomon, J. A., & Morgan, M. J. (2018). Precues' elevation of sensitivity is not only preattentive, but largely monocular. Attention, Perception, and Psychophysics. https://doi.org/10.3758/s13414-018-1564-1

Solomon, Joshua A. (2004). The effect of spatial cues on visual sensitivity. In Vision Research. https://doi.org/10.1016/j.visres.2003.12.003

Summerfield, C., & Egner, T. (2009). Expectation (and attention) in visual cognition. Trends in Cognitive Sciences, 13(9), 403–409. https://doi.org/10.1016/j.tics.2009.06.003

van den Berg, R., Zylberberg, A., Kiani, R., Shadlen, M. N., & Wolpert, D. M. (2016). Confidence Is the Bridge between Multi-stage Decisions. Current Biology. https://doi.org/10.1016/j.cub.2016.10.021

Wilimzig, C., Tsuchiya, N., Fahle, M., Einhäuser, W., & Koch, C. (2008). Spatial attention increases performance but not subjective confidence in a discrimination task. Journal of Vision, 8(5), 7. https://doi.org/10.1167/8.5.7

Zizlsperger, L., Sauvigny, T., & Haarmeier, T. (2012). Selective attention increases choice certainty in human decision making. PLoS ONE. https://doi.org/10.1371/journal.pone.0041136

Zizlsperger, L., Sauvigny, T., Händel, B., & Haarmeier, T. (2014). Cortical representations of confidence in a visual perceptual decision. Nature Communications. https://doi.org/10.1038/ncomms4940

Zylberberg, A., Wolpert, D. M., & Shadlen, M. N. (2018). Counterfactual Reasoning Underlies the Learning of Priors in Decision Making. Neuron. https://doi.org/10.1016/j.neuron.2018.07.035

CHAPTER 2 | METACOGNITION OF ATTENTIONAL DISENGAGEMENT

Metacognition of attentional disengagement

Samuel Recht [1], Vincent de Gardelle [2], Pascal Mamassian [1], Niko A. Busch [3], Laura Dugué[4]

[1] Laboratoire des systèmes perceptifs, Département d'études cognitives, École normale supérieure, PSL University, CNRS, 75005 Paris, France

[2] CNRS and Paris School of Economics, Paris, France

[3] Institute of Psychology, Westfälische Wilhelms-Universität and Otto Creutzfeldt Center for Cognitive and Behavioral Neuroscience, Westfälische Wilhelms-Universität, Münster, Germany

[4] CNRS (Integrative Neuroscience and Cognition Center, UMR 8002) and Université Paris Descartes, Sorbonne Paris Cité, Paris, France

In the previous chapter, we observed that confidence was able to track the early effects of exogenous cueing on accuracy. This result suggests that despite the use of unpredictive, task-irrelevant pre-cues, the gain in accuracy induced by these irrelevant transients was still detectable in confidence judgments. However, exogenous orienting is not the only situation in which the effects of spatial attention can be considered task-irrelevant. Another such situation would be when attending voluntarily and solely to a specific location is not relevant anymore for the task at hand. Spatial attention would thus need to be disengaged from the initial location. Despite disengagement, when a local event occurs soon after the endogenous attentional episode has ceased, the former might still benefit from the later. Here, we define 'attentional disengagement' as the process of progressively deallocating covert endogenous attention from one location before eventually reorienting it to another location. In this sense, attentional disengagement is a transition phase between two stable attentional orienting states. We will see in this chapter that attentional disengagement can take longer than reorienting, which can occur in a very short time span.

The following experiment was initially designed to test two aspects of selective attention and confidence: (a) The overall effect of attentional disengagement on confidence judgments and (b) the more fined-grained, rhythmic temporal structure of selective attention and its effects on confidence. We will only be presenting the first aspect of this work in the present chapter, the second aspect, being intricately fine-grained, requires large amounts of data collection, and is thus still in preparation.

In this chapter, we therefore investigate how confidence tracks the effects of attentional reorienting and disengagement following endogenous attentional orienting. The experimental protocol was specifically designed to probe perceptual performance at both attended and unattended locations on each trial, as well as participant's confidence for these two estimates. Importantly, we use a bias-free measure of confidence to extrapolate metacognitive sensitivity from bias.

Material & methods

Participants

5 adult volunteers participated in the study (M ± SD = 26.8 ± 2.3, 4 females). They all provided informed written consent prior to the experiment. Participants were compensated for their time at a rate of 10€ per hour. The experiment consisted in roughly five 3-hour sessions; the total number of sessions varied per participant (total number of trials per participant: 5500). All procedures were approved by the CERES (Conseil d'Évaluation Éthique pour les Recherches En Santé) ethics committee of Paris Descartes University. All research was performed in accordance with the relevant guidelines and regulations from the committee.

Apparatus

Observers sat in a dark room, 57.5 cm from a calibrated and linearized CRT monitor (refresh rate: 85 Hz; resolution: 1280 × 1024 pixels), their head maintained with a chin-rest. Visual stimuli were generated and presented using MATLAB (MathWorks, Natick, MA) and the MGL toolbox (http://gru.stanford.edu/doku.php/mgl/overview). Eye fixation was monitored online using an eye-tracker (Eyelink 1000, SR Research).

Stimuli & procedure

The present experimental protocol was adapted from a recent study (Senoussi, Moreland, Busch, & Dugué, 2019), to incorporate a continuous reproduction task with confidence judgments. On each trial, participants were first presented with a black fixation cross on a grey background (fig. 1). After a variable delay (sampled from a uniform distribution between 1000ms and 2000ms), the fixation cross turned white and a central pre-cue was displayed during 50ms. This pre-cue was pointing toward the left or right bottom quadrant. Following a fixed 350ms stimulus onset asynchrony (SOA) meant to maximize covert endogenous orienting of attention at the pre-cued location (Carrasco, 2011), two tilted sinusoidal gratings (4° diameter, 12% contrast, 2 cpd), windowed by a raised-cosine function, were presented within each bottom quadrant at 4° eccentricity. The orientation of the tilt for each grating (clockwise versus counter-clockwise) was assigned independently and randomly on each trial. The angle of the tilt was determined for each participant prior to the experiment using an adaptive staircase procedure to achieve 75% average accuracy ("1 up/2 down"). The pre-cue was neutral during the staircase procedure, and consisted of two diagonal lines pointing toward each grating location (each measuring half the size of the pre-cue). Given the duration of the experiment and the number of sessions, to mitigate the effect of learning, the staircase procedure was re-initiated each time the participant was departing too systematically (< 65% or >85% correct responses for a block of 130 trials) from the target accuracy level. The two gratings were displayed for 60ms together with a response cue indicating one of them (i.e., the "target") for subsequent report by the participant. The pre-cue was predictive of the target location in 70% of trials. For these trials, the pre-cue was considered 'valid'. For the remaining 30% of the trials, the pre-cue was not predictive of target location, but rather foil location, and was considered 'invalid'.

After a variable inter-stimulus interval (ISI), sampled from 13 possible intervals (from 40ms to 520ms, by step of 40ms), two Landolt Cs (diameter: 1.3°, white colour, 30° of aperture size, hereafter 'probes' were displayed for 130ms. The orientation of each probe was random on each trial. After the probes' offset, the colour of the fixation cross changed to black, inviting the participant to report the orientation of the target grating using the left, for counter-clockwise, and right, for clockwise, arrow keys on the keyboard, with no time pressure. Once the key was pressed, two disks (diameter: 1.3°, white colour) were displayed at each probe location. After participants clicked on one of the disks, an aperture (30° aperture size) was displayed at the cursor location. Participants were instructed to reproduce the orientation of the Landolt Cs using the mouse cursor. The release of the mouse button after adjustment registered the participants' response for the considered probe. After having oriented both probes, they were requested to select which of the two probes they were the most confident about. This 2-alternative forced choice for confidence report has been proposed as a criterion-free measure of metacognition (Barthelme & Mamassian, 2010; Barthelmé & Mamassian, 2009; de Gardelle et al., 2016; de Gardelle & Mamassian, 2014). Once clicked, the selected probe turned green and the next trial was initiated. The order of report for each probe was enforced by a small yellow line presented below the probe. The order of report was random in each trial. Importantly, participants were requested to be as precise as possible for both probes: they were specifically informed that performance was estimated in light of both reports. They were also specifically requested to prioritize the first task, in order to ensure a strong initial endogenous attentional orienting.

Every 20 trials, participants got feedback on their performance on the grating discrimination task, and had the opportunity to take a short, 20s break. A longer break was offered to the participants approximately every 25min. Participants completed roughly 5500 trials in 5-10 sessions, each session having a 3h maximum duration. To enforce fixation and prevent eyeblinks and saccadic shifts preceding cue onset, any trial during which participants blinked or move their gaze away from a 2° window centred on the fixation cross were automatically aborted, and a new sample of the trials was added at the end of the block. It was to check for successful fixation, because our primary interest was the covert deployments of endogenous attention, in the absence of gaze or head movements.

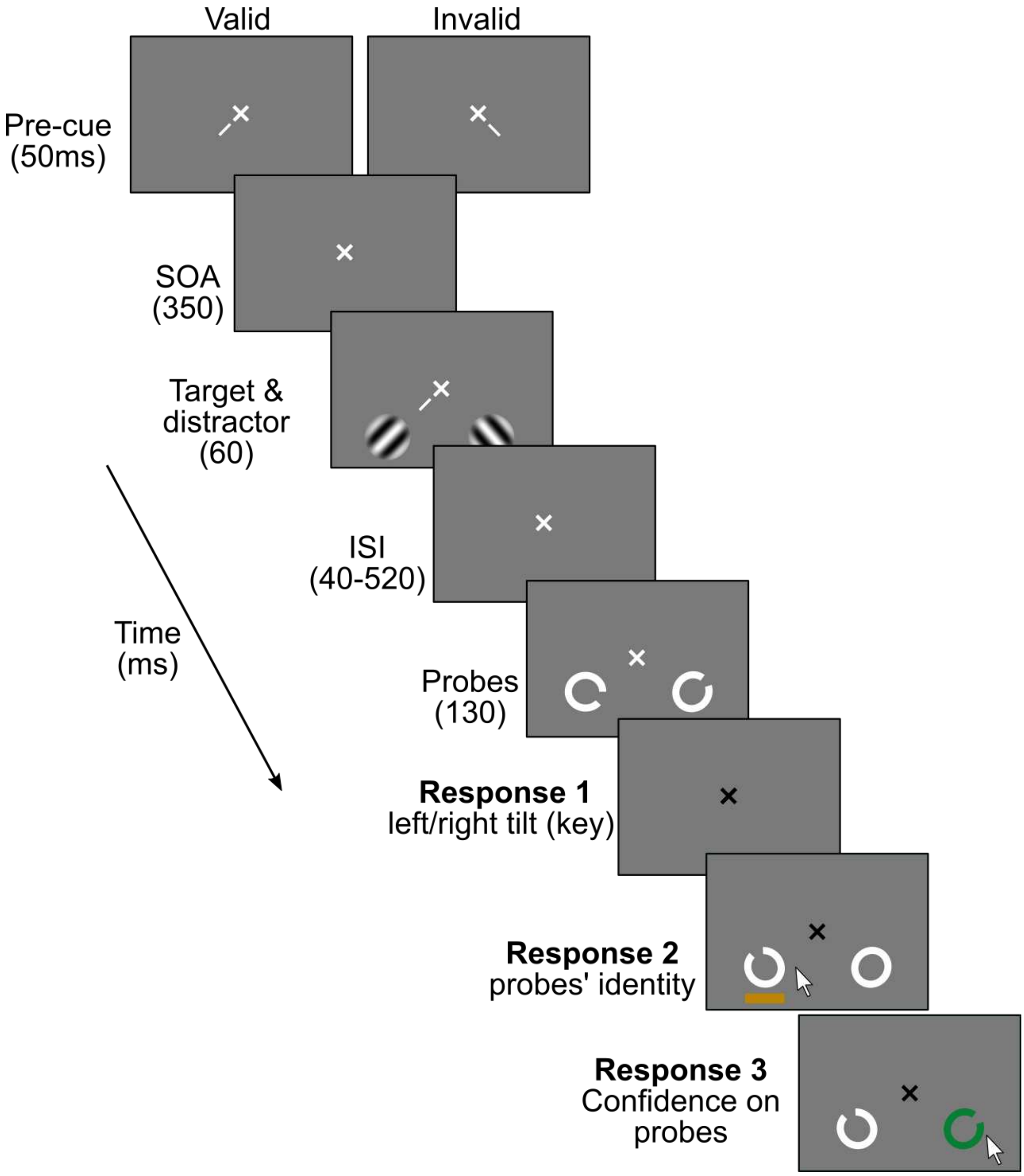

Figure 1. Experimental protocol. On each trial, a pre-cue was presented for 50ms either at the target (70% of the trials) or distractor (30%) location. After 350ms, two oriented grating patches were presented, together with the cue indicating the target to be reported. After a variable inter-stimulus interval (40-520ms), two Landolt Cs (or "probes") with random orientation were displayed at each grating location. Following probes offset, participants were requested to indicate the orientation of the target grating using the keyboard. Then, participants had to reproduce the orientation of each probe using the mouse cursor. The order of report was randomized across trials and indicated by a yellow line below the probe. Finally, participants had to select the probe for which they were the most confident by clicking on it.

ANALYSES

For clarity, and because we were mostly interested in within – not between – participants variability, the error bars for averaged data are based on the standard error (SEM) of the within-participant variability, unless otherwise stated. These SEM were calculated using the Cousineau-Morey intervals (Baguley, 2012; Cousineau, 2005; Morey, 2008).

For the first discrimination task, the oriented gratings accuracy was used to ensure an endogenous orienting of attention at cued location. Signal Detection Theory (SDT) was used to estimate sensitivity (d') as a bias-free measure of accuracy. Trials were grouped using the clockwise-oriented category as signal, leading to four categories of trials: (a) hits, where a CW target was correctly reported as CW; (b) misses, in which a CW target was reported as CCW; (c) false alarms, where a CCW target was reported as CW; (d) correct rejections, where CCW was reported as CCW. This grouping was conducted for valid and invalid trials and for each participant separately, and then sensitivity (d') was calculated as the difference in z-scores between hit rates, and false alarm rates.

For the second, reproduction task, two Landolt Cs were used to probe the quality of perceptual processing at the two locations. The error for a considered probe was calculated as the absolute distance between the true orientation of the probe and the reported orientation by the participant (in degrees). Here, we used the circular mean of the error. Our experimental protocol allowed us to measure the error at both the target and distractor locations, as two probes were presented on each trial. This approach allowed us to study trial-by-trial attentional disengagement, as a function of delay and pre-cue validity, and its effect on confidence. Trials in which the response time for either of the probes was 4 times higher than the standard deviation of any one given participant were discarded.

In all analyses, we used a repeated-measures ANOVA and t-tests. When necessary, ANOVAs were corrected using the Greenhouse-Geisser adjustment and t-tests were corrected using the Welch-Satterthwaite adjustment. We report

Student t-tests with a lowercase t when Shapiro-Wilk normality test did not fail, and Wilcoxon signed ranked test using uppercase T otherwise.

Results

Orienting of endogenous attention

We first performed a sanity check test on the covert orienting of endogenous attention in the grating discrimination task. Sensitivity was significantly greater in the valid compared to the invalid condition (t(4) = 3.66, p = 0.021), as expected. To rule out any speed-accuracy trade-off, we also checked response times. Response times were significantly faster in the valid compared to the invalid condition (t(4) = -4.41, p = 0.011). Therefore, these results show that attention was successfully manipulated with no speed-accuracy trade-off.

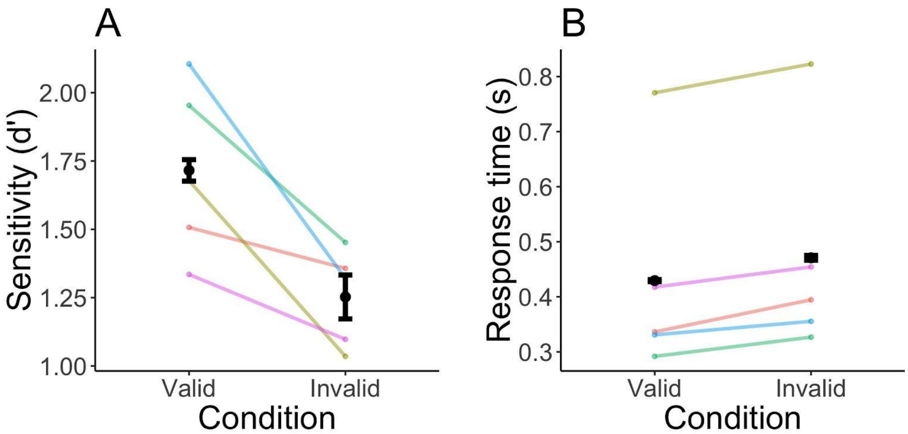

Figure 2. Endogenous orienting of attention. (A) Sensitivity (d') for valid and invalid trials. Valid cues elicited greater sensitivity. (B) The average response times for valid and invalid trials. Coloured lines represent individual participants. Black lines represent group averages. The error bars are within-participant ±1 SEM.

Measuring attentional disengagement

First, we tested the effect of validity and probe location on the average circular error, to determine whether validity affected average error on the reproduction task (fig. 3, A and B). A repeated-measure 3-way ANOVA with average error as the dependent variable, validity (valid/invalid), probe location (target/distractor side) and ISI as independent variables, revealed a significant main effect of validity

($F(1,4) = 13.58$, MSE = 43.90, $p = 0.021$), probe location ($F(1,4) = 10.31$, MSE = 1083.29, $p = 0.033$), and ISI ($F(12,48) = 5.95$, MSE = 72.41, $p < 0.001$). Only one interaction was significant, the ISI x probe location ($F(12,48) = 4.71$, MSE = 25.96, $p < 0.001$), indicating that the difference in error between each location was impacted by ISI, but neither the validity x probe ($F(1,4) = 4.71$, MSE = 25.96, $p=0.596$) nor the validity x ISI x probe ($F(12,48) = 1.64$, MSE = 6.01, $p = 0.111$) interaction were significant.

Taken together, these results demonstrate that the average error was affected by validity, probe location, and ISI, but that error difference between locations was likely mainly driven by the ISI.

CONFIDENCE

We tested the effect of condition (validity and ISI) on average confidence (fig. 3C and 3D). Since confidence judgment was about selecting one probe, we used a repeated-measure 2-way ANOVA with validity (valid/invalid) and ISI as independent variables, and the probability of selecting the probe at the target location as dependent variable (the probability of selecting the distractor probe being 1-p). The ANOVA revealed no main effect of validity ($F(1,4) = 3.74$, MSE = 0.006, $p = 0.125$), a main effect of ISI ($F(12,48) = 2.64$, MSE = 0.006, $p = 0.009$), and no validity x ISI interaction ($F(12,48) = 1.17$, MSE = 0.002, $p = 0.331$). Confidence thus did not seem to be affected by validity but monotonically decreased with the decay in error difference observed for longer ISIs.

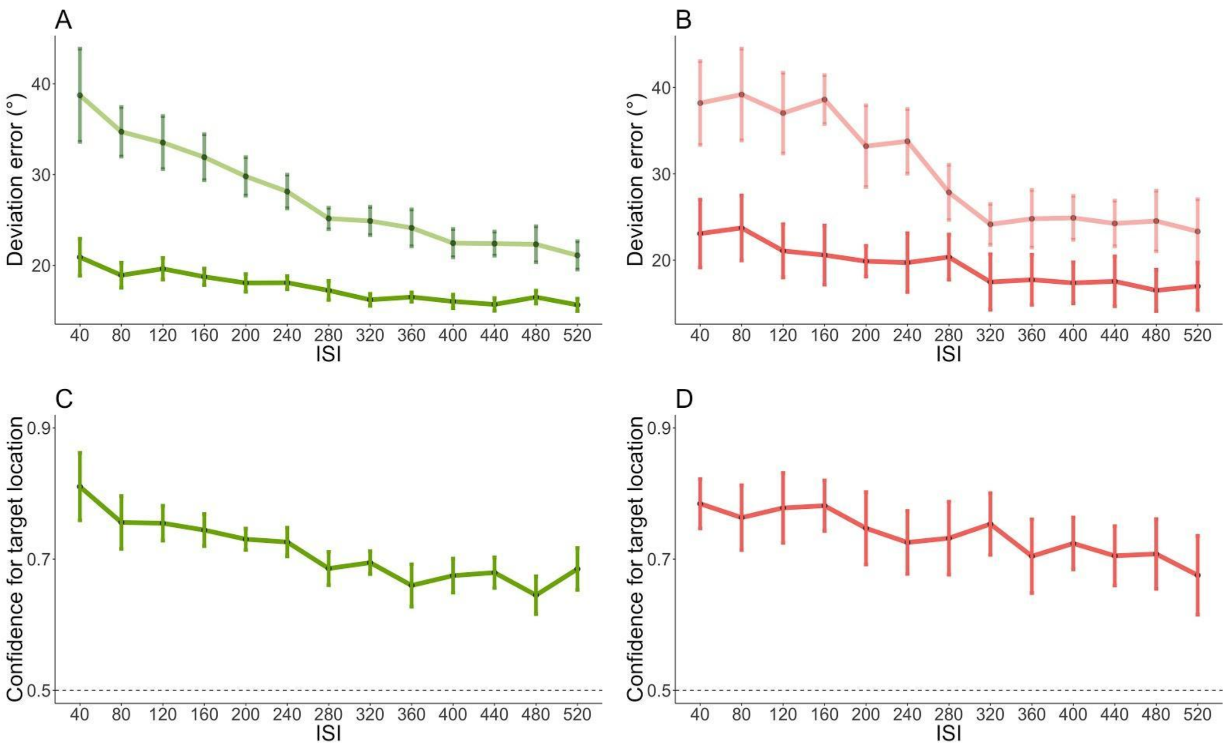

Figure 3. Disengagement of attention and confidence. (A) Average circular error (in degrees) for valid trials. Dark green represents the error at the target location and light green represents the error at the distractor location. (B) Average circular error (in degrees) for invalid trials. Dark red represents the error at the target location and light red represents the error at the distractor location. (C) Average probability of selecting with high confidence the probe on the target side. The probability at the distractor location is equivalent to 1-p (target location). (D) Average probability of selecting with high confidence the probe on the target side in the invalid condition. The error bars are within-participant ±1 SEM.

OVERALL METACOGNITION

The ability of confidence to accurately reflect performance has been coined metacognitive ability (Fleming & Lau, 2014; Fleming & Daw, 2017; Mamassian, 2016). A first rough estimate of metacognitive ability can be calculated by simply grouping probe reporting errors into two categories: high and low confidence. Figure 4A shows the average circular error for high and low confidence, for each participant, as a function of validity. Figure 4B shows the same variables, but as a function of ISI instead of validity. All participants exhibited overall metacognitive ability, with lower average error for the high confidence probes. A repeated-measures 3-way ANOVA revealed a significant main effect of confidence (F(1,4) = 28.67, MSE = 348.54, p = 0.006), validity (F(1,4) = 12.43, MSE = 27.77, p = 0.024), and ISI (F(12,48) = 6.35, MSE = 49.91, p<0.001). We found a confidence x ISI interaction (F(12,24) = 4.38, MSE = 21.26, p<0.001), indicating a progressive decrease in overall confidence difference, but no interaction between confidence and validity (F(1,4) = 3.58, MSE = 8.36, p = 0.131), validity x ISI (F(12,48) = 1.67, MSE = 5.40, p = 0.104) or confidence x validity x ISI (F(12,48) = 1.28, MSE = 3.68, p = 0.263). Therefore, validity affected the average error within each confidence category to a similar extent, and did not interact with time. However, the difference in error between high and low confidence decreased with time.

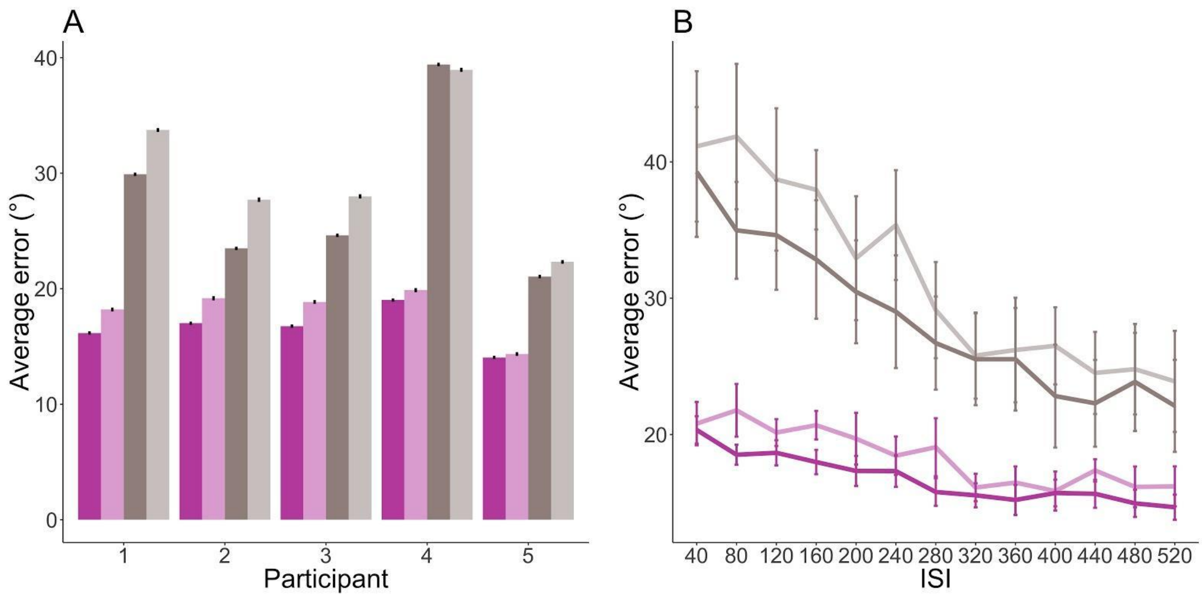

Figure 4. Overall metacognition. (A) Average error for the high (purple) and low (grey) confidence probe report for the valid (dark colour) and invalid (light colour) conditions. Each participant is depicted separately. (B) The average error for the high (purple) and low (grey) confidence probes for the valid (dark color) and invalid condition, as a function of ISI, at the group level. The error bars are within-participant ±1 SEM.

Trial-by-trial metacognition

We devised whether participants had access to the precision of their response on each trial, to some degree. If this were so, they should be using an estimate of the difference in error magnitudes between each probe to decide which one to select with greater confidence. Metacognitive ability can therefore be estimated using the objective difference in absolute error for each probe. We used a logistic regression to predict the probability of selecting the left-side probe during confidence judgment as a function of the difference in absolute errors between the left and right location probes (the 'subtraction model'). The selection of the left-side probe is arbitrary in this equation.

$$p(High\ confidence\ |Left\ probe) = a + \ \beta\Delta\varepsilon$$

$$\Delta\varepsilon = |\varepsilon_{Left}| - |\varepsilon_{Right}|$$

Where a and β are the intercept and the slope of the model, and ε_{Left} and ε_{Left} are the probe report error at left and right locations, respectively. A negative value of $\Delta\varepsilon$ would indicate a greater error for the left side, and a positive value, a greater error for the right side. It is possible that instead of relying strictly on the absolute difference in errors, confidence would also be sensitive to the overall error amplitude (that is, the sum of the errors). Even when the difference between two errors remains unchanged, it could be more difficult to decide between them when both errors are large. This form of scaling is observed for first order decisions (Shepard, 1987); it has likewise been proposed for confidence (C. Peirce & Jastrow, 1884; Van Den Berg, Shin, Chou, George, & Ma, 2012). We therefore tested an alternative 'scaling' model, where $\Delta\varepsilon$ was divided by the sum of the probe errors.

$$p(High\ confidence\ |Left\ probe) = a + \ \beta\frac{\Delta\varepsilon}{\Sigma\varepsilon}$$

$$\frac{\Delta\varepsilon}{\Sigma\varepsilon} = \frac{|\varepsilon_{Left}| - |\varepsilon_{Right}|}{|\varepsilon_{Left}| + |\varepsilon_{Right}|}$$

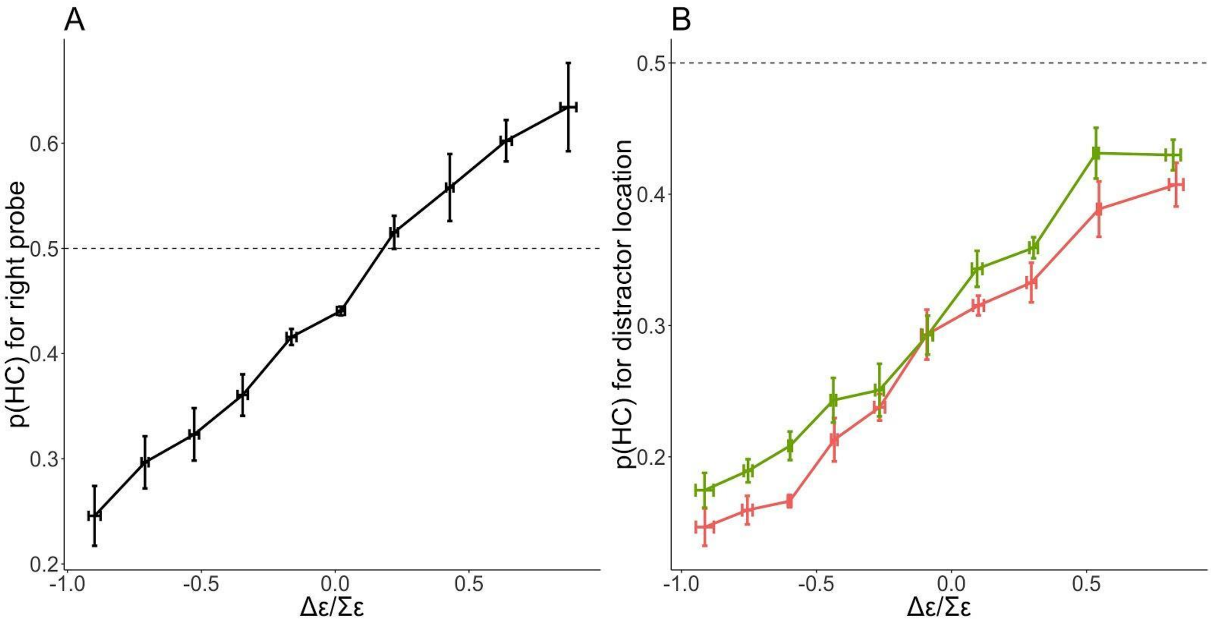

Figure 5. Trial-by-trial metacognitive ability. (A) Probability of selecting the probe on the right during confidence judgment, as a function of the scaled error difference between the two probes. A negative error represents greater error for the probe on the right. For illustration, the errors have been grouped in 10 quantiles. (B) Probability of selecting the probe at the distractor location (during confidence judgment), as a function of the weighted error difference between the probes at target and distractor locations. The valid condition is depicted in green, and invalid condition in red. A negative error represents a greater error for the probe at the distractor location. All values are below 0.5, indicating an overall metacognitive bias towards the target location. The errors have been grouped by quantiles. The error bars are within-participant ±1 SEM.

For each of these two measures, a logistic regression model (logit) was used to predict confidence judgments based on the considered error metric. This was done for each participant separately. The scaling model significantly outperformed the subtraction model for all participants ($\chi^2(0)$ = [343.81; 122.38; 112.4; 182.63; 75.826], with all $p<0.001$). We therefore selected the scaling model for all following analyses.

Furthermore, the positive slope at the group level ($t(4) = 6.66$, $p = 0.003$), confirms that participants were using the scaled difference in error magnitudes ($\Delta\varepsilon/\Sigma\varepsilon$) for their confidence judgments. Figure 5A shows the probability of a high confidence judgment for a given probe as a function of the scaled error difference between the two probes. Confidence increased monotonically while the relative error for the considered probe decreased.

Next, to test the effect of condition on metacognition, we used a mixed-effects logistic regression comparison approach. We constructed a model where confidence at the distractor location was predicted using the scaled error difference between distractor and target locations.

$$p(High\ confidence\ |Distractor\ location) = a + \beta \frac{\Delta\varepsilon}{\Sigma\varepsilon}$$

$$\frac{\Delta\varepsilon}{\Sigma\varepsilon} = \frac{|\varepsilon_{Distractor}| - |\varepsilon_{Target}|}{|\varepsilon_{Distractor}| + |\varepsilon_{Target}|}$$

In the simplest version of the model, confidence was predicted by the scaled error and there was an intercept as a random effect for each participant. A model with validity effect significantly outperformed the simplest model ($\beta = 0.14$; $\Delta AIC = -13$; $\Delta BIC = -4$; $\chi^2(1) = 14.31$, $p<0.001$). Including ISI improved the model further ($\beta = 0.15$; $\Delta AIC = -90$; $\Delta BIC = -82$; $\chi^2(1) = 91.98$, $p<0.001$). However, adding an interaction term between validity and scaled error did not enhance the model ($\beta = -0.04$; $\Delta AIC = 2$; $\Delta BIC = 10$; $\chi^2(1) = 0.30$, $p=0.585$). A model with validity x ISI interaction was not better either ($\beta = 0.03$; $\Delta AIC = 1$; $\Delta BIC = 9$; $\chi^2(1) = 0.82$, $p=0.365$). However, we found a significant negative interaction between ISI and scaled error ($\beta = -0.10$; $\Delta AIC = -9$; $\Delta BIC = -1$; $\chi^2(1) = 11.17$, $p<0.001$). Finally, the validity x ISI x scaled error interaction was not significant ($\beta = 0.03$; $\Delta AIC = 5$; $\Delta BIC = 28$; $\chi^2(3) = 1.46$, $p=0.692$).

These results suggest that confidence at the distractor location is greater on average for the valid condition, and has a tendency to increase with ISI. Importantly, there is still a strong confidence bias in favour of the target location up to 520 ms post-target. However, the relation between confidence and scaled error (that is, our proxy for metacognitive evidence) remained unaffected by validity. We nonetheless found a negative impact of ISI on metacognitive ability (via the ISI x scaled error interaction), suggesting that longer ISI were defined by lower metacognition.

Discussion

In this work, we investigated the effects of the disengagement of endogenous attention from a spatial location on both response precision (Type-1 decision) and confidence (Type-2 decision). For this purpose, we used two successive tasks within each trial. A first task, in which a pre-cue predicted the location of an upcoming target, was used to induce covert, endogenous orienting of attention to one of two possible locations. At different delays after the offset of the stimuli from the first task (i.e., discrimination of grating patches), we presented the participant with two probes: one at the attended, and the other at the unattended location. Participants had to report the grating target, but also the identity of both probes. All the reports occurred at the end of the trial. While orienting attention to the pre-cued location was meaningful for the first task, it was not relevant anymore for the second task since a probe appeared on both sides. In order to report the two probes as precisely as possible, disengaging spatial attention from first-task was therefore crucial. Importantly, in 30% of the trials, the pre-cue was invalid: in these trials, participants had to reorient their attention towards the opposite location to succeed in the first task. It should be noted that the 'attentional disengagement' in the current context was not intended to constitute a new spatial reorienting, because the probes are distributed over two distinct locations and visual fields, while being both equally relevant. In order to report both probes as accurately as possible, attention would have at least to spread to both locations, a transition which can also be encompassed by the term 'disengagement'. This manipulation allowed us to test the effect of recent reorienting (i.e., to the invalid location in the grating task) of voluntary attention on upcoming attentional disengagement and confidence.

Spatiotemporal signatures of attentional disengagement

Remarkably, the pattern of errors in the reproduction task did not change as a function of validity in the first task: in both valid and invalid trials, the lower error was always at the target location (fig. 3, A & B). This observation put some constraints on the potential mechanism involved. It is quite unlikely that attention was able to reorient to the target location in 100ms (target presentation duration

+ 40ms ISI), unless it had an exogenous component (Carrasco, 2011), which was not the case here. In our experiment, we presented both a target and a distractor in the first task, and the only way for the participant to distinguish the target from the distractor was to use the central cue presented concomitantly (fig. 2). Central cues, however, are known to elicit endogenous, but not exogenous orienting of attention (Carrasco, 2011). If attention did not have enough time to reorient voluntarily before target offset, how did the target location always elicit lower error, even when the probes followed the grating target by only 40ms?

At first glance, one may want to think of endogenous attention as a purely proactive mechanism, and yet this need not be so. Attention may act at the level of sensory/iconic memory, on the low-level stimulus footprint remaining in the sensory cortices. The existence of an attention benefit when using post or retro-cues (cues appearing after the stimulus is gone) has been demonstrated in numerous studies, suggesting a flexible temporal window around which a stimulus can be selected and prioritized even after offset (Dugué, Merriam, Heeger, & Carrasco, 2018; Griffin & Nobre, 2003; Ruff, Kristjánsson, & Driver, 2007; C. Sergent et al., 2013; C Sergent, Ruff, & Barbot, 2011).

Furthermore, the temporal structure of the task was such that it gave priority to the encoding of the first task's stimuli, followed by the second task stimuli, and the effect of prior entry might have been consolidated by the order of the reports at the end of each trial. The order of report of the two probes was randomized, the participant always had to report the grating first, and was instructed to prioritize this task. As such, the prioritization and enhancement of sensory signal induced by spatial attention would therefore occur first at target location, and could 'leak' toward a temporally and spatially adjacent stimulus. This is particularly interesting because it suggests a serial processing of different locations with an upper bound in the temporal precision of selective attention.

Metacognition successfully monitors attentional disengagement

In the present task, the probability of selecting the probe at the target location decreased monotonically with time (fig. 3, C and D), mirroring the shrinking in error difference between the probes at target and distractor locations. Despite a strong confidence bias favouring target location for all ISIs, confidence was sensitive to variation in estimation error across conditions. Indeed, when considering the relation between confidence and error, participants were able to monitor their precision, selecting more often the probe with the smaller error on average (fig. 4). This ability was sustained across ISIs, suggesting no strong impact of attentional disengagement on metacognition. The capacity to monitor the strong initial spatial bias induced by attention and its following decay was confirmed when analysing trial-level metacognitive ability. Confidence judgments were predictable from the difference in errors between the two probes on a trial-by-trial basis. In the spatial domain, confidence has been shown to adapt to change in performance following endogenous attention manipulation in some studies (Denison et al., 2018; Kurtz et al., 2017). However, other studies found systematic differences between accuracy and confidence when endogenous attention is involved (D. Rahnev et al., 2011; Wilimzig et al., 2008; Zizlsperger et al., 2012, 2014). Yet, most of these studies did not investigate the time course of attention per se, but rather focus on confidence when attention was in its prime. In the present work, we specifically investigated the dynamics of confidence regarding post-orientating attentional mechanisms, when spatial attention is disengaging from a given location. It appeared that confidence did adequately track attentional dynamics, with no particular cost when attention was just recently reoriented.

We did however find a slight but significant decrease in metacognitive ability for longer ISI. This result cannot be solely reduced the decrease in attentional bias between target and distractor locations. One could argue that since the confidence judgment was based on the difference between the two probes, any decrease in this error difference would result in lower evidence for the metacognitive judgment. Yet, we controlled for this potential confound, by using the measure of metacognitive ability with trial-by-trial scaled error difference, and checking for a clear linear relationship between the probability of high confidence judgment and this relative error. If the errors' range is not strongly affecting the slope of the model, another mechanism must be involved. One possible explanation could be the intrinsic relation metacognition has with the timing of both temporal and spatial attention: we recently found that metacognition is oblivious to the latency of selective attention, a phenomenon leading to dissociations between accuracy and confidence (Recht, Mamassian, & de Gardelle, submitted; Recht, Mamassian, & de Gardelle, 2019). Moreover, a decrease in metacognitive ability was observed at the boundaries of an attentional episode: during the orienting of spatial attention (Recht et al., submitted), and during the reallocation of temporal attention to a second target (Recht et al., 2019). A similar metacognitive cost might be occurring here during the disengagement of spatial attention. Hence, this cost appeared qualitatively modest, and should not overshadow the notable ability of metacognition to track the effect of attentional disengagement.

Perspectives and further work

The notion of reorienting to novel stimuli has been presented as an important process in human decision making (Corbetta & Shulman, 2002; Posner, 1980; Sara & Bouret, 2012). Reorienting can be defined as the redirecting of attention and/or other cognitive resources for the processing of a new and unexpected stimulus (Corbetta et al., 2008). From a purely semantic perspective, orienting and reorienting processes could be considered as similar, since any orienting of attention should, in principle, be following a previous orienting episode. In the present study, we had primarily focused on the general linear trend of spatial attention disengagement and its effects on confidence. Yet, recent work on the fine-grained temporal dynamics of attention suggests that spatial attention samples the environment rhythmically at approximately 8 Hz in the theta frequency band (Dugué, McLelland, Lajous, & VanRullen, 2015; Dugué et al., 2016; Fiebelkorn, Saalmann, & Kastner, 2013; Landau & Fries, 2012; Senoussi et al., 2019). In this context, each relevant location in the visual field will be processed sequentially, at a speed of about 250ms when two locations are monitored. This attention-dependent rhythmic sampling has been shown to induce behavioural oscillations in accuracy, with the accuracy rate fluctuating in anti-phase between the two locations (Dugué et al., 2015, 2016; Fiebelkorn et al., 2013; Landau & Fries, 2012; Senoussi et al., 2019). Interestingly, these oscillations of performance are mainly observable following the early reorienting of endogenous attention when the pre-cue is invalid (Dugué et al., 2016; Senoussi et al., 2019). How does confidence react to such periodic effects remain a largely open question. While a version of our paradigm has been previously used to study attentional rhythms (Senoussi et al., 2019), the number of participants in the present study (n = 5) does not give enough statistical power to conduct such a spectral analysis of behavioural data. In a follow-up experiment, we are planning to use the present paradigm to study the potential effect of attentional rhythms on confidence and metacognitive ability with a larger cohort of participants.

For this further objective, our paradigm specifically targeted the invalid condition in the grating discrimination task. In the analyses presented here, the invalid condition showed greater probe report errors overall compared to the valid condition, suggesting a slight cost in reorienting attention to the novel, target location. Yet, both the valid and invalid conditions elicited lower errors at the target location compared to the distractor location. This result suggests that early reorienting did not significantly change the time course of attentional disengagement. Interestingly, the confidence pattern was also similar for both valid and invalid conditions, with only a monotonic decrease in time mirroring attentional disengagement. Confidence judgments were therefore able to adapt to the early reorienting of attention with no significant metacognitive cost.

To tie our results in with the existent literature, spatial and temporal attention have been often considered in separated experimental contexts. In the temporal domain, our results might be compatible with the idea that the visual stream is divided into temporal perceptual episodes. This process can be described as the deployment of a selection window (Gaussian-smoothed) sliding over stimuli as time passes, and putting an upper bound limit on the individuation of stimuli presented at high speed. The role of temporal attention in the shaping of these episodes is, however, subject to ongoing debate. One line of thought proposes that attention directly determines the onset and size of the perceptual episode (e.g., Wyble Brad et al., 2011), or/and involves a trade-off between different points in time (Denison et al., 2017). Another account suggests that these episodes are perceptual in nature, and that attention is not directly affecting the size of the selection window, but rather freezes - like a snapshot - the episode considered the most temporally relevant (Martini, 2012; Snir & Yeshurun, 2017). In the latter case, attention would only select, but not alter, the perceptual content of an episode. In the present study, it would not possible to distinguish one account from the other, and both explanations could very well hold. If we consider the invalid condition, in both cases the attentional/perceptual episode would have its peak after target offset. Accordingly, the episode would encompass not only the target of the first task, but also the upcoming probe sharing the same location. Such a process could explain why the probe on the target side was more accurately reported than the probe on the distractor side, if we assume that perceptual episodes are spatially selective (Wyble et al., 2009). In the present work, both the temporal and spatial dimensions were affecting the selection process: the enhancement at target location was estimated as a function of the error at distractor location across time. Thus, perceptual episodes could be restrictive in both time and space: when attention is engaged at one location, disengagement takes time and might lead to the residual facilitation of an incoming stimulus at previous location. These residual effects were detectable up to 520 ms after target offset. The notable length of this time interval might be explained by the absence of any masks/distractors intercalated between the first and the second task, which might hinder the selection episode from expanding (Wyble et al., 2011). More generally, the observed bond that seems to exist between temporal and spatial attention brings to question the need for a strict taxonomy differentiating the two processes (Anna C. Nobre & van Ede, 2017).

Conclusion

Here, we observed that confidence was able to monitor the progressive disengagement of attention from a previous covert endogenous episode. In particular, metacognitive judgments were predictive of the trial-by-trial fluctuation in error difference between target and distractor locations.

Metacognitive ability decreased with disengagement, suggesting a specific role of selective attention on metacognition. Finally, confidence also adapted to the abrupt reorienting of attention elicited by invalid pre-cues, confirming the tight bound that is likely to exist between confidence and spatiotemporal attentional mechanisms.

References

Baguley, T. (2012). Calculating and graphing within-subject confidence intervals for ANOVA. Behavior Research Methods. https://doi.org/10.3758/s13428-011-0123-7

Barthelme, S., & Mamassian, P. (2010). Flexible mechanisms underlie the evaluation of visual confidence. Proceedings of the National Academy of Sciences, 107(48), 20834–20839. https://doi.org/10.1073/pnas.1007704107

Barthelmé, S., & Mamassian, P. (2009). Evaluation of Objective Uncertainty in the Visual System. PLoS Computational Biology, 5(9), e1000504. https://doi.org/10.1371/journal.pcbi.1000504

Carrasco, M. (2011). Visual attention: The past 25 years. Vision Research. https://doi.org/10.1016/j.visres.2011.04.012

Corbetta, M., & Shulman, G. L. (2002). Control of goal-directed and stimulus-driven attention in the brain. Nature Reviews Neuroscience. https://doi.org/10.1038/nrn755

Cousineau, D. (2005). Confidence intervals in within-subject designs: A simpler solution to Loftus and Masson's method. Tutorials in Quantitative Methods for Psychology. https://doi.org/10.20982/tqmp.01.1.p042

de Gardelle, V., Le Corre, F., & Mamassian, P. (2016). Confidence as a Common Currency between Vision and Audition. PloS One, 11(1), e0147901. https://doi.org/10.1371/journal.pone.0147901

de Gardelle, V., & Mamassian, P. (2014). Does Confidence Use a Common Currency Across Two Visual Tasks? Psychological Science, 25(6), 1286–1288. https://doi.org/10.1177/0956797614528956

Denison, R. N., Adler, W. T., Carrasco, M., & Ma, W. J. (2018). Humans incorporate attention-dependent uncertainty into perceptual decisions and confidence. Proceedings of the National Academy of Sciences, 201717720. https://doi.org/10.1073/pnas.1717720115

Denison, R. N., Heeger, D. J., & Carrasco, M. (2017). Attention flexibly trades off across points in time. Psychonomic Bulletin & Review, 24(4), 1142–1151. https://doi.org/10.3758/s13423-016-1216-1

Dugué, L., McLelland, D., Lajous, M., & VanRullen, R. (2015). Attention searches nonuniformly in space and in time. Proceedings of the National Academy of Sciences, 112(49), 15214–15219. https://doi.org/10.1073/pnas.1511331112

Dugué, L., Merriam, E. P., Heeger, D. J., & Carrasco, M. (2018). Endogenous and exogenous attention distinctly modulate fMRI activity in visual cortex. BioRxiv.

Dugué, L., Roberts, M., & Carrasco, M. (2016). Attention reorients periodically. Current Biology, 26(12), 1595-1601. https://doi.org/10.1016/j.cub.2016.04.046

Fiebelkorn, I. C., Saalmann, Y. B., & Kastner, S. (2013). Rhythmic sampling within and between objects despite sustained attention at a cued location. Current Biology, 23(24), 2553–2558. https://doi.org/10.1016/j.cub.2013.10.063

Fleming, S. M., & Daw, N. D. (2017). Self-evaluation of decision performance: A general Bayesian framework for metacognitive computation. Psychological Review, 124(1), 1–59. https://doi.org/10.1002/smll.))

Fleming, S. M., & Lau, H. C. (2014). How to measure metacognition. Frontiers in Human Neuroscience, 8(July), 1–9. https://doi.org/10.3389/fnhum.2014.00443

Griffin, I. C., & Nobre, A. C. (2003). Orienting Attention to Locations in Internal Representations. Journal of Cognitive Neuroscience. https://doi.org/10.1162/089892903322598139

Kurtz, P., Shapcott, K. A., Kaiser, J., Schmiedt, J. T., & Schmid, M. C. (2017). The Influence of Endogenous and Exogenous Spatial Attention on Decision Confidence. Scientific Reports, 7(1). https://doi.org/10.1038/s41598-017-06715-w

Landau, A. N., & Fries, P. (2012). Attention samples stimuli rhythmically. Current Biology, 22(11), 1000–1004. https://doi.org/10.1016/j.cub.2012.03.054

Mamassian, P. (2016). Visual Confidence. Annual Review of Vision Science, 2(1), 459–481. https://doi.org/10.1146/annurev-vision-111815-114630

Martini, P. (2012). Sources of bias and uncertainty in a visual temporal individuation task. Attention, Perception, and Psychophysics, 75(1), 168–181. https://doi.org/10.3758/s13414-012-0384-y

Morey, R. D. (2008). Confidence Intervals from Normalized Data: A correction to Cousineau (2005). Tutorials in Quantitative Methods for Psychology. https://doi.org/10.20982/tqmp.04.2.p061

Nobre, A. C., & van Ede, F. (2017). Anticipated moments: temporal structure in attention. Nature Reviews Neuroscience. https://doi.org/10.1038/nrn.2017.141

Peirce, C., & Jastrow, J. (1884). On small differences in sensation. Retrieved from https://philarchive.org/archive/PEIOSD

Posner, M. I. (1980). Orienting of attention. Quarterly Journal of Experimental Psychology. https://doi.org/10.1080/00335558008248231

Rahnev, D., Maniscalco, B., Graves, T., Huang, E., de Lange, F. P., & Lau, H. (2011). Attention induces conservative subjective biases in visual perception. Nature Neuroscience, 14(12), 1513–1515. https://doi.org/10.1038/nn.2948

Recht, S., Mamassian, P., & de Gardelle, V. (submitted). Orienting spatial attention weakens metacognition.

Recht, Samuel, Mamassian, P., & de Gardelle, V. (2019). Temporal attention causes systematic biases in visual confidence. Scientific Reports, 9(1), 11622. https://doi.org/10.1038/s41598-019-48063-x

Ruff, C. C., Kristjánsson, Á., & Driver, J. (2007). Readout from iconic memory and selective spatial attention involve similar neural processes. Psychological Science. https://doi.org/10.1111/j.1467-9280.2007.01998.x

Sara, S. J., & Bouret, S. (2012). Orienting and Reorienting: The Locus Coeruleus Mediates Cognition through Arousal. Neuron. https://doi.org/10.1016/j.neuron.2012.09.011

Senoussi, M., Moreland, J. C., Busch, N. A., & Dugué, L. (2019). Attention explores space periodically at the theta frequency. Journal of Vision. https://doi.org/10.1167/19.5.22

Sergent, C., Ruff, C. C., Barbot, A., Driver, J., & Rees, G. (2011). Top–down modulation of human early visual cortex after stimulus offset supports successful postcued report. Journal of Cognitive Neuroscience, 23(8), 1921-1934.

Sergent, C., Wyart, V., Babo-Rebelo, M., Cohen, L., Naccache, L., & Tallon-Baudry, C. (2013). Cueing attention after the stimulus is gone can retrospectively trigger conscious perception. Current Biology. https://doi.org/10.1016/j.cub.2012.11.047

Shepard, R. N. (1987). Toward a universal law of generalization for psychological science. Science. https://doi.org/10.1126/science.3629243

Snir, G., & Yeshurun, Y. (2017). Perceptual episodes, temporal attention, and the role of cognitive control: Lessons from the attentional blink. In Progress in Brain Research. https://doi.org/10.1016/bs.pbr.2017.07.008

Van Den Berg, R., Shin, H., Chou, W. C., George, R., & Ma, W. J. (2012). Variability in encoding precision accounts for visual short-term memory limitations. Proceedings of the National Academy of Sciences of the United States of America. https://doi.org/10.1073/pnas.1117465109

Wilimzig, C., Tsuchiya, N., Fahle, M., Einhäuser, W., & Koch, C. (2008). Spatial attention increases performance but not subjective confidence in a discrimination task. Journal of Vision, 8(5), 7. https://doi.org/10.1167/8.5.7

Wyble, B., Bowman, H., & Nieuwenstein, M. (2009). The Attentional Blink Provides Episodic Distinctiveness: Sparing at a Cost. Journal of Experimental Psychology: Human Perception and Performance, 35(3), 787–807. https://doi.org/10.1037/a0013902

Wyble Brad, B., Potter, M. C., Bowman, H., & Nieuwenstein, M. (2011). Attentional episodes in visual perception. Journal of Experimental Psychology: General. https://doi.org/10.1037/a0023612

Zizlsperger, L., Sauvigny, T., & Haarmeier, T. (2012). Selective attention increases choice certainty in human decision making. PLoS ONE. https://doi.org/10.1371/journal.pone.0041136

Zizlsperger, L., Sauvigny, T., Händel, B., & Haarmeier, T. (2014). Cortical representations of confidence in a visual perceptual decision. Nature Communications. https://doi.org/10.1038/ncomms4940

CHAPTER 3 | TEMPORAL ATTENTION CAUSES SYSTEMATIC BIASES IN VISUAL CONFIDENCE

In the previous chapter, our results attest the potent role of the temporal structure of spatial attention in shaping perceptual confidence. Yet, to better understand this influence, we would need to manipulate the timing of attention independently of task requirements, in order to induce conflicts between the state of attention and the ability to perform in the task. In the present chapter, we adapt a classic Attentional Blink paradigm to induce discontinuities in the orientation of temporal attention. This approach thus enables us to investigate how confidence reacts when attention is pushed to its limits, selecting the wrong stimulus in time.

Temporal attention causes systematic biases in visual confidence

Authors: Samuel Recht (1)*, Pascal Mamassian (1), Vincent de Gardelle (2)

Affiliations:

(1) Laboratoire des systèmes perceptifs, Département d'études cognitives, École normale supérieure, PSL University, CNRS, 75005 Paris, France

(2) CNRS and Paris School of Economics, Paris, France

* Corresponding author: samuel.recht@gmail.com

Abstract

Temporal attention enhances the perceptual representation of a stimulus at a particular point in time. The number of possible attentional episodes in a given period is limited, but whether observers' confidence reflects such limitations is still unclear. To investigate this issue, we adapted an "Attentional Blink" paradigm, presenting observers with a rapid visual stream of letters containing two targets cued for subsequent perceptual reports and confidence judgments. We found three main results. First, when two targets fell within the same attentional episode, the second target underwent a strong under-confidence bias. In other words, confidence neglected that a single attentional episode can benefit to both targets. Second, despite this initial bias, confidence was strongly correlated with response probability. Third, as confidence was yoked to the evidence used in perceptual reports, it remains blind to delays in response selection for the second target. Notably, the second target was often mistaken with a later item associated with higher confidence. These results suggest that confidence does not perfectly evaluate the limits of temporal attention in challenging situations.

Paper published in *Scientific Reports* (2019).

Introduction

Visual confidence is the subjective estimation of the accuracy of a decision made about a visual stimulus (Mamassian, 2016). It typically correlates with the objective accuracy of the decision, and can be used to regulate behavior (Desender, Boldt, & Yeung, 2018; Guggenmos, Wilbertz, Hebart, & Sterzer, 2016; Hainguerlot, Vergnaud, & De Gardelle, 2018). However, humans do not always monitor their performance perfectly, and dissociations between confidence and performance have been documented (Graziano & Sigman, 2009; Koizumi, Maniscalco, & Lau, 2015; Maniscalco, Peters, & Lau, 2016; Peters et al., 2017; D. Rahnev et al., 2011). Here, our goal is to assess how observers' confidence and performance are affected when temporal attention is challenged, and whether confidence tracks the limits of temporal attention.

Temporal attention enhances a stimulus at a particular point in time (Coull & Nobre, 1998) and inhibits other time points (Denison et al., 2017), much like spatial attention does in space (Carrasco, 2011). Both attention and confidence are related to accuracy: attention increases the signal-to-noise ratio of the stimulus, while confidence ideally reflects this increase. Attention and confidence have already been studied together in the spatial domain, leading to mixed findings: some studies observed a dissociation between the two (D. Rahnev et al., 2011; Schoenherr et al., 2010; Wilimzig et al., 2008), while others suggested that spatial attention is well incorporated into confidence (Denison et al., 2018; Samuel Recht, de Gardelle, & Mamassian, 2017; Zizlsperger et al., 2012, 2014). In the time domain, this link between temporal attention and confidence remains largely unexplored. This question is particularly relevant given the possibility that attention and confidence might operate at different time scales (D. Rahnev et al., 2015).

In some circumstances, temporal attention can be suppressed, delayed or misplaced. One robust finding regarding the limits of temporal attention is the "Attentional Blink" (Broadbent & Broadbent, 1987; Raymond et al., 1992). Specifically, when two targets are embedded in a rapid serial visual presentation stream, the second target T2 is often missed when it appears soon (150-300ms) after the first target T1. When temporal selection is not simply suppressed in the case of missed T2 targets, it is delayed, such that an item following T2 would be reported instead. These selection delays, sometimes known as "post-target error intrusions" (M M Chun, 1997; Vul, Hanus, et al., 2008) are a second feature of the Attentional Blink. Finally, when T2 is presented immediately after T1 (60-100ms), then both targets are on average accurately reported. This effect, coined the "lag-1 sparing" (Hommel & Akyürek, 2005) is a third feature of the Attentional Blink. These three features can be accounted for by a variety of models (Dux & Marois, 2009; Martens & Wyble, 2010). However, whether confidence tracks these three features remains an open empirical question.

To address this question, we used an Attentional Blink paradigm in combination with confidence judgments, in order to evaluate whether participants' confidence judgments about T2 reports would reflect the suppression of accuracy during the Attentional Blink, the sparing of accuracy at lag-1, and the delay in temporal selection that follows the Attentional Blink. We also collected confidence judgments for T1 as a comparison baseline. To measure errors and delays in temporal selection, we presented participants with a rapid stream of letters, and indicated two letters in the stream for later report. The serial position of each letter in the stream provided critical information on the point in time at which attention was deployed (Goodbourn et al., 2016; Martini, 2012; Vul, Nieuwenstein, et al., 2008). In other words, the present work proposes to investigate whether participants accurately evaluate the limits of their ability to deploy their attention at the right moment in time.

MATERIAL AND METHODS

PARTICIPANTS

39 adult volunteers were recruited from the Laboratoire d'Economie Expérimentale de Paris (LEEP) pool of participants (M ± SD = 25.5 ± 2.9 years old, 17 females). They all provided informed written consent prior to the experiment. The sample size was based on a recent study involving a highly similar

Attentional Blink paradigm (Goodbourn et al., 2016). The present experiment was also replicated with a similar sample size (see Experiment 2 in Supplementary Material). Four observers were discarded because of a technical problem, and three participants were removed because of extremely low accuracy rate for target 1 or 2 (exclusion criterion: <10% accuracy), leaving 32 participants for analysis. Observers were paid a base sum (10 EUR) plus a bonus depending on their performance in the task (up to 10 EUR in addition). The average payoff was 16.43 EUR (SD = 1.89) for a single 1.5 hours session. The experimental procedure received approval from the Paris School of Economics (PSE) ethics review board and adhered to the principles of the Declaration of Helsinki.

Apparatus and stimuli

Participants sat approximately 60 cm from the screen (1280 ×1024 pixels, 60 Hz refresh rate). Stimuli were generated using the Python programming language and the PsychoPy toolbox (J. W. Peirce, 2007a) on a Windows XP computer. On each trial, participants were presented with a rapid serial visual presentation (RSVP) stream of the 26 English letters (Courier New, white font, 2.5° of visual angle) in the center of a black screen background (Fig. 1). Letters were randomized, and each letter was presented for 33ms (2 frames) with an inter-stimulus interval of 50ms (3 frames). Two letters in the stream were targets surrounded by a visual cue (white annulus, inner/outer diameter: 2.9°/3.1°), which appeared simultaneously with the target. The first target (T1) was located between the 5th and the 10th item in the stream, while the second target occurred at the 1st, 2nd, 3rd, 6th or 9th position after T1. Both target positions where counterbalanced with a full factorial design.

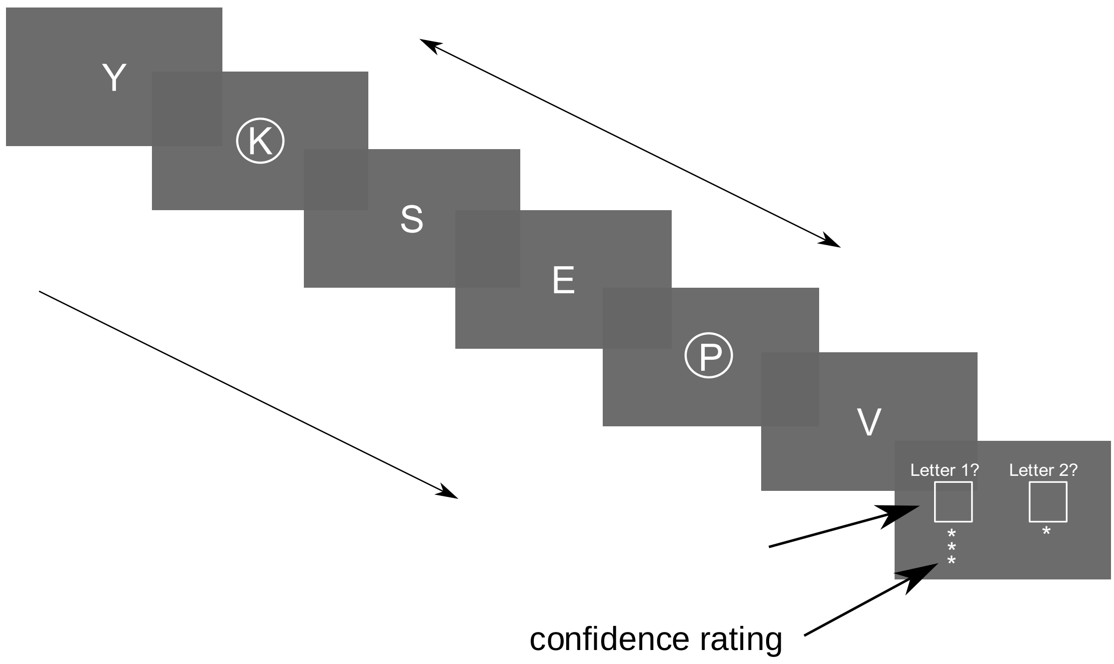

Figure 1. Experiment design. Participants were required to report the two cued letters in the RSVP, and rate their confidence for each reported letter (Experiment 1) or for only one of the letter (Experiment 2, see Supplementary Material) on a three-point scale. The distance in items (or lag) between the first target (T1) and second target (T2) was varied across trials (lag-3 depicted here). Each letter appeared for 33ms, followed by a 50ms ISI.

The lags between T1 and T2 were chosen in order to sample the different periods of the Attentional Blink: lag-1 (83ms after T1), where lag-1 sparing is known to occur; lags 2 and 3 (166ms and 249ms), which usually show strong drop in T2 reporting accuracy; and finally lags 6 through 9 (498ms and 747ms) that demonstrate a progressive recovery in accuracy.

PROCEDURE

At the end of each trial, participants had to report each target letter, in order of appearance, as well as their confidence for each report, using a French keyboard. Duplicates of the same letter were not accepted, given that each letter only appeared once in the stream. Confidence ratings were given on a 3-point scale using the numerical pad. For T1 confidence, keys 1, 4 and 7 corresponded to low, medium and high confidence. For T2 confidence, keys 3, 6 and 9 corresponded to low, medium, and high confidence. The confidence rating given to each target was displayed as one to three stars appearing below each of the reported letters. Participants could correct their response and confidence as needed. Participants validated their responses by pressing the Shift key.

Confidence was also incentivized. Specifically, participants were informed that each of their responses would generate 1, 2 or 3 points depending on their confidence rating. Points will be considered "good" if the response is correct and worth 0.5 EUR, and "bad" for incorrect responses and worth 0 EUR. Every 25 trials, the computer would randomly draw one point from those generated by the participant in the past 25 trials. The randomly drawn point, which could be "good" or "bad", determines the reward for these 25 of trials. This approach was applied separately to T1 and T2 responses. At the end of the experiment, the sum of these draws was used to estimate the monetary reward of the participant. The goal of this procedure was to engage participants in using confidence rating scale as accurately as possible during the whole experiment. High accuracy and good confidence estimates were therefore decisive to maximize payoff. Participants did not receive accuracy feedback until the very end of the experiment.

Before the main experiment, participants completed 10 practice trials, the first half without confidence judgments. The main session then consisted in 500 trials, with a 10-seconds break every 60 trials.

Analyses

All the analyses were carried out using the R programming language. Mixed effects models were built using the Lme4 R package. Accuracy and average confidence of T1 and T2 reports were analyzed using standard ANOVAs. In the current paradigm, the position of the reported item is also of interest. To analyze how reports and confidence depended on this serial position, a mixed effects model comparison approach was used. Specifically, a regression with fixed effects of position (and possibly other factors) and participants as random intercepts was compared to a regression without the fixed effect of position. When necessary, a third model including an interaction was added to the comparison.

Statistical results involving serial positions were systematically confirmed using permutation analysis, given the unbalanced nature of the dataset in this case. Serial positions were randomly shuffled for each participant and lag separately (for the whole dataset) and the relevant statistical analysis was applied to these surrogates data. The process was repeated 3,000 times, and the resulting distribution was compared to the test result obtained on the original data. P-values obtained through this method are reported as 'p_{RAND}'.

When necessary, ANOVAs were corrected using the Greenhouse-Geisser adjustment and t-tests were corrected using the Welch-Satterthwaite adjustment. We report Wilcoxon signed ranked test using uppercase T when the Shapiro-Wilk normality test failed, and Student test using lowercase t otherwise.

RESULTS

OVERVIEW

We start our result section by focusing on the first target (T1), which constitutes a baseline to evaluate how confidence links to reports when attention is unchallenged. In brief, for T1 we found that reports were distributed around the true position, and that confidence for these reports decreased with the distance to the target, following a bell-shaped profile similar to the one seen in report probability.

We then turn to our main results, which concern the second target (T2), known to be affected by the Attentional Blink. There are three main findings. First, both confidence and accuracy drop at lag-2 and lag-3, and confidence failed to reflect the sparing of accuracy at lag-1. Second, confidence was strongly correlated with the frequency of item selection (as was found for T1). A simple model for this correlation will be detailed in the discussion and simulations for this model can be found in the Supplementary Material. Our third result is that confidence was oblivious to the delays in item selection: after the Attentional Blink and up to lag-9, reports were systematically delayed relative to the target, and confidence was also shifted towards delayed responses, consistently with the correlation between confidence and frequency.

T1: Probability of report and confidence are strongly correlated

Overall, T1 targets were identified correctly 43% of the time. As can be seen on Figure 2A, and as documented previously (Vul, Nieuwenstein, et al., 2008), errors were not random guesses. The letter presented just before or just after the target was reported in 18% of the trials, largely exceeding the guess rate of 1/26≈4% ($t(31)=21.7$ $p<0.001$). Focusing on the 5 serial positions around T1 (included), we further tested how report frequency can be predicted from the position, the lag and their interaction (using mixed models, see Analyses). Including item position as a predictor outperformed a model without the position effect ($\chi^2(4)=1058$, $p_{RAND}<0.001$). Including the lag x position interaction improved the model even further ($\chi^2(16)=43.3$, $p_{RAND}=0.003$), but this interaction seemed specifically driven by the lag-1 as it disappeared when excluding this lag from the analysis ($\chi^2(16)=5.6$, $p_{RAND}=0.95$). The interaction between lag and position might reflect the confusion and order reversals that occur at lag-1 (see Supplementary Material).

One striking feature of the data is that confidence followed a profile similar to report frequency: when a specific position was reported more frequently, these reports were also associated with greater confidence (Fig. 2B). Confidence was significantly affected by item position ($\chi^2(4)=240$, $p_{RAND}<0.001$). Including the interaction between lag and position however did not improve the model ($\chi^2(16)=15.8$, $p_{RAND}=0.48$). We replicated these analyses while excluding correct responses, to confirm that these results did not merely reflect the ability to discriminate between correct and erroneous responses.

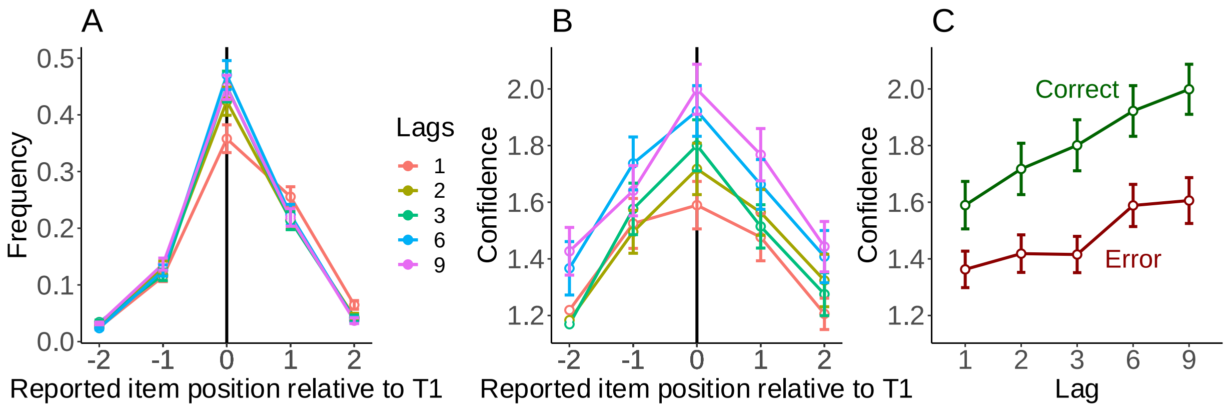

Figure 2. Reports and confidence for the first target. (A) The frequency of reports for item around target true position, separately for each lag. (B) The average confidence per position, for each lag. (C) The average confidence level for correct responses and errors, which provides an estimate of metacognition. Error bars represent standard error of the mean across participants.

To directly evaluate the similarity between confidence and report frequency, confidence was averaged for each participant by grouping all lags together, and we correlated this average confidence to the report frequency, across the 5 report positions centered on the target (including the target's position). The mean r coefficient was 0.86, across participants (95% CI=[0.82 0.90]; t(31)=44.2, p_{RAND}<0.001). Thus, it appears that participants' confidence is closely linked to the probability with which the reported letter is selected.

One typical signature of metacognition is the difference of confidence between correct and incorrect reports, with higher confidence for correct responses. Figure 2C illustrates this measure for the different lags. A repeated-measures ANOVA with lag and trial type (correct vs. error) revealed a main effect of trial type (F(1,31)=77.8, MSE=0.11, p<0.001), a main effect of lag (F(2.04,63.4)=38.2, MSE=0.06, p<0.001), as well as a lag x type interaction (F(3.35,104)=5.7, MSE=0.02, p<0.001). Overall, participants gave higher confidence to correct than to incorrect T1 responses. This difference between trial types increased with the lag between T1 and T2, but was present for all lags (all p<0.01, alpha = 0.05/5, Bonferroni-corrected for 5 comparisons).

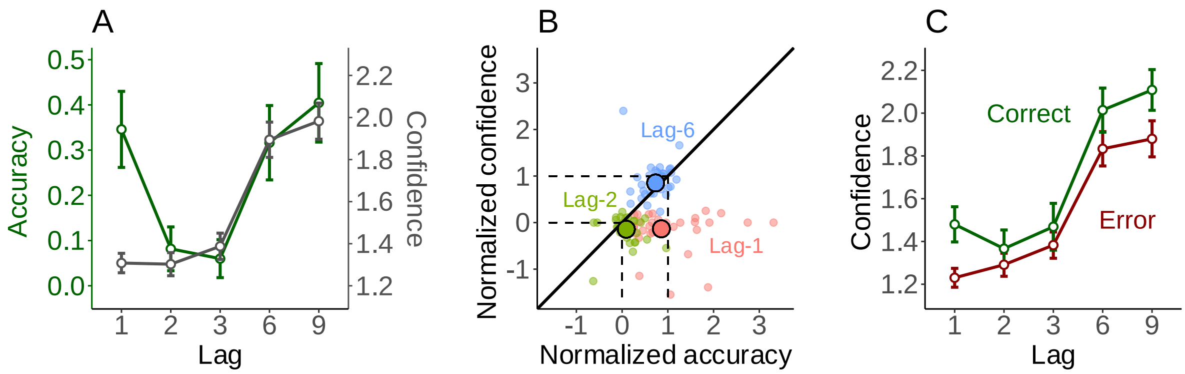

Figure 3. Attentional Blink and early confidence bias. (A) T2 average accuracy (in green) and confidence (in grey) as a function of the lag between T1 and T2. (B) The systematic under-confidence occurring at lag-1 (83ms after the first target) is illustrated by representing accuracy and confidence for lag-1 (in red) in the space from lag-3 to lag-9. The dashed lines represent (0,0) coordinates corresponding to lag-3 and (1,1) coordinates corresponding to lag-9 in this space. As a comparison, lag-2 (in green) and lag-6 (in blue) are pictured as well. Each colored point is a participant in the considered condition. The means for each condition are black-circled. Points below the diagonal represent under-confidence. (C) The average confidence level for correct T2 reports and errors, for each lag. Metacognitive sensitivity is conserved at lag-1 despite a bias for low confidence ratings. Error bars represent standard error of the mean across participants.

T2: Confidence tracks the Attentional Blink but not Lag-1 sparing

We then analyzed reports and confidence judgment about T2 targets (see Fig. 3A). To make sure of a successful initial attentional capture by T1, we analyzed only trials in which T1 was correctly reported. In these trials, 23% of T2 reports were correct. Figure 3A shows T2 accuracy and confidence for the different T1-T2 lags. T2 accuracy was affected by the T1-T2 lag ($F(2.14,66.5)=67.2$, $MSE=0.02$, $p<0.001$) and exhibited the classical Attentional Blink effect: it dropped for lag-2 and lag-3 relative to longer lags (2-3 vs. 6-9: $T(31)=0$, $p<0.001$). Confidence was also affected by lag ($F(1.88,58.4)=92.4$, $MSE=0.08$, $p<0.001$) and dropped for lags 2-3 relative to longer lags (2-3 vs. 6-9: $T(31)=0$, $p<0.001$), paralleling accuracy. Thus, participants were able to acknowledge the drop of performance at lags 2-3 relative to longer lags.

Importantly however, participants' confidence was strongly dissociated from accuracy at lag-1. Confidence seemed blind to lag-1 sparing, a classical phenomenon where T2 accuracy at lag-1 is much higher than during the blink period (1 vs. 2-3: $T(31)=528$, $p<0.001$) and indistinguishable from long lags (1 vs. 6-9: $T(31)=260$, $p=0.95$). Indeed, lag-1 confidence was as low as for lag 2-3 ($T(31)=197$, $p=0.66$) and much lower than for long lags (1 vs. 6-9: $T(31)=0$, $p<0.001$).

To further quantify this "lag-1 under-confidence", we asked whether the increase in accuracy at lag-1 relative to lag-3 was accompanied by the corresponding increase in confidence. Specifically, for each participant we regressed confidence against accuracy using lag-3 and lag-9 average data. The predicted confidence at lag-1 was then interpolated from the accuracy at lag-1, using this regression. Across participants, the observed confidence was significantly lower than the predicted confidence level (M=0.63, 95% CI=[0.45 0.81]; $t(31)=7.1$, $p<0.001$, alpha=0.05/3). For comparison, we also applied this approach to lag 2 and lag 6. Some under-confidence was found for lag-2 (M=0.14, 95% CI=[0.07 0.21]; $t(31)=3.9$, $p<0.001$, alpha=0.05/3). For lag-6 we found no difference between predicted and observed confidence (M=-0.07, 95% CI=[-0.13 0.003]; $t(31)=-1.9$, $p=0.06$, alpha=0.05/3).

Figure 2B illustrates this analysis by plotting confidence against accuracy, in the lag-3-to-9 space. For each participant, normalized accuracy was calculated as $(x_1-x_3)/(x_9-x_3)$, where x_k is the accuracy at lag-k, and the same procedure was done for confidence. For lag-1, all participants are located below the diagonal, suggesting that they are less confident than what could be expected given their accuracy. Figure 2B further illustrates how lag-6 and lag-1 differ in terms of confidence but not in terms of accuracy, whereas lag-2 and lag-1 differ in terms of accuracy but not in terms of confidence.

We then focused on metacognition, defined above as the difference in confidence between correct reports and errors. Because some participants had no correct answers at lag-2, only a subset of participants was considered here (N=25). As can be seen from Figure 3C, participants overall expressed higher confidence when they were correct and higher confidence at longer lags. A repeated-measures ANOVA with lag and trial type (correct vs. error) confirmed these two main effects (error vs correct: F(1,24)=11, MSE=0.15, p=0.002; lag: F(2.37,56.92)=58.5, MSE=0.15, p<0.001) and indicated an interaction (F(3.46,83.1)= 3.28 MSE=0.05, p=0.02). Post-hoc Bonferroni-corrected tests (alpha=0.05/5) showed that the difference in confidence between correct reports and errors was significant for lag-1 (t(24)=3.7, p=0.001), lag-6 (t(24)=3.1, p=0.004) and lag-9 (t(24)=4.3, p<0.001) but not for lag-2 (t(24)=1.4, p=0.18) or lag-3 (t(24)=0.1, p=0.89). In other words, the ability to detect objective errors was diminished specifically during the Attentional Blink period. Note that this is not surprising given the well-known relation between metacognitive sensitivity and task performance (Stephen M. Fleming & Lau, 2014). Interestingly, it did not disappear at lag-1, despite the low level of confidence.

T2: Probability of report and confidence are strongly correlated

Similarly, to T1, errors for T2 reports were not random guesses but distributed around the correct target position. In particular, items appearing just before or just after the target were reported more often than chance (17%, with a 95% CI=[0.16 0.18]; vs. chance level at 4%: t(31)=19.5, p<0.001). Comparing Figures 4A and 4B, we note that for each lag confidence and report frequency typically peak at the same item position, even when this item position is not the target position. This similarity between confidence and report frequency across positions was examined for each individual participant, by considering 5 positions centered on T2, after averaging across lags. Figure 4C shows a representative participant and Figure 4D shows the distribution of correlation coefficients at the group level, which confirm the strong relation between confidence and report frequency (Mean r coefficient: 0.82, 95% CI=[0.76 0.89]; t(31)=25.3, p_{RAND}<0.001). A correlation between confidence and log-frequency provided equivalent results.

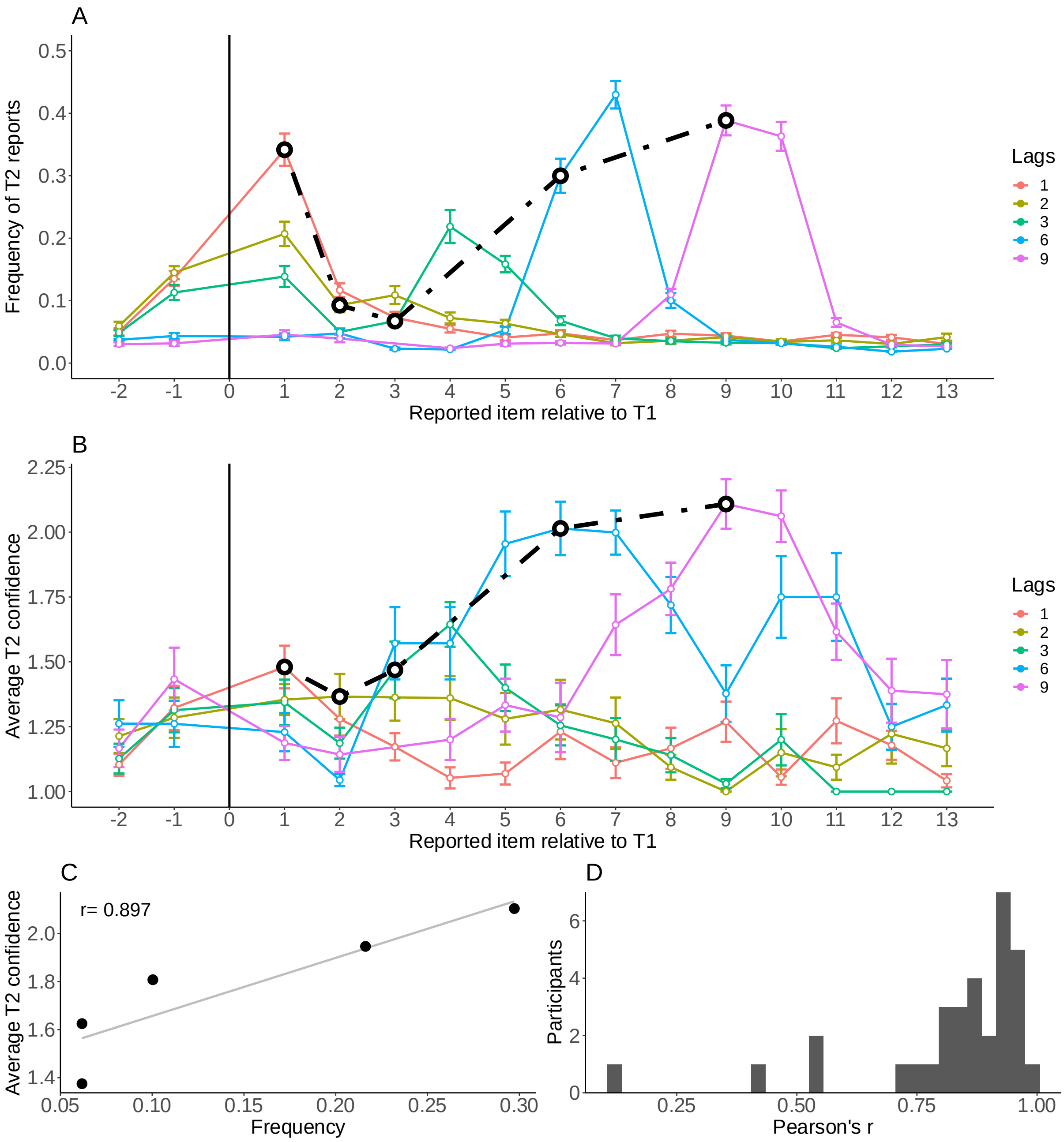

Figure 4. Reports and confidence for the second target. (A) The frequency of T2 reports as a function of the position of the reported item relative to T1, for each lag. Note that T1 position has no value, given that only trials in which T1 is correctly reported were considered here (hence T2 reports cannot correspond to T1 position). The black line connects the points corresponding to accurate T2 reports. (B) Confidence of the T2 reports, as a function of the position of the reported item relative to T1, for each lag. The black line connects the points corresponding to accurate T2 reports. Error bars represent standard error of the mean across participants. (C) Regression between frequency and confidence with 5 positions centered on T2, collapsed across lags, for a representative participant. (D) Histogram of the correlation coefficients for all the participants. The confidence-frequency relation is strong and holds for most participants.

T2: CONFIDENCE DOES NOT CORRECT FOR ATTENTIONAL DELAY

Attention is typically delayed after the Attentional Blink, as participants tend to report items that follow the target rather than the target itself. To analyze the delay in selection and confidence induced by the reorienting of attention (T2), we calculated the average position of the reported item relative to the target position, in an 11-items window centered on the target position. This measure, called the "center of mass" was positive for lags 6 and 9, showing that a delay occurred in item selection, as found in previous studies (Goodbourn et al., 2016; Vul, Hanus, et al., 2008; Vul, Nieuwenstein, et al., 2008) (see Supplementary Material). Given that confidence was correlated with report frequency, we investigated whether confidence was similarly shifted towards delayed selections. To do so, we calculated the average confidence for reports corresponding to late selections ("post-target" errors) minus the average confidence for early selections ("pre-target" errors). This "confidence shift" (Fig. 5) was evaluated over an 11-items window centered on (but excluding) the target position, separately for each lag. A model comparison approach confirmed that including the pre-target/post-target factor as a predictor for average confidence significantly outperformed the null model ($\chi^2(1)=27.1$, $p_{RAND}<0.001$). The interaction between lag and shift was also significant ($\chi^2(4)=34.8$, $p_{RAND}<0.001$). T-tests (Bonferroni-corrected for 5 lags with alpha=0.05/5) confirmed a significant delay for lag-3 ($t(31)=3.13$, $p=0.004$), lag-6 ($T(31)=406$, $p<0.001$) and lag-9 ($T(31)=354$, $p<0.001$) but not for lag-1 and lag-2 (all $p>0.3$). For comparison, this analysis showed no confidence shift when applied to T1 ($\chi^2(4)=0.3$, $p_{RAND}=0.5$).

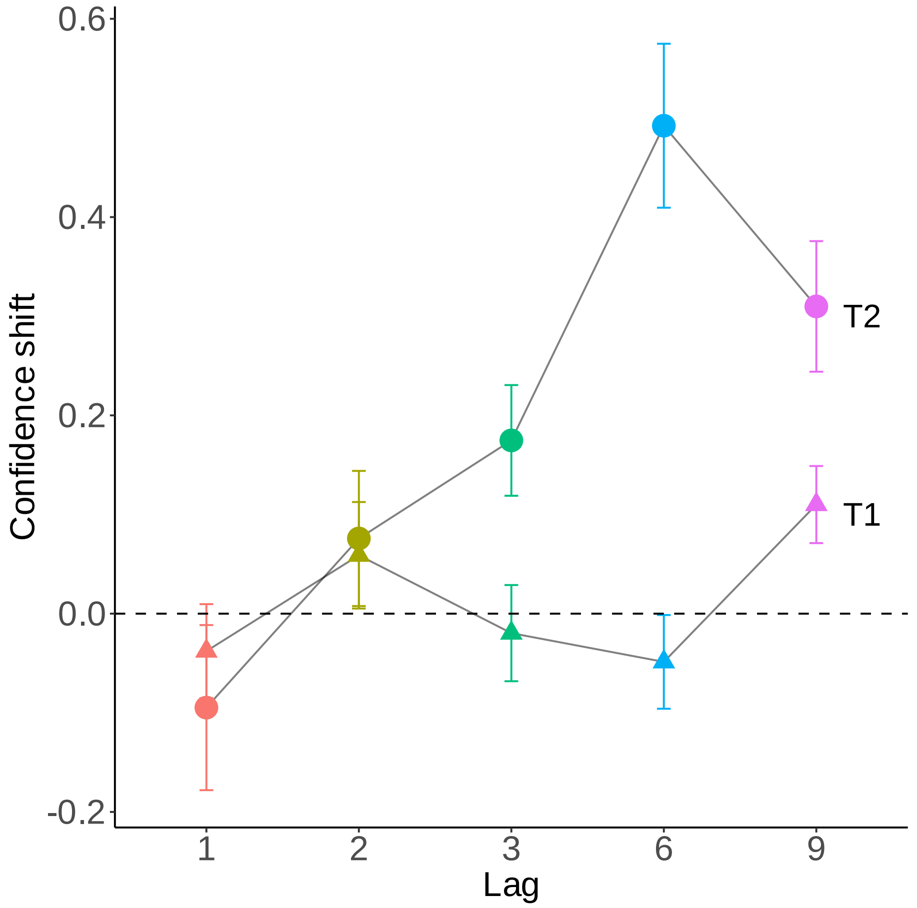

Figure 5. Confidence for T2 is delayed. Confidence shift is the average confidence in post-target minus pre-target errors, evaluated separately for each lag and for T1 (triangles) and T2 (dots). A positive value corresponds to greater confidence for post-target errors, that is, a shift of the confidence peak towards more delayed items. T2 confidence is delayed for lags 3, 6 and 9, reproducing the delay generally observed in items selection after the Attentional Blink period (see Fig. 4A and Supplementary Material). Error bars represent standard error of the mean across participants.

A REPLICATION WITH A REDUCED METACOGNITIVE LOAD (EXP. 2)

In Experiment 1, participants reported their confidence for both T1 and T2 targets in each trial. The high demand put on the metacognitive system during the task might explain why confidence failed to track the lag-1 sparing or the delays in item selection induced by the Attentional Blink. To address this possibility, we conducted a second experiment in which we lowered the demands put on the metacognitive system, by asking only one confidence estimate per trial. In experiment 2, participants (N=29) gave their confidence about T1 in the first half of the experiment and their confidence about T2 in the other half (or vice-versa, counterbalanced across participants). All other parameters were identical to Experiment 1, and performance levels in Experiment 2 were similar to Experiment 1, with an average accuracy at 40% for T1 and at 22% for T2 after a correct T1 response (see Supplementary Fig. S1 and S2).

Critically, in Experiment 2 we replicated the three main findings of Experiment 1, as summarized below (for details see the Supplementary Material). First, participants were oblivious to lag-1 sparing and exhibited a clear under-confidence at lag-1 for their T2 reports (see Supplementary Fig. S3). Second, we replicated the finding that confidence was tied to report frequency for T1 (Supplementary Fig. S4). Hence, when a particular item was more likely to be selected, it was also reported with a greater confidence. Finally, both temporal selection and confidence were delayed after the Attentional Blink (Supplementary Fig. S5). In other words, whereas the metacognitive task was less demanding, participants were not better at acknowledging the lag-1 sparing or delays in temporal selection induced by the Attentional Blink.

Discussion

The present study considered how human observers could evaluate their own performance in a task in which temporal attention has to be oriented towards two targets (T1 and T2) presented in close succession. To do so, confidence judgments were introduced within an Attentional Blink paradigm, and we analyzed how such judgments would track the limits of performance typically observed in this paradigm. We obtained three main results. First, participants failed to notice the early sparing of accuracy at lag-1, despite being able to detect the drop of accuracy at lag-2 and lag-3. Second, participants' confidence when reporting an item systematically followed the probability of selecting this item in the sequence. Third, and likely because of this confidence-probability coupling, participants were oblivious to the delays in temporal selection induced by the

Attentional Blink. All these results were replicated in a second experiment in which we only collected one confidence judgment (either for T1 or for T2), to reduce the demands put on the metacognitive system.

CONFIDENCE IS BLIND TO LAG-1 SPARING

Surprisingly, confidence was not able to track the sparing of accuracy known to occur when the two targets are very close in time. However, we note that metacognition was not particularly altered during lag-1: participants still discriminated between correct responses and errors, and between different errors (Fig. 3). This under-confidence is therefore not due to participants being unable to use their metacognition. Nonetheless, confidence did not adjust to lag-1 sparing, despite its ability to track the drop in accuracy during lag-2 and lag-3, and the progressive recovery for longer lags. A confidence cost was systematically applied to all responses for lag-1, and this early under-confidence bias was present for almost every participant.

One possibility is that the under-confidence bias at lag-1 results from participants being aware of possible order reversals, where T1 would be reported as T2 and vice-versa due to temporal selection uncertainty (see Supplementary Material). Order reversals have been documented in the literature, and it has been suggested that at lag-1, T2 would actually benefit from the T1 attentional episode, the two targets being often perceived as a single object (Akyürek et al., 2012; Goodbourn et al., 2016; Hommel & Akyürek, 2005), at the cost of an increased uncertainty about their relative order. This increased uncertainty could lead participants to express lower confidence.

Our confidence data at lag-1 seem to mirror what was found for visibility in a recent study that suggested lower visibility despite high accuracy (Pincham, Bowman, & Szucs, 2016) at lag-1. However, another study (C. Sergent & Dehaene, 2004) found that subjective visibility during lag-1 is spared. Besides these mixed findings for visibility, one might consider that confidence and visibility do not always go hand-in-hand, and can be dissociated both conceptually and empirically (Rausch & Zehetleitner, 2016; Rosenthal, 2018).

A SIMPLE MODEL OF THE CONFIDENCE-FREQUENCY RELATION

The second major result of our study is that confidence generally follows report frequency across the items in the sequence. This robust correlation was observed on both T1 and T2, and irrespectively of the T1-T2 lag or the delays induced by the Attentional Blink. This finding speaks to the ongoing debate regarding whether the same evidence signal is used for decisions and confidence, and the observed dissociations between confidence and accuracy (Graziano & Sigman, 2009; Koizumi et al., 2015; Maniscalco et al., 2016; Peters et al., 2017; D. Rahnev et al., 2011). In our study, the under-confidence at lag 1 illustrates such a dissociation, but seems to exist on top of the strong relation between confidence and reports, suggesting that decisions and confidence judgments are also relying on the same evidence signal (Fleming & Daw, 2017; Mamassian, 2016).

The robust confidence-frequency relation found in the present work could be well accounted for by a simple attentional selection mechanism within a RSVP stream, based on the Attentional Gating Model (Reeves & Sperling, 1986). In this model, the letters presented in the RSVP stream lead to a short-lasting activation of the corresponding letter-detectors in the perceptual system. When the cue appears, it triggers an attentional boost that enhances the response of the letter-detectors. This boost is smoothly distributed in time over several items. At the end of the sequence, the evidence for each item is the integral of the activity of the corresponding letter-detector, corrupted by random perturbations (i.e., noise). The item selected for report will be the one with maximum evidence. In fact, under the simple assumption that confidence relates to the amplitude of this evidence, a correlation between confidence and report frequency would occur across trials. To understand why, note that noise on evidence levels would move the peak evidence away from the correct target, thereby producing errors distributed around the target. These perturbations would also affect the confidence in these reports. Simulations of this process produced a correlation between confidence and report frequency across positions, as was found in our data. Details of this model are presented in the Supplementary Material (see Supplementary Fig. S6 – S10).

This proposed model accounts for (i) the correlation between report confidence and report frequency, (ii) the related observation that confidence is higher for correct responses than for errors, (iii) the finding that this metacognition is present mostly outside of the Attentional Blink and (iv) the result that confidence was blind to selection delays. However, it is important to highlight that this mechanism linking confidence and reports does not account for the under-confidence at lag-1. We believe that accommodating this last result would require additional components. Incorporating this mechanism within a full computational model of the Attentional Blink is a task for future research.

CONFIDENCE DOES NOT CORRECT FOR ATTENTIONAL DELAY

Our last result relates to the delayed attentional selection induced by the Attentional Blink. We found for both experiments a long-lasting delay in selection after the Attentional Blink, at lag-6 and lag-9, replicating previous findings (M M Chun, 1997; Vul, Hanus, et al., 2008). Confidence remained fully oblivious to this fundamental limitation of the attentional system, an expected result given the correlation found between confidence and report frequency (Fig. 4D).

There is a striking similarity between the present finding about confidence in the Attentional Blink paradigm and a finding about introspective response times in the Psychological Refractory Period paradigm (Corallo, Sackur, Dehaene, & Sigman, 2008; Marti, Sackur, Sigman, & Dehaene, 2010). In this paradigm, two tasks have to be conducted in short succession in time, and the decision process for the second task is postponed until the first decision process has been completed. Interestingly however, introspective estimates of response times are blind to this delay. It has been suggested that the Attentional Blink and Psychological Refractory Period paradigm involve a similar central bottleneck (Marti, Sigman, & Dehaene, 2012; Wong, 2002). Indeed, introspective measures of performance (respectively, confidence and subjective estimates of response times) appear to be oblivious to the delays presumably induced by this central bottleneck in both paradigms. To expand this research, future work might investigate whether introspection is blind to central delays in different paradigms, or to other constraints of central processing stages (e.g., the discrete/symbolic nature of information processing at central stages (de Gardelle, Charles, & Kouider, 2011; de Gardelle, Kouider, & Sackur, 2010)).

Conclusion

The strong correlation between frequency of reports and confidence during temporal selection (T1), which holds when attention has to reorient to a second point in time (T2), suggests that decision and confidence are mostly sharing the same evidence signal during the temporal orienting of attention. This tight coupling might prevent confidence from accessing delays in selection induced by the Attentional Blink, as shown in the present work. In addition, confidence seems to be affected by a heuristic penalizing a target that is too close in time from a prior attentional episode, a penalty that would account for the lag-1 under-confidence. These multiple phenomena suggest that confidence does not perfectly evaluate the state of temporal attention in challenging situations, likely because of late heuristic bias and the fact that confidence is yoked in time to temporal attention.

Funding

This research was supported by the Agence Nationale de la Recherche (ANR-16-CE28-0002 and ANR-16-ASTR-0014 to VdG, ANR-17-EURE-0017 to PM) and by PSL University (doctoral scholarship to SR).

Competing interests

The authors declare no competing interests.

Acknowledgements

We would like to thank Dobromir Rahnev, Monica Barbir, Shannon M. Locke and Thomas Schaffhauser for helpful comments during the drafting of this manuscript.

Contributions

SR, PM and VdG designed the experiment. SR conducted the experiment. SR, PM and VdG analyzed the data and wrote the manuscript.

Data availability

Data for both experiments have been made publicly available via Open Science Framework and can be accessed at https://osf.io/xjh2v/.

References

1. Mamassian, P. Visual Confidence. Annu. Rev. Vis. Sci. 2, 459–481 (2016).

2. Guggenmos, M., Wilbertz, G., Hebart, M. N. & Sterzer, P. Mesolimbic confidence signals guide perceptual learning in the absence of external feedback. Elife (2016). doi:10.7554/eLife.13388

3. Hainguerlot, M., Vergnaud, J. C. & De Gardelle, V. Metacognitive ability predicts learning cue-stimulus associations in the absence of external feedback. Sci. Rep. (2018). doi:10.1038/s41598-018-23936-9

4. Desender, K., Boldt, A. & Yeung, N. Subjective Confidence Predicts Information Seeking in Decision Making. Psychol. Sci. (2018). doi:10.1177/0956797617744771

5. Graziano, M. & Sigman, M. The Spatial and Temporal Construction of Confidence in the Visual Scene. PLoS One 4, e4909 (2009).

6. Rahnev, D. et al. Attention induces conservative subjective biases in visual perception. Nat. Neurosci. 14, 1513–1515 (2011).

7. Peters, M. A. K. et al. Perceptual confidence neglects decision-incongruent evidence in the brain. Nat. Hum. Behav. (2017). doi:10.1038/s41562-017-0139

8. Maniscalco, B., Peters, M. A. K. & Lau, H. Heuristic use of perceptual evidence leads to dissociation between performance and metacognitive sensitivity. Attention, Perception, Psychophys. (2016). doi:10.3758/s13414-016-1059-x

9. Koizumi, A., Maniscalco, B. & Lau, H. Does perceptual confidence facilitate cognitive control? Attention, Perception, Psychophys. (2015). doi:10.3758/s13414-015-0843-3

10. Coull, J. T. & Nobre, A. C. Where and when to pay attention: the neural systems for directing attention to spatial locations and to time intervals as revealed by both PET and fMRI. J. Neurosci. 18, 7426–7435 (1998).

11. Denison, R. N., Heeger, D. J. & Carrasco, M. Attention flexibly trades off across points in time. Psychon. Bull. Rev. 24, 1142–1151 (2017).

12. Carrasco, M. Visual attention: The past 25 years. Vision Research 51, 1484–1525 (2011).

13. Wilimzig, C., Tsuchiya, N., Fahle, M., Einhäuser, W. & Koch, C. Spatial attention increases performance but not subjective confidence in a discrimination task. J. Vis. 8, 7 (2008).

14. Schoenherr, J. R., Leth-Steensen, C. & Petrusic, W. M. Selective attention and subjective confidence calibration. Attention, Perception, Psychophys. (2010). doi:10.3758/APP.72.2.353

15. Denison, R. N., Adler, W. T., Carrasco, M. & Ma, W. J. Humans incorporate attention-dependent uncertainty into perceptual decisions and confidence. Proc. Natl. Acad. Sci. 201717720 (2018). doi:10.1073/pnas.1717720115

16. Zizlsperger, L., Sauvigny, T. & Haarmeier, T. Selective attention increases choice certainty in human decision making. PLoS One (2012). doi:10.1371/journal.pone.0041136

17. Zizlsperger, L., Sauvigny, T., Händel, B. & Haarmeier, T. Cortical representations of confidence in a visual perceptual decision. Nat. Commun. (2014). doi:10.1038/ncomms4940

18. Recht, S., de Gardelle, V. & Mamassian, P. Exogenous cues and visual confidence. J. Vis. 17, 670 (2017).

19. Rahnev, D., Koizumi, A., McCurdy, L. Y., D'Esposito, M. & Lau, H. Confidence Leak in Perceptual Decision Making. Psychol. Sci. (2015). doi:10.1177/0956797615595037

20. Broadbent, D. E. & Broadbent, M. H. P. From detection to identification: Response to multiple targets in rapid serial visual presentation. Percept. Psychophys. 42, 105–113 (1987).

21. Raymond, J. E., Shapiro, K. L. & Arnell, K. M. Temporary Suppression of Visual Processing in an RSVP Task: An Attentional Blink? J. Exp. Psychol. Hum. Percept. Perform. 18, 849–860 (1992).

22. Chun, M. M. Temporal binding errors are redistributed by the attentional blink. Percept. Psychophys. 59, 1191–9 (1997).

23. Vul, E., Hanus, D. & Kanwisher, N. Delay of selective attention during the attentional blink. Vision Res. 48, 1902–9 (2008).

24. Hommel, B. & Akyürek, E. G. Lag-1 sparing in the attentional blink: Benefits and costs of integrating two events into a single episode. Q. J. Exp. Psychol. Sect. A Hum. Exp. Psychol. 58, 1415–1433 (2005).

25. Dux, P. E. & Marois, R. The attentional blink: a review of data and theory. Atten Percept Psychophys 71, 1683–1700 (2009).

26. Martens, S. & Wyble, B. The attentional blink: Past, present, and future of a blind spot in perceptual awareness. Neuroscience and Biobehavioral Reviews 34, 947–957 (2010).

27. Goodbourn, P. T. et al. Reconsidering Temporal Selection in the Attentional Blink. Psychol. Sci. 27, 1146–1156 (2016).

28. Martini, P. Sources of bias and uncertainty in a visual temporal individuation task. Attention, Perception, Psychophys. 75, 168–181 (2012).

29. Vul, E., Nieuwenstein, M. & Kanwisher, N. Temporal selection is suppressed, delayed, and diffused during the attentional blink. Psychol. Sci. 19, 55–61 (2008).

30. Peirce, J. W. PsychoPy-Psychophysics software in Python. J. Neurosci. Methods 162, 8–13 (2007).

31. Fleming, S. M. & Lau, H. C. How to measure metacognition. Front. Hum. Neurosci. 8, 1–9 (2014).

32. Akyürek, E. G. et al. Temporal target integration underlies performance at Lag 1 in the attentional blink. J. Exp. Psychol. Hum. Percept. Perform. 38, 1448–1464 (2012).

33. Pincham, H. L., Bowman, H. & Szucs, D. The experiential blink: Mapping the cost of working memory encoding onto conscious perception in the attentional blink. Cortex 81, 35–49 (2016).

34. Sergent, C. & Dehaene, S. Is Consciousness a Gradual Phenomenon?: Evidence for an All-or-None Bifurcation During the Attentional Blink. Psychol. Sci. 15, 720–728 (2004).

35. Rosenthal, D. Consciousness and confidence. Neuropsychologia (2018). doi:10.1016/j.neuropsychologia.2018.01.018

36. Rausch, M. & Zehetleitner, M. Visibility Is Not Equivalent to Confidence in a Low Contrast Orientation Discrimination Task. Front. Psychol. 7, 591 (2016).

37. Mamassian, P. Visual Confidence. Annu. Rev. Vis. Sci. 2, annurev-vision-111815-114630 (2016).

38. Fleming, S. M. & Daw, N. D. Self-evaluation of decision performance: A general Bayesian framework for metacognitive computation. Psychol. Rev. 124, 1–59 (2017).

39. Reeves, A. & Sperling, G. Attention gating in short-term visual memory. Psychol. Rev. (1986). doi:10.1037/0033-295X.93.2.180

40. Corallo, G., Sackur, J., Dehaene, S. & Sigman, M. Limits on Introspection. Psychol. Sci. (2008). doi:10.1111/j.1467-9280.2008.02211.x

41. Marti, S., Sackur, J., Sigman, M. & Dehaene, S. Mapping introspection's blind spot: Reconstruction of dual-task phenomenology using quantified introspection. Cognition (2010). doi:10.1016/j.cognition.2010.01.003

42. Marti, S., Sigman, M. & Dehaene, S. A shared cortical bottleneck underlying attentional blink and psychological refractory period. Neuroimage (2012). doi:10.1016/j.neuroimage.2011.09.063

43. Wong, K. F. E. The relationship between attentional blink and psychological refractory period. J. Exp. Psychol. Hum. Percept. Perform. (2002). doi:10.1037/0096-1523.28.1.54

44. de Gardelle, V., Kouider, S. & Sackur, J. An oblique illusion modulated by visibility: Non-monotonic sensory integration in orientation processing. J. Vis. (2010). doi:10.1167/10.10.6

45. de Gardelle, V., Charles, L. & Kouider, S. Perceptual awareness and categorical representation of faces: Evidence from masked priming. Conscious. Cogn. (2011). doi:10.1016/j.concog.2011.02.001

CHAPTER 3 | SUPPLEMENTARY MATERIAL

Temporal attention causes systematic biases in visual confidence

Supplementary Material

Authors: Samuel Recht (1)*, Pascal Mamassian (1), Vincent de Gardelle (2)

Affiliations:

(1) Laboratoire des systèmes perceptifs, Département d'études cognitives, École normale supérieure, PSL University, CNRS, 75005 Paris, France

(2) CNRS and Paris School of Economics, Paris, France

*Corresponding author: samuel.recht@gmail.com

Experiment 1

T1: Position-based metacognition

In a finer analysis, we tested whether participants' confidence could discriminate between different errors across different serial positions, not just between correct and incorrect responses. Excluding correct T1 responses, we found that a regression model with position effect and lag outperformed the null model without the position for predicting confidence ($\chi^2(3)=101.2$, $p_{RAND}<0.001$), with no significant interaction between position and lag ($\chi^2(12)=7.99$, $p_{RAND}=0.78$). Participants are thus sensitive to the difference between various position errors, even if this distinction is irrelevant to succeed in the present task.

T2: Delay in attentional selection

To analyze the delay in selection and confidence following reorienting of attention to T2, we calculated the average position of the reported item relative to the target position, in an 11-items window centered on the target position. This measure, called the "center of mass" (Goodbourn et al., 2016; Vul, Nieuwenstein, & Kanwisher, 2008) is positive when a delay occurs in item selection. Figure S1 illustrates the average center of mass across participants, separately for each lag, and shows that T2 item selection is delayed specifically after the Attentional Blink (at lags 6 and 9), replicating previous findings (Goodbourn et al., 2016; Vul, Nieuwenstein, et al., 2008). A model comparison approach confirmed that including the lag as a predictor for the center of mass significantly outperformed the null model ($\chi^2(4)=56.9$, $p_{RAND}<0.001$). Bonferroni-corrected t-tests (alpha=0.05/5) confirmed a significant effect at lag-2 ($t(30)=-3.3$, $p=0.002$), lag-6 ($T(30)=506$, $p<0.001$) and lag-9 ($T(30)=527$, $p<0.001$), but not for lag 1 and 3 (all $p>0.6$). The non-linearity observed from lags 1 to 3 should be considered with caution: it could reflect both the interaction with T1 attentional episode (Goodbourn et al., 2016) and the bi-modality of lag-3 reports distribution (see Fig. 4A). A similar analysis on T1 confirmed a significant effect of lag on the center of mass as well ($\chi^2(4)=19.4$, $p_{RAND}<0.001$). This positive center of mass for T1 was not necessary predicted by the literature (Goodbourn et al., 2016; Vul, Hanus, & Kanwisher, 2008; Vul, Nieuwenstein, & Kanwisher, 2008) although some datasets show a similar tendency (e.g., Fig. S3 in the Supplementary Material of Goodbourn et al., 2016 and in particular the distribution of T1 latency for the

"Western", "Berkeley", and "Sydney words" datasets, as well as estimated delays in Martini, 2012). Interestingly, this delay disappeared in our replication with lowered metacognitive load (Exp. 2). The hypothesis that the observed T1 delay is the effect of (meta)cognitive load on selection would require further investigations. This positive delay, however, did not affected confidence (see below).

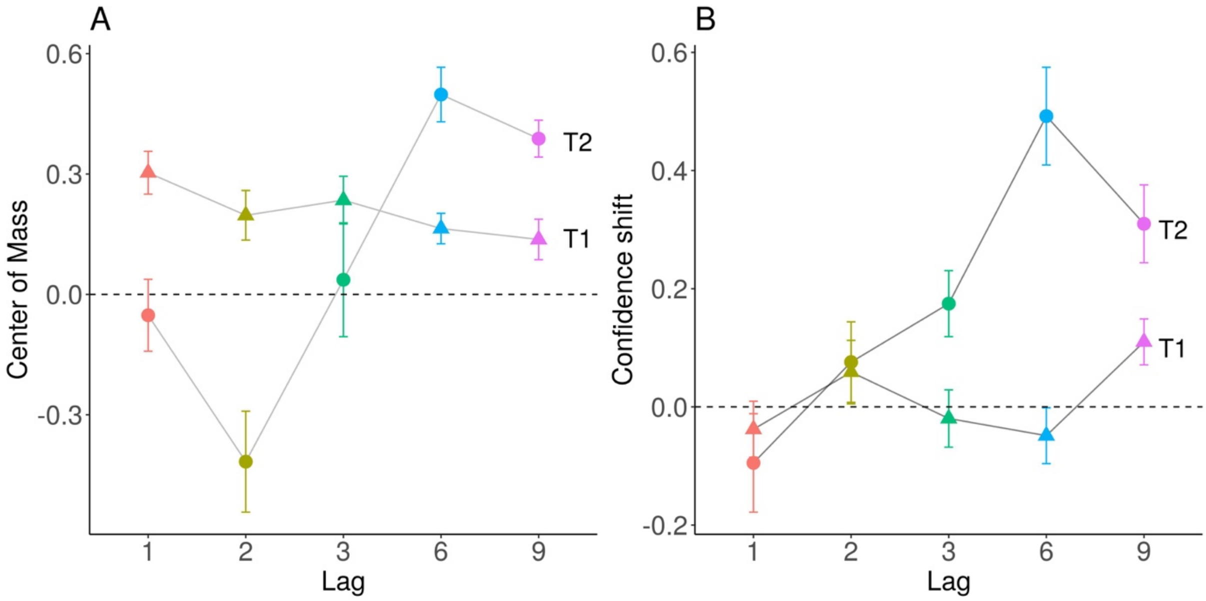

Figure S1. Delay is temporal selection. (A) The average center of mass for T1 (rectangles) and T2 (dots) as a function of lag. T2 center of mass is specifically delayed for lag-6 and lag-9. For T1, selection is slightly delayed but this remains stable across lags. (B) The confidence shift, which is the difference in average confidence between post-target and pre-target errors for T1 (triangles) and T2 (dots). A positive value corresponds to higher confidence for post-target errors, that is, a shift of the confidence peak towards more delayed items. Error bars represent standard error of the mean across participants.

Order reversals between T1 and T2

Order reversals occur at lag-1 when participants report both T1 and T2 but in the reverse order. In our data, order reversals occurred on average in 7.68% (SE ± 4.71%) of lag-1 trials. For comparison, correct report of both T1 and T2 in the correct order occurred in 12% (SE ± 5.77%) of lag-1 trials. To evaluate whether participants were aware of such reversals, the confidence between trials in which both T1 and T2 were correctly reported was compared to the confidence in reversed trials. One participant was discarded from this analysis due to no order reversal trial. No difference in confidence was found between these two types of trials, neither for T1 ($t(30)=1.07$, $p=0.29$) nor for T2 ($t(30)=1.20$, $p=0.24$). Thus, it seems that participants were not specifically aware of the occurrence or non-occurrence of a reversal on a trial-by-trial basis. However, it is still possible that participants could be aware of the possibility of order reversals at lag-1 relative to longer lags, and that being aware of this possibility would be responsible for the lag-1 under-confidence.

Experiment 2: a replication with lowered metacognitive load

Material & methods

Participants

35 adult volunteers were recruited from the Laboratoire d'Economie Expérimentale de Paris (LEEP) pool of participants (M ± SD = 24.5 ± 3.06 years old, 18 females). They all provided informed written consent prior to the experiment. One observer was discarded for not finishing the experimental session, and 6 participants were removed because of extremely small accuracy rate for target 1 or 2 (exclusion criterion: <10% accuracy), leaving 29 participants for analysis. Observers were paid a base sum (10 EUR) plus a bonus depending on their performance in the task (up to 10 EUR in addition). The average payoff was 14.89 EUR (SD = 2.09) for a single 1.5 hours session. The experimental procedure received approval from the Paris School of Economics (PSE) ethics review board and adhered to the principles of the Declaration of Helsinki.

Apparatus and stimuli

Identical apparatus, stimuli and parameters were used for both experiments. The only difference being that for Experiment 2, confidence judgments was required only for T1 on half of the 500 trials, and only for T2 on the other half. Participants were divided into two groups to control for possible order effects. Participants were left uninformed that they will have to estimate their confidence for the other target until the end of the first half of the experiment.

Analysis

For the following analyses, trials were grouped by confidence probe: one group of trials for T1 confidence (250 trials per participant) and one group of trials for T2 confidence (250 trials). Therefore, even when accuracy only was considered, the average concerns the subset of trials related to the target where confidence judgment was requested.

RESULTS

T1: DISTRIBUTION OF REPORTS

The results from Experiment 1 were successfully replicated, with a significant effect of lag on accuracy ($F(3.4, 5.134)=9.1$, $MSE=0.005$, $p<0.001$) and confidence ($F(1.65,46.17)=17.5$, $MSE=0.08$, $p<0.001$). We found that letters presented just before or just after the target were reported on 19% of the trials (18% in Exp. 1), which exceeded the guess rate of 1/26 that is about 4% (mean corrected for guess rate: 0.15, 95% CI=[0.13 0.17]; $t(28)=14.7$ $p<0.001$).

To quantify how report frequency depended on serial position, we focused on serial positions from 2 items before to 2 items after T1 (included) and tested how report frequency can be predicted from the lag, the position and their interaction as fixed effects. Including item position as a predictor outperformed a model without the position effect ($\chi^2(4)=565.0$, $p_{RAND}=0.002$). Including the interaction between lag and position did not improved the model over a model without the interaction ($\chi^2(16)=21.6$, $p_{RAND}=0.36$), contrary to Exp. 1.

T1: PROBABILITY OF REPORT AND CONFIDENCE ARE CORRELATED

Similarly to Exp. 1, confidence was affected by item position ($\chi^2(4)=94.03$, $p_{RAND}=0.003$). Including the interaction between lag and position however did not improve the model ($\chi^2(16)=26.0$, $p_{RAND}=0.16$). Given that for T1 data, both report frequency (Fig. S2A) and confidence (Fig. S2B) were affected by position in similar manners, we directly evaluated the correlation between confidence and report frequency. To do so, for each participant we averaged confidence over lags, and correlated this average confidence to the report frequency across 5 report positions centered on target (including target's true position). The mean r coefficient was 0.71 across participants (95% CI=[0.59 0.83]; $t(27)=11.9$, $p_{RAND}<0.001$), replicating Exp. 1.

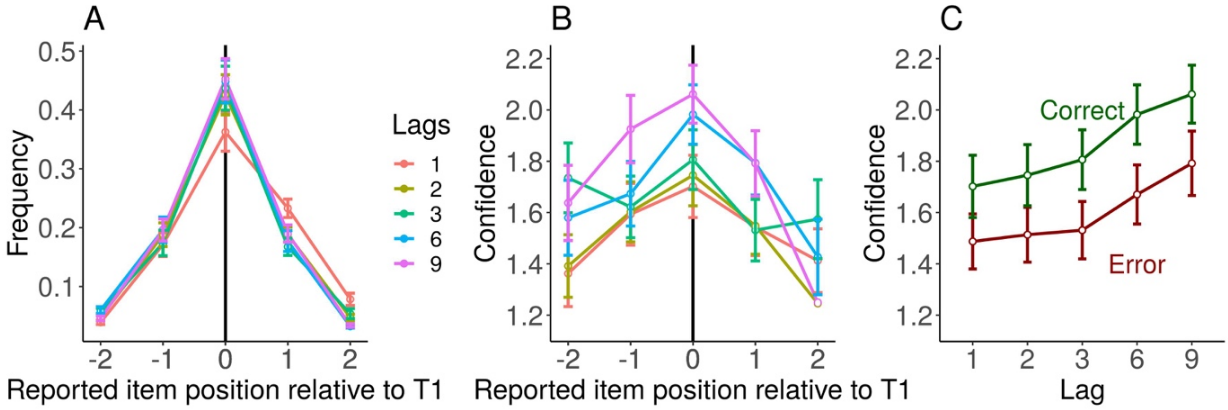

Figure S2. Reports and confidence about T1. (A) The frequency of report for item around target true position. (B) The corresponding average confidence per position. (C) The average confidence level for correct responses and errors, which provides an estimate of metacognition. Error bars represent standard error of the mean across participants.

Overall, T1 targets in Exp. 2 – as for Exp. 1 – were identified correctly 43% of the time. A main effect of trial type (error versus correct trial) was found ($F(1,28)=39.0$, $MSE=0.13$, $p<0.001$) but no interaction between lag and trial type ($F(3.598,100.737)=1.2$, $MSE=0.02$, $p=0.31$), confirming that participants had stable error-based metacognition for T1. Participants therefore gave higher confidence to correct than to incorrect T1 responses.

T2: Confidence tracks the Attentional Blink but not lag-1 sparing

Overall, 22% of T2 reports were correct when T1 was correctly reported. Figure S3A shows T2 accuracy and confidence for the different T1-T2 lags. As expected, the accuracy of T2 reports (i.e., in green) was affected by the lag between T1 and T2 ($F(2.9,81.13)=41.4$, $MSE=0.03$, $p<0.001$). In particular, the drop for lag 2 and lag 3 relative to longer lags (2-3 vs. 6-9: $T(28)=7$, $p<0.001$) indicated a classical Attentional Blink effect. Confidence was also affected by lag ($F(2.59,72.42) = 37.2$, $MSE=0.10$, $p<0.001$) and dropped for lags 2-3 relative to longer lags (2-3 vs. 6-9: $T(28)=0$, $p<0.001$), paralleling the drop observed for accuracy. Thus, participants seem able to acknowledge the drop of performance during the Attentional Blink that occurs at lags 2-3, in a similar manner as for Exp. 1.

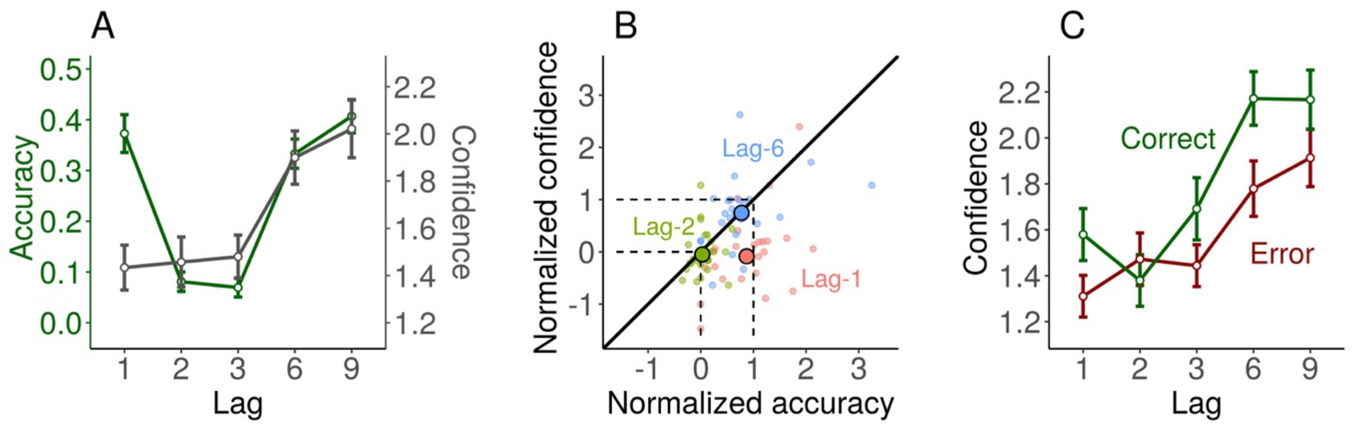

Figure S3. Attentional Blink and early confidence bias under lowered metacognitive load. (A) T2 average accuracy (in green) and confidence (in grey) as a function of the lag between T1 and T2. (B) The systematic under-confidence occurring at Lag-1 (see Fig. 3B) was also found in Experiment 2. Each point is a participant. (C) The average confidence level for correct T2 reports and errors, for each lag.

Participants' confidence, similar to Exp. 1, seemed blind to lag-1 sparing. Indeed, the lag-1 sparing effect was also found in our data: T2 accuracy was spared when T2 was presented immediately after T1. Accuracy at lag-1 was much higher than during the blink period (1 vs. 2-3: T(28)=378, p<0.001) and was in fact indistinguishable from accuracy at long lags (1 vs. 6-9: T(28)=238, p=0.67). By contrast, confidence was as low at lag-1 as it was for lag 2-3 (T(28)=160, p=0.70) and much lower than confidence at long lags (1 vs. 6-9: T(28)=9, p<0.001). All these results were fully coherent with what was found in Exp. 1.

Figure S3B shows confidence and accuracy at lag-1, in the lag-3-to-9 space, where lag-3 and lag-9 have (0,0) and (1,1) coordinates, respectively. Most participants are located below the diagonal, suggesting that they are less confident at lag-1 than what would be expected given their accuracy level at lag-1. This lag-1 under-confidence, calculated as the average difference between predicted and observed lag-1 confidence, was significant at the group level (T(28)=325, p<0.001, alpha=0.05/3). To confirm that this linear approach could nonetheless be used to predict confidence at another lag, we applied the same analysis to lag-2 and lag-6. The difference was significant neither for lag-2 (t(28)=248, p=0.7, alpha=0.05/3) nor for lag-6 ($t(28)=0.13$, $p=0.9$, alpha=0.05/3). These results suggest that probing confidence only for T2 did not alter the pattern found in Experiment 1.

T2: Confidence in correct responses vs. errors

Because some participants had no correct answers during the Attentional Blink, only half of participants were considered here (N=14). As can be seen from Figure S3C, participants overall expressed higher confidence when they were correct relative to their errors, with a main effect of trial type (error vs correct, $F(1,13)=16.5$, MSE=0.10, $p=0.001$) and a main effect of lag ($F(3.1,40.7)= 21.8$ MSE=0.16, $p<0.001$), but no interaction ($F(1.9,24.1)=1.5$, MSE=0.24, $p=0.2$). This difference between Exp 1 and Exp 2 might relate to the difference in samples (250 vs 500) and the low number of participants in the present analysis (N=14).

T2: Probability of report and confidence are correlated

The similarity between confidence and report frequency was tested by looking at their correlation across lags for 5 positions centered on T2, but contrary to T1, the correlation was not reaching significance (Mean r coefficient: 0.55, 95% CI=[0.38 0.73]; $t(28)=6.5$, $p_{RAND}=0.06$), as shown on Figure S4D. Figure S4C plots the regression on one representative participant for illustrative purpose. The smaller correlation found in Exp. 2 compared to Exp. 1 might be the result of the reduced number of samples (half of Exp.1 samples for T2 confidence).

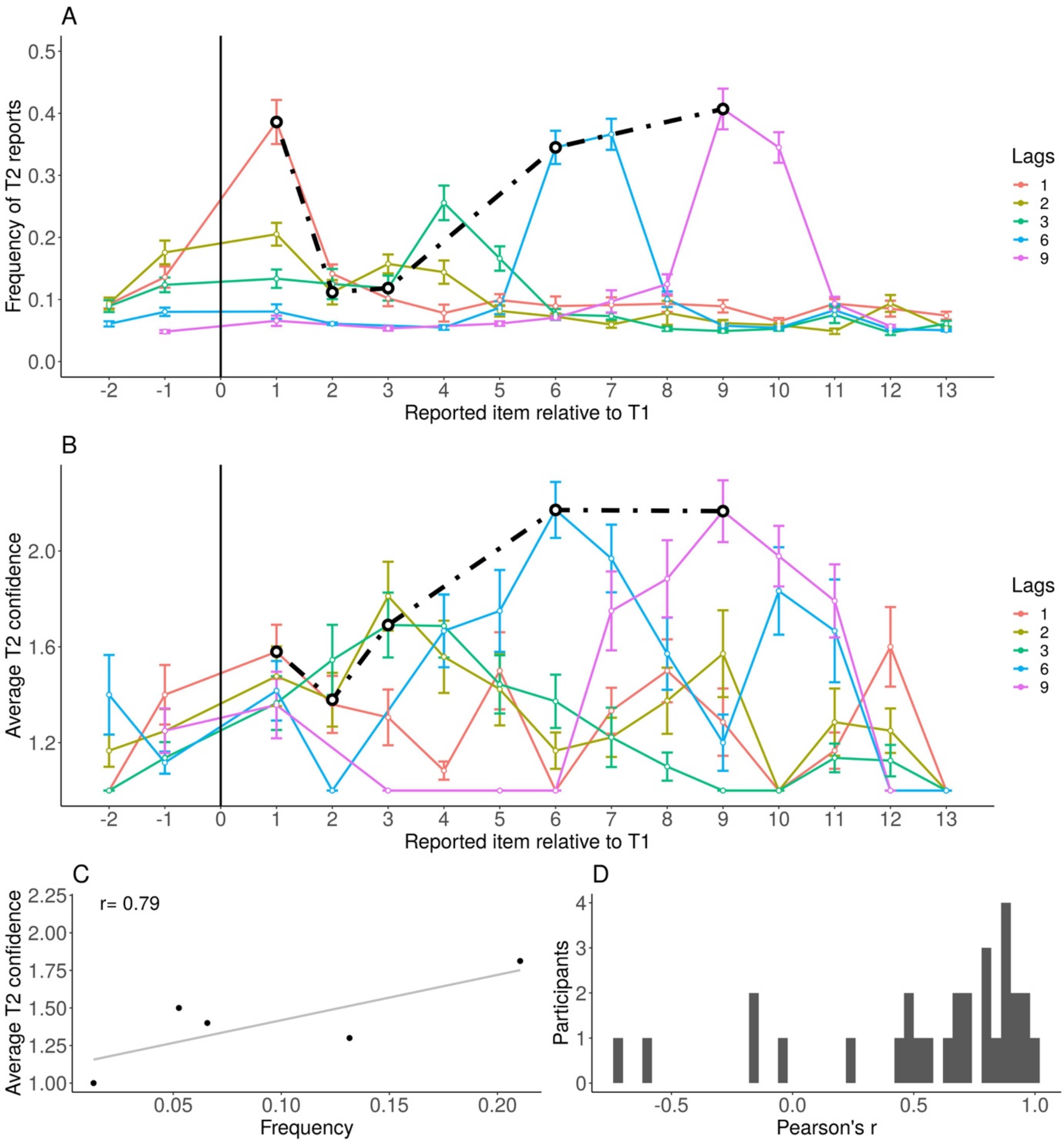

Figure S4. Reports and confidence about T2. (A) The frequency of T2 reports as a function of the position of the reported item relative to T1, for each lag. Note that T1 position has no value, given that only trials in which T1 is correctly reported were considered here (hence T2 reports cannot correspond to T1 position). The black line connects the points corresponding to accurate T2 reports. (B) Confidence of the T2 reports, as a function of the position of the reported item relative to T1, for each lag. The black line connects the points corresponding to accurate T2 reports. Error bars represent standard error of the mean across participants. (C) Regression between frequency and confidence with 5 positions centered on T2, collapsed across lags, for a representative participant. (D) Histogram of the correlation coefficients for all the participants.

T2: Delay in temporal selection and confidence

As for Exp. 1, items appearing just before or just after the T2 were more likely to be reported than chance (17%, with a 95% CI=[0.15 0.18]; vs. chance level at 4%: $t(28)=19.0$, $p<0.001$). Hence, errors were not random guesses but samples that are close to the actual T2 target. A model comparison approach confirmed that including the lag as a predictor for the center of mass significantly outperformed the null model for T2 ($\chi^2(4)=18.4$, $p_{RAND}<0.001$). Replicating Exp. 1, selection appears to be systematically too late for lags 6 and 9 (Fig. S5A). Bonferroni-corrected t-tests (alpha=0.05/5) confirmed an effect of lag on the center of mass for lag-6 ($t(28)=5$, $p<0.001$) and lag-9 ($t(28)=5.2$, $p<0.001$), but not for lag 1, 2 and 3 (all $p>0.15$).

For T1, the effect of lag on the center of mass was also significant ($\chi^2(4)=24.4$, $p_{RAND}<0.001$), but Bonferroni-corrected t-tests (alpha=0.05/5) confirmed that it was specifically driven by lag-1 ($t(28)=3.2$, $p<0.001$), but not by other lags (all $p>0.3$). This lag-1 effect on T1 selection delay could be resulting from order reversals (see below).

To analyze confidence, a model comparison approach confirmed that including the pre/post-target factor (or “shift”) as a predictor for average confidence significantly outperformed the null model ($\chi^2(1)=18.3$, $p_{RAND}<0.001$). The interaction between lag and shift was, however, not significant ($\chi^2(4)=6.6$, $p_{RAND}=0.08$). In other words, confidence is oblivious to the delays induced by the Attentional Blink and biased towards items selected later. A reduced metacognitive load in Exp. 2 did not enhance delay introspection (Fig. S5B). For comparison, we found no effect of shift on confidence for T1 ($\chi^2(1)=0.3$, $p_{RAND}=0.5$).

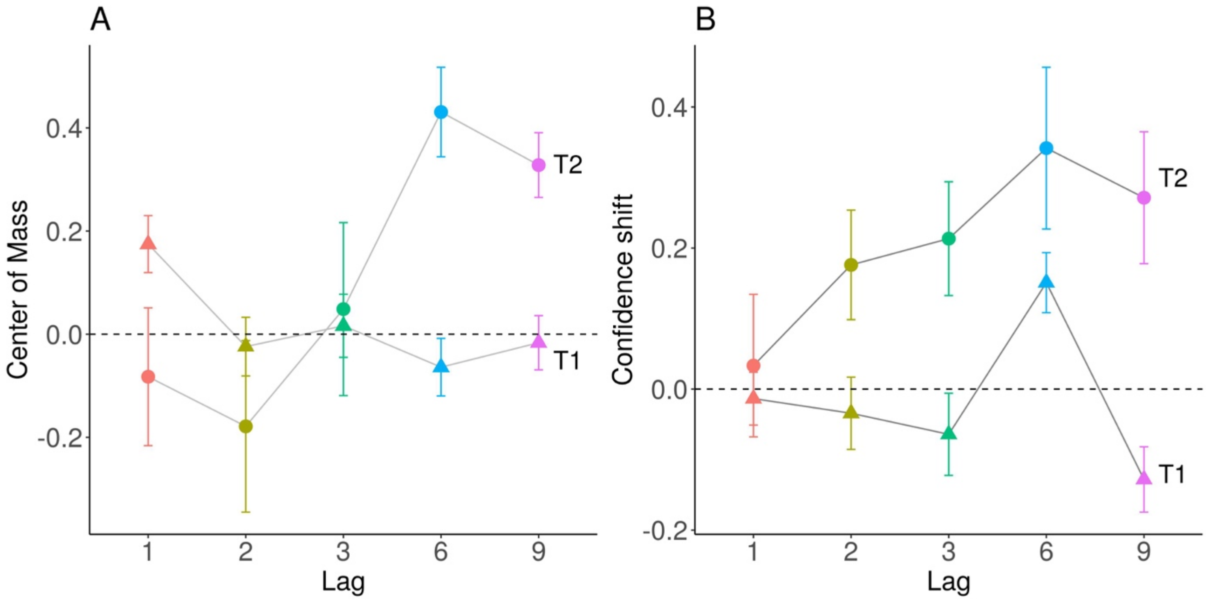

Figure S5. Confidence does not correct for attentional delay. (A) The average center of mass for T1 (rectangles) and T2 (dots) as a function of lag. Note the delay in T2 selection following lag-3. (B) The confidence shift, which is the difference in average confidence between post-target and pre-target errors for T1 (triangles) and T2 (dots). A positive value corresponds to higher confidence for post-target errors, that is, a shift of the confidence peak towards more delayed items. Error bars represent standard error of the mean across participants.

ORDER REVERSAL BETWEEN T1 AND T2

In Exp.2, order reversals occurred on average in 5.2% (SE ± 4.1%) of lag-1 trials. For comparison, correct report of both T1 and T2 in the correct order occurred in 13.2% (SE ± 6.3%) of lag-1 trials. To evaluate whether participants were aware of such reversals, the confidence between trials in which both T1 and T2 were correctly reported was compared to the confidence in reversed trials. Seven participants were discarded from the later analysis due to no order reversal trial for the T1 confidence block, and two participants were discarded for the T2 confidence block. No difference in confidence was found between these two types of trials for T1 ($t(21)=0.4$, $p=0.70$), or for T2 ($t(26)=2.2$, $p=0.04$), after Bonferroni correction (alpha=0.05/2 for testing T1 and T2). Thus, it seems that when metacognitive load is reduced, participants were not more able to notice the occurrence or non-occurrence of a reversal on a trial-by-trial basis.

Descriptive model for attention and confidence

Here, we propose a simple implementation of a single target selection model inspired by the Attentional Gating Model (Reeves & Sperling, 1986), that could produce the relation between confidence and report frequency found in our data. The model has 3 components: a sensory stage, an attentional modulation, and a decision stage.

The sensory stage consists in a set of letter detectors or channels. Each channel has a preferred letter and when this letter is presented on the screen the channel is activated for a short period of time (Eq. 1). The activity s_c of each channel *c* at the sensory stage is defined as a Gaussian function of time t, with parameters μ representing the time at which the letter is presented, and τ the duration of the channel's response.

$$s_c(t) = \frac{1}{\tau\sqrt{2\pi}}\, e^{-\frac{1}{2}\left(\frac{t-\mu}{\tau}\right)^2} \quad \text{(Eq. 1)}$$

When a cue is presented on the screen, an attentional modulation is triggered that will amplify the activity of all channels for a brief period of time. The attentional modulation a(t) involves a strength parameter A, and follows a Gaussian function of time (Eq. 2), with parameters μ_A and τ_A representing the center and spread in time of the attentional window. Note that the attentional modulation can be suppressed (e.g., for T2 at lag 3), which will be captured by the strength parameter A being reduced. This attentional modulation can also be delayed relative to the true position of the cue, which will be represented by the parameter μ_A.

$$a(t) = A\, \frac{1}{\tau_A\sqrt{2\pi}}\, e^{-\frac{1}{2}\left(\frac{t-\mu_A}{\tau_A}\right)^2} \quad \text{(Eq. 2)}$$

At the end of the trial, the resulting activity of the channel c (noted y_c) is the cumulated response of channel c over time, corrupted by normally distributed noise with standard deviation σ (Eq. 3).

$$y_c = \int_0^T s_c(t)\, a(t) dt + \varepsilon, \varepsilon \sim N(0, \sigma) \quad \text{(Eq. 3)}$$

Finally, the response corresponds to the letter associated with the channel with maximal activity, and the confidence associated with this response corresponds to the activity of this channel.

$$response = argmax_c(y_c)$$

$$confidence = max_c(y_c)$$

We simulated this process independently for T1 and for T2 at different lags. The duration of the sensory response and attentional boost, and the noise at the decision stage were kept constant across simulations ($\tau = 60$, $\tau_A = 80$, $\sigma=0.001$). The values for A and μ_A were defined separately for T1 ($A = .95$ and $\mu_A = 0$) and for T2 at the different lags (see Fig. S6 D and E), in order to roughly reproduce our behavioral results. For comparison with our actual data, the simulated confidence was binned into 3 values across all lags, separately for T1 and T2.

The R script for the model can be found on OSF: https://osf.io/xjh2v

Applying our analyses to these simulated data (see Fig. S7 – S10), we found that the model qualitatively produces the correlation between confidence and report frequency across positions (Fig S9), as anticipated. Unsurprisingly, this model was also able to reproduce the associated observations that confidence judgments for T2 are blind to delays in response selection (Fig. S10), and that they are higher for correct responses than for errors for T1 (Fig. S10) and for T2 (Fig.

S10). We found also that as in our real data, the simulated T2 confidence was higher at longer lags (Fig. S9), although this presumably reflects the choice of parameter values across lags and should not be taken as a key aspect of our model. It is also clear that this simple model does not reproduce one main result of our study, which is the under-confidence found at lag-1 for T2. We anticipated that this model would not show such under-confidence at lag-1, as it implements a strong link between confidence and accuracy, and no factor that would affect lag-1 specifically. This result might require an additional component to the model.

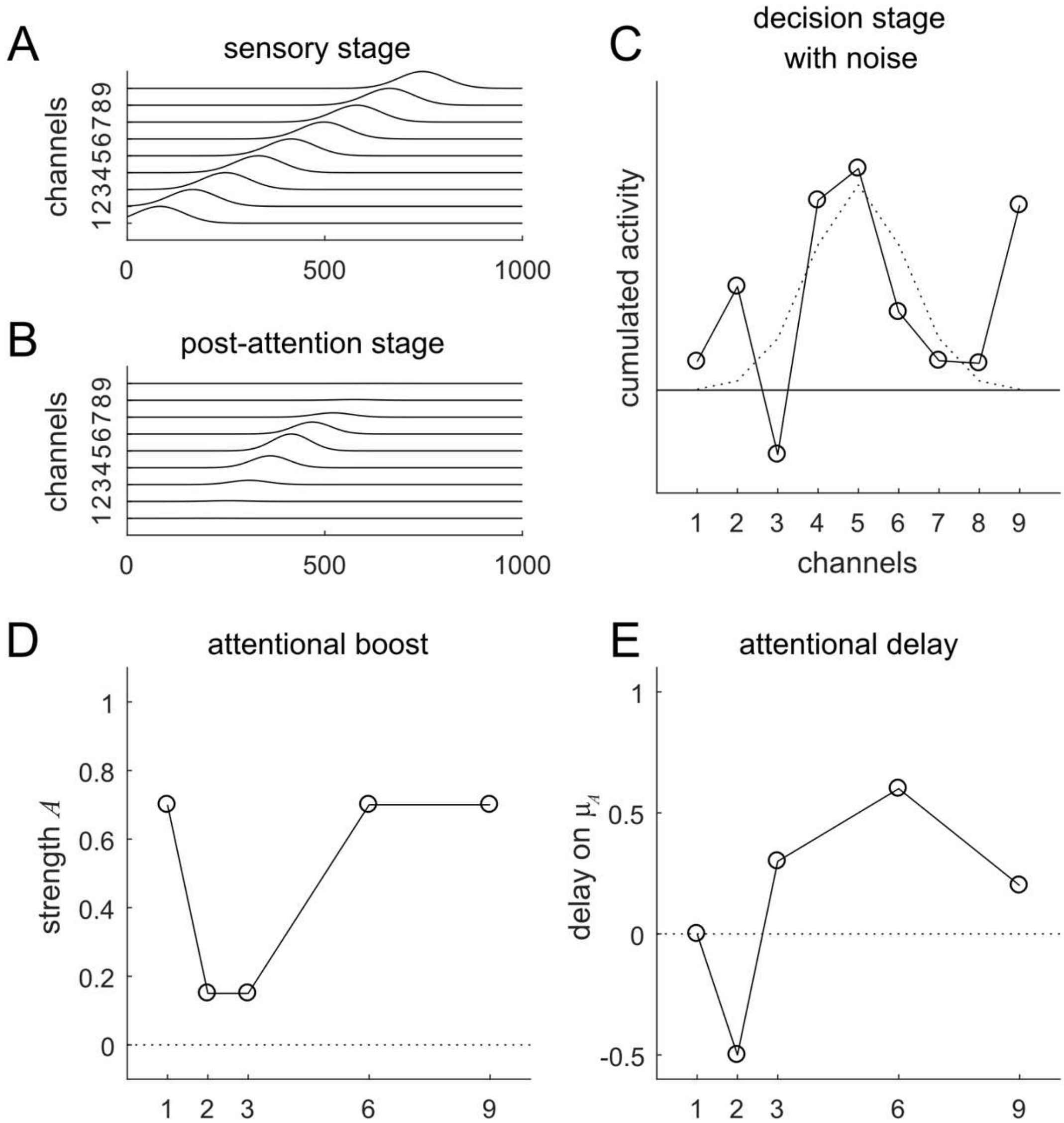

Figure S6. Illustration of the descriptive model. (A) Time-course of activity in channels at the sensory stage. Only 9 channels are represented, which corresponded to the first 9 letters in the stream. (B) Time-course of activity in channels after attentional modulation triggered at the 5th letter in the stream. (C) At the end of each trial, the activity in each channel (dotted line) is summed over time, and corrupted with additive noise (solid line). The identity of the best-responding channel, here channel 6, on a trial gives the response for that trial, and its activity gives the confidence. (D) The profile of attentional modulation across lags, used for our simulations of T2. (E) The profile of delay in attentional modulation across lags, used for our simulations of T2.

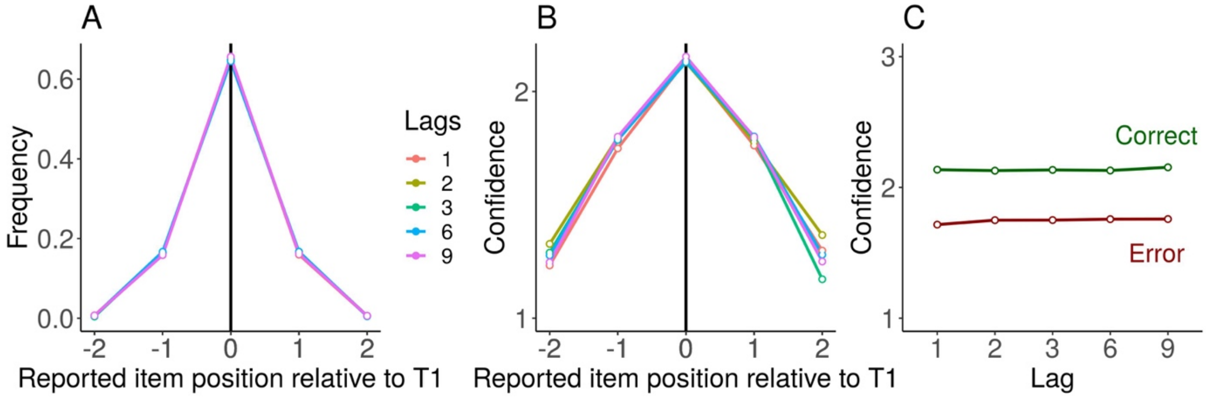

Figure S7. simulated reports and confidence about T1. (A) The frequency of report for item around target true position. (B) The corresponding average confidence per position. (C) The average confidence for correct and error trials.

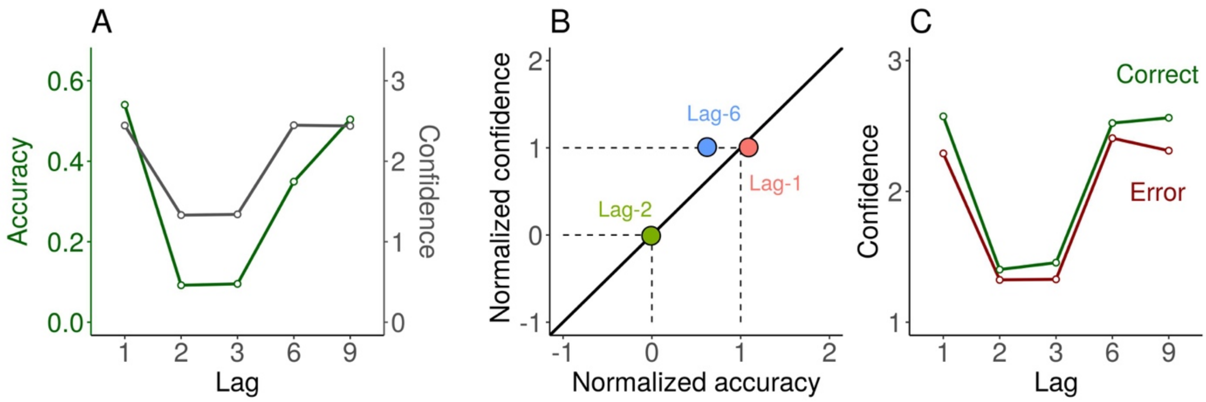

Figure S8. Attentional Blink but no early confidence bias for our simulated data. (A) T2 average accuracy (in green) and confidence (in grey) as a function of the lag between T1 and T2. (B) Simulated confidence and accuracy, normalized to the lag-3 to lag-9 interval. Note that the model does not produce the under-confidence at lag-1. (C) The average confidence level for correct responses and errors, for the different lags.

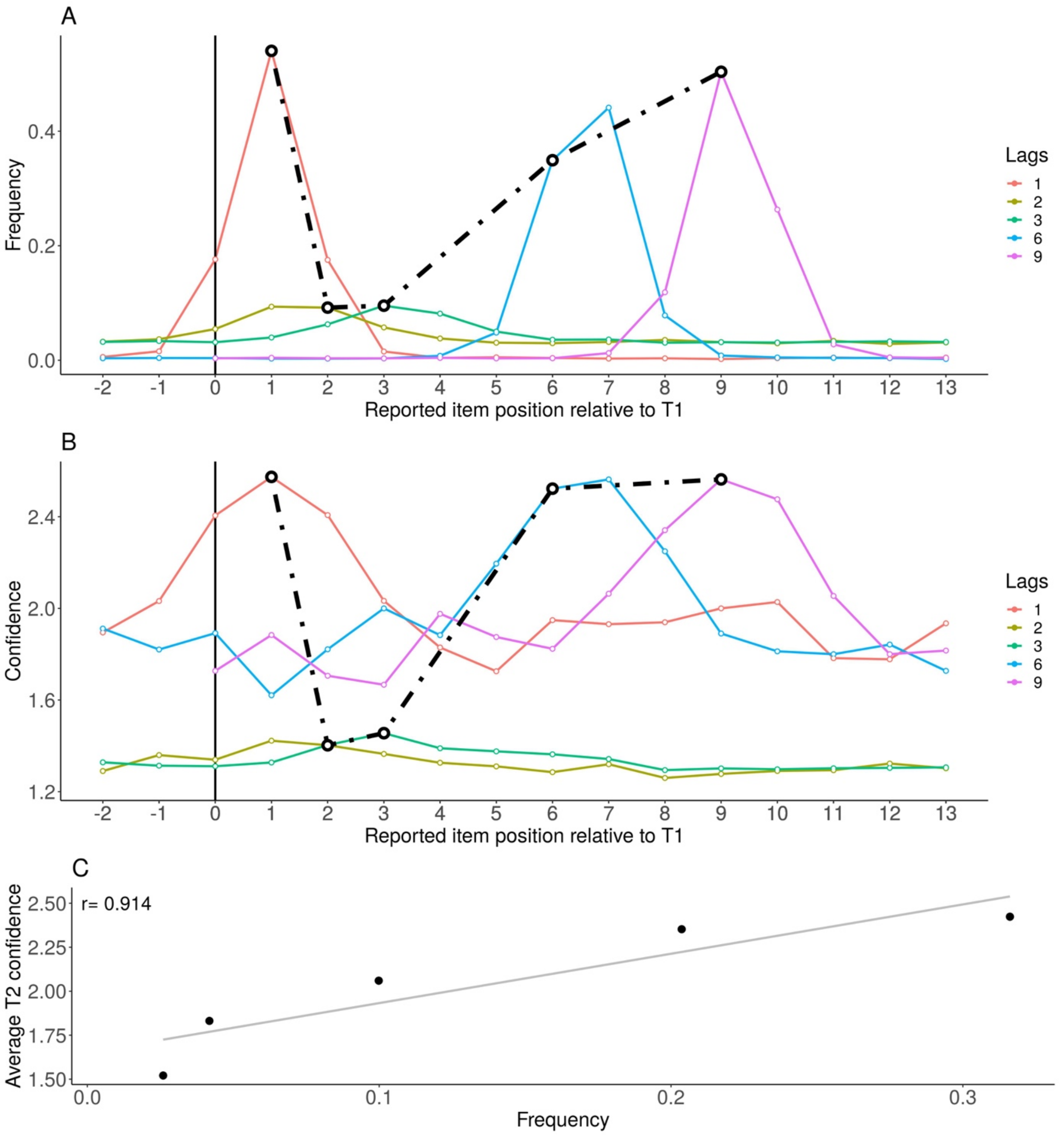

Figure S9. Simulated reports and confidence about T2. (A) The frequency of simulated T2 reports as a function of the position of the reported item relative to T1, for each lag. Note that T1 position has no value, given that only trials in which T1 is correctly reported were considered here (hence T2 reports cannot correspond to T1 position). The black line connects the points corresponding to accurate T2 reports. (B) Confidence of the simulated T2 reports, as a function of the position of the reported item relative to T1, for each lag. The black line connects the points corresponding to accurate T2 reports. (C) Regression between frequency and confidence with 5 positions centered on T2, collapsed across lags, for our simulation.

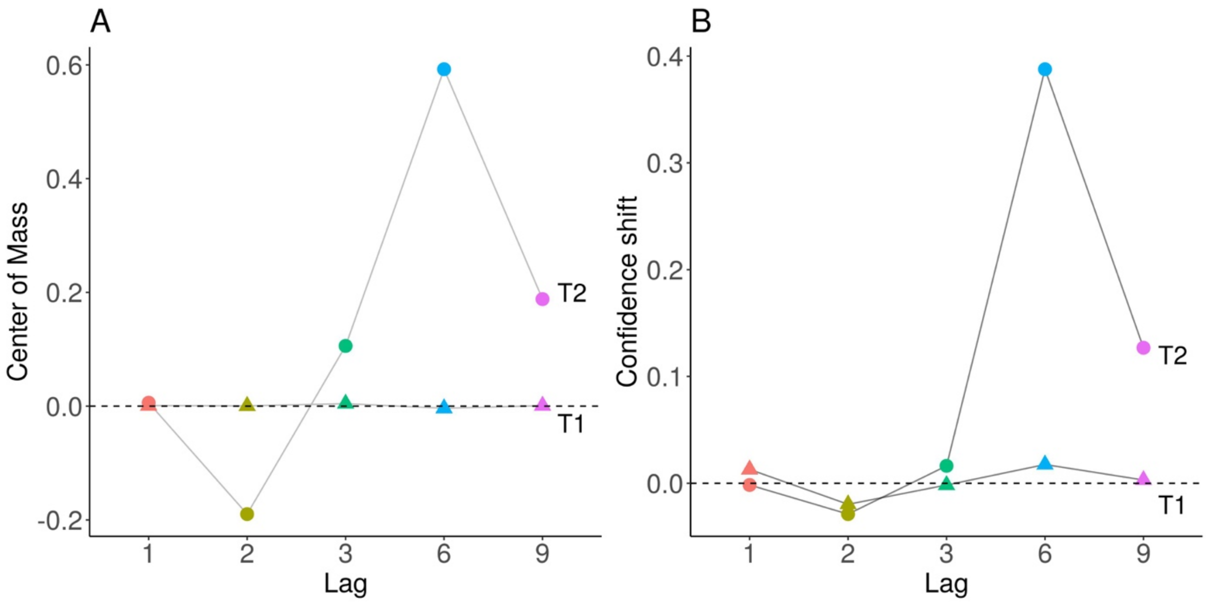

Figure S10. Simulated confidence does not correct for attentional delay. (A) The average center of mass for T1 (rectangles) and T2 (dots) as a function of lag, in our simulated data. Note the delay in selection following lag-3 for T2. (B) The confidence shift, which is the difference in average confidence between post-target and pre-target errors for T1 (triangles) and T2 (dots). A positive value corresponds to higher confidence for post-target errors, that is, a shift of the confidence peak towards more delayed items.

References

Goodbourn, P. T., Martini, P., Barnett-Cowan, M., Harris, I. M., Livesey, E. J., & Holcombe, A. O. (2016). Reconsidering Temporal Selection in the Attentional Blink. Psychological Science, 27(8), 1146–1156. https://doi.org/10.1177/0956797616654131

Martini, P. (2012). Sources of bias and uncertainty in a visual temporal individuation task. Attention, Perception, and Psychophysics, 75(1), 168–181. https://doi.org/10.3758/s13414-012-0384-y

Reeves, A., & Sperling, G. (1986). Attention gating in short-term visual memory. Psychological Review. https://doi.org/10.1037/0033-295X.93.2.180

Vul, E., Hanus, D., & Kanwisher, N. (2008). Delay of selective attention during the attentional blink. Vision Research, 48(18), 1902–1909. https://doi.org/10.1016/j.visres.2008.06.009

Vul, E., Nieuwenstein, M., & Kanwisher, N. (2008). Temporal Selection is Suppressed, Delayed, and Diffused During the Attentional Blink. Psychological Science, 19(1), 55–61. https://doi.org/10.1111/j.1467-9280.2008.02046.x

Chapter 4 | Orienting spatial attention weakens metacognition

In the previous chapters, we observed that confidence was able to detect changes in accuracy when attention was oriented to the right moment in time (Chapter 3, first target) or when it progressively disengaged from the right location (Chapter 2). In contrast, when attention was oriented to the wrong moment in time (Chapter 3, second target), confidence ceased to purely reflect accuracy. On the contrary, it continued to trust attention as a reliable provider of evidence, a pattern responsible for a drop in metacognitive ability. If there is this much dependency between metacognitive ability and attention, what about the mechanism drawing attention to a given location, what about the orienting process itself? In this last chapter, we adapted a 'Wundt clock' paradigm to investigate the effect of the trial-by-trial variability in attentional orienting on confidence. Wundt described his original paradigm aptly as follows: "Let, e.g., an index-hand move over a circular scale with uniform and sufficiently slow velocity, so that the impressions it gives will not fuse, but permit its position at any instant to be distinctly seen. Let the clockwork which turns it have an arrangement which rings a bell once in every revolution, but at a point which can be varied, so that the observer need never know in advance just when the bell-stroke takes place. (…) The bell-stroke can be perceived either exactly at the moment to which the index points when it sounds—in this case there will be no time-displacement; or we can combine it with a later position of the index—(…) positive time-displacement, as we shall call it (…)” (cited in James, 1887, p. 415). In the present paradigm, we simply replaced the bell sound by a visual transient, and we capitalised on the effect attention has on what Wundt refers to as ‘positive time-displacement’.

Orienting spatial attention weakens metacognition

Samuel Recht (1)*, Pascal Mamassian (1)+, Vincent de Gardelle (2)+

(1) Laboratoire des systèmes perceptifs, Département d'études cognitives, École normale supérieure, PSL University, CNRS, 75005 Paris, France

(2) CNRS and Paris School of Economics, Paris, France

* Corresponding author: samuel.recht@gmail.com

+ these authors share senior authorship

Abstract

How does orienting of attention in space affects a person's ability to evaluate her performance? Previous work has considered cases in which spatial attention is already fully deployed, but less is known about metacognition during attentional deployment. Here, we investigate how the timing of attention affects metacognitive ability. To probe both exogenous and endogenous visuo-spatial attention, we adapted a "Wundt clocks" paradigm. This design builds on the robust finding that attention has been shown to alter the latency between objective and perceived events (i.e., "flash-lag" effect). Participants looked at 6 clocks at a fixed eccentricity rotating at a fixed speed but at different phases. At a random time, one of the clocks was either cued peripherally (exogenous) or centrally (endogenous), and when the clocks stopped, participants were requested to report the hand position at cue onset. The moment of attentional orienting was manipulated using a "pre-cue" condition, such that attention could either be deployed at the cued location or still be undeployed. Each two trials, participants chose the one they felt more confident to be correct. The average reported times were delayed in accordance with exogenous/endogenous attention. Surprisingly, confidence was not correlated to these attention-induced delays. However, confidence judgments correlated with the relative error between each trial in the pair, suggesting that participants were able to estimate their internal deviation at the trial level. Importantly, endogenous orienting of attention reduced this confidence-error relation compared to the pre-cue condition. A control experiment confirmed that this metacognitive cost could not be reduced to pure sensorimotor uncertainty. These results demonstrate that the very process of orienting attention in space weakens metacognitive ability.

Introduction

Visual confidence is the subjective reliability of a preceding decision in the visual domain; broadly, it is the self-evaluation of performance (Mamassian, 2016). Confidence is an important second order judgement that allows to objectively evaluate the quality of a first-order judgment: a confidence estimate can be directly matched to first-order accuracy. One's ability to reliably track self-performance through confidence judgments has been coined 'metacognitive ability', or simply metacognition (Mamassian, 2016). Metacognition plays a critical role in adaptive learning (Guggenmos et al., 2016; Hainguerlot et al., 2018; Zylberberg, Wolpert, & Shadlen, 2018), information seeking (Boldt, Blundell, & De Martino, 2019; Desender et al., 2018), and the integration of multiple decision stages (van den Berg, Zylberberg, Kiani, Shadlen, & Wolpert, 2016). In general, performance variability is well tracked by confidence judgments, but dissociations have been documented in the literature (for a review, see Fleming & Daw, 2017). These, so called, 'limits' of metacognitive ability, have important ramification for the understanding of perceptual decision-making.

One way to tackle these limits is to root the investigation in an important and reliable predictor of performance, and then investigate the effects of this mediator on changes in accuracy. Given the limitations of cognitive resources and the vast amount of peripheral sensory information, visual selective attention is thought to play the critical role of a filter: it selects, prioritizes and amplifies some specific sensory information for further processing (Buschman & Kastner, 2015; Carrasco, 2011). Such a selective mechanism provides an organism with an important tool to optimize resource allocation in space and in time. Ideally, an individual would use some sort of metacognition to evaluate the state of the attentional system, in order to make online adjustments. Specifically, Spatial attention would enhance a stimulus at a particular point in space, and inhibit other locations in the visual field (Carrasco, 2011). The relationship between spatial attention and metacognition, however, is less clear. Some studies have found a dissociation between the two (D. Rahnev et al., 2011; Wilimzig et al., 2008), while others show a tight positive association between metacognition and the effects of spatial attention (Denison et al., 2018; Zizlsperger et al., 2012, 2014). When attention is oriented to a point in time rather than to a point in space (Coull & Nobre, 1998), we found in a recent study a robust under-confidence bias, whereby visual confidence neglected that a single attentional episode can benefit to multiple targets (Recht et al. 2019). Interestingly, confidence was also oblivious to the latency of temporal attention, suggesting that metacognition does not perfectly track the limits of the attentional system in the temporal domain. The same may well hold for the temporal dynamics of spatial attention. However, most of these studies on spatial attention considered how perceptual and metacognitive judgments are varying from attended to unattended locations, but did not investigated how the temporal dynamics of spatial attention might affect confidence.

When accuracy and confidence are evaluated after a valid or invalid pre-cue, the metric assesses whether the benefit in accuracy is accompanied by a change in confidence. There is another aspect of attentional deployment that has been overlooked, that is, whether observers can evaluate the time it takes for spatial attention to be deployed.

The temporal structure of spatial attention is usually considered through the lens of its processing types. Classical taxonomy in the literature differentiates exogenous from endogenous attention. Exogenous stands for an involuntary, early and short-lasting orienting of attention, while endogenous relates to a voluntary, late and long-lasting allocation (Carrasco, 2011). The nature of an attentional episode is therefore defined primarily by the time it takes to emerge, with exogenous attention taking roughly 100ms to be effective while approximately 300ms are necessary for endogenous attention to be allocated. Therefore, time is an essential element of attention, and yet little is known about how the fluctuations of attentional timing affect confidence and metacognition.

Here, we adapted a "Wundt clocks" paradigm where participants have to reproduce the phase of a clock at probe onset. Crucially, this continuous report is known to be affected by attention, and has been considered to be a proxy for attentional timing (Carlson, Hogendoorn, & Verstraten, 2006; Chakravarthi & VanRullen, 2011; Hogendoorn, Carlson, VanRullen, & Verstraten, 2010). By anchoring stimulus features to attentional timing, this design enabled us to record a signature of the temporal fluctuation of spatial attention, and to study its effect on confidence judgments. We did this by asking participants to (indirectly) estimate how sensory processing time was affected by attention in a perceptual task. To ensure that this process could not be explained by metacognition of sensorimotor uncertainty, we compared these results to a simple detection task in which participants had to estimate their own response times. Our study revealed three major findings. First, visual confidence ignored the latency of both exogenous and endogenous attention. Secondly, metacognition was specifically altered during but not after endogenous orienting of attention to a particular location. Finally, metacognitive ability in the main task was not correlated to metacognition of response times, suggesting that metacognition of temporal variability in the first task cannot be reduced to metacognition of sensorimotor uncertainty.

Material & methods

Participants

20 adult volunteers were recruited from the French RISC pool of participants (age M ± SD = 25.85 ± 2.30, 14 females). They all provided informed written consent prior to the experiment. Participants were compensated for their time at a rate of 10€ per hour. The experiment consisted in two 1-hour sessions. The experimental procedure was approved by the ethics review board of the Paris School of Economics (PSE).

Apparatus & stimuli

Main task

Participants sat 60cm from the screen (11 in front of a 1280x1024 pixels CRT monitor, 85Hz refresh rate, and 9 in front of a 1920x1080 pixels monitor, 60Hz refresh rate) with their head maintained with a chin-rest. Stimuli were generated using the Python programming language and the PsychoPy library (J. W. Peirce, 2007b), on a Mac OS computer. On each trial, participants were presented with a fixation dot on a grey background for 1000ms. After this delay, six clocks (black outline, inner/outer diameter: 4.1°/3.9°) were presented on the right and left side of the fixation dot (0.4° rectangle) for a total duration of 4 seconds. Four clocks were displayed at a 4° eccentricity from the center of the screen on its two diagonals, and two clocks were displayed at 6° eccentricity from the center on the horizontal midline. The center of each clock consisted in a black dot (diameter: 0.2°). The hand was made of a line starting 0.2° from the clock's center (length: 1.4°). They rotated at a fixed speed of 1 revolution per second. Hand positions at clocks onset were random for each clock and trial.

In the pre-cue condition, a green line (length: 1°) was displayed centrally for the whole trial duration, indicating with 100% validity the clock to attend (randomly assigned per trial). In the "exogenous" and "endogenous" conditions, no pre-cue was displayed.

At a random time (sampled from a uniform distribution in the range 1000-2000ms after clocks' onset) in the pre-cue and exogenous conditions, a clock was peripherally cued for approx 20ms with a red annulus surrounding the clock (inner/outer diameter: 4.2°/4°). In the "endogenous" condition, a central probe (black line, length: 1°) pointing towards the clock was presented for ~20ms. The moment of the central probe onset was sampled from the same uniform distribution as the two other conditions. The clocks offset occurred 4000ms after the initial onset for all conditions. The probability to be probed for a considered clock was at chance level (1/6).

Reaction times (RT) task

The stimuli were identical to the main task, but this task only included the "pre-cue" and "exogenous" conditions.

Procedure

For both tasks, participants were instructed to fixate the center of the screen during the whole trial period and their gaze was monitored online using an eye-tracker (EyeLink 1000 Hz, SR Research). To enforce fixation and prevent eyeblinks and saccadic shifts preceding cue onset, any trial during which participants blinked or move their gaze away from a 1° window centered on the fixation dot were automatically aborted, and a new sample of the trials pair was added at the end of the block. Fixation was enforced from 200ms before cue onset. Before each task, participants completed 10 practice trials for each condition. The order of the two tasks was counterbalanced for each session. The full experiment consisted in 432 trials for the main task and 288 trials for the RT task.

MAIN TASK

Each session consisted in 3 blocks of 72 trials. The order of the blocks was randomized. Participants were instructed to fixate the centre of the screen, to monitor all the clocks (exogenous/endogenous conditions) or only the pre-cued clock (pre-cue condition), and to register the phase of the relevant clock at cue onset. After each trial, participants were requested to use the mouse to indicate the phase of the considered clock at the cue's onset (Type-1 continuous response). Every two trials, participants were asked to select which of the two previous responses they felt more confident about by clicking on one of two rectangles (2°x2°) displayed at 6° eccentricity on each side below the fixation cross, flanked by “1” and “2” (for first or second trial in the pair). As such, for each pair of trials, one trial was labelled as “high confidence” and the other one as “low confidence”. Participants were not instructed to make speeded responses. At the end of each ~15min block, participants had the opportunity to take a break.

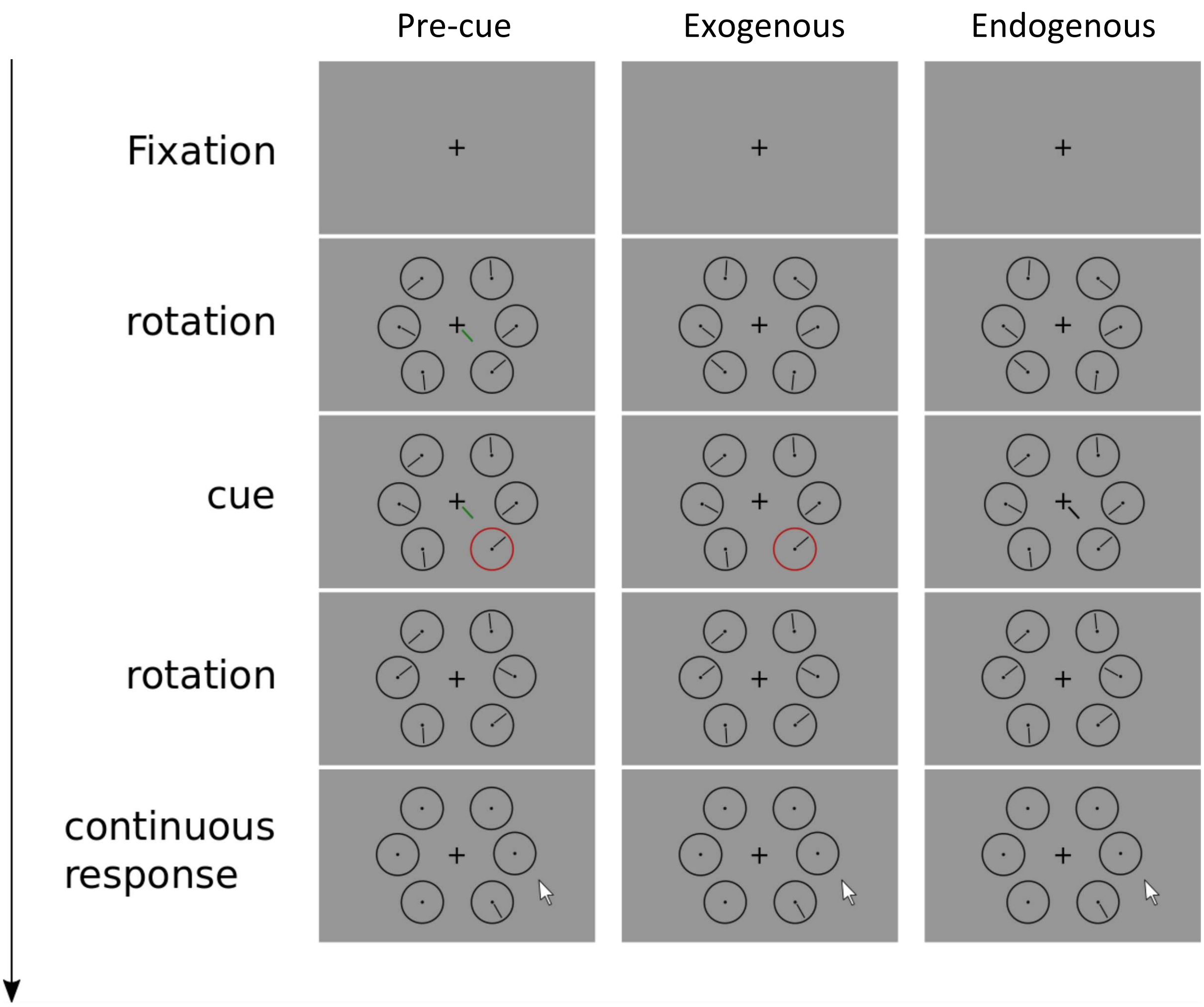

Figure 1. Experimental protocol. Main task: On each trial, the stimuli consisted in a grid of clocks rotating at a fixed speed but with random phases. After a variable delay, one of the clocks was cued, either peripherally ("exogenous" condition) or centrally ("endogenous" condition). A third ("pre-cue") condition included a central pre-cue through the whole trial, indicating with 100% validity the to-be-cued clock. After 4 seconds, a response prompt was displayed, and the participant had to reproduce the phase of the clock at cue onset. Every two trials, participant was requested to choose the best of the two preceding responses (confidence judgment).

RT TASK

Participants were requested to make a speeded response by pressing a key at cue onset. Every two trials, participants were asked to select which of the two previous responses had the lowest reaction time (confidence 2AFC). Each session consisted in one block of 72 trials for the pre-cue condition and one block of 72 trials for the exogenous condition. We did not test the endogenous condition in the RT task. The order of the blocks was randomized.

ANALYSES

For each trial, an estimation error (hereafter referred to as error) was calculated as the difference between objective hand's position at probe onset and the position reported by the participant. This angular error was then converted in milliseconds. Given the circular nature of the data, Von Mises distributions were fitted on angular values using maximum likelihood estimation (MLE) separately for each participant and condition (see Supplementary Material). The location parameter of the distribution (equivalent to the mean of a normal distribution) is an estimate for the latency of response errors, since it relates to the average time difference between the objective event and the perceived event. The concentration parameter (or "kappa", equivalent to $1/\sigma^2$ in a normal distribution) is an estimate for the precision of the responses, and is inversely proportional to the variance. We chose this model to account for our data. Yet, there are other possible models in the literature to account for subjective reports of continuous variables such as orientation or colour. Two prominent alternatives are a model including a guess rate (Zhang & Luck, 2008) and a model with variable precision (e.g., Van Den Berg, Shin, Chou, George, & Ma, 2012). The former involves a mixture of Von Mises and uniform distribution, in order to account for possible guesses in certain trials (Zhang & Luck, 2008). However, this model did not improve our fit, and the estimation of precision between conditions remained unaltered when opting for such model. The second, variable-precision model (Van Den Berg, Shin, Chou, George, & Ma, 2012), did not alter the pattern of estimated precision either (see Supplementary Material).

We also conducted a pupillary analysis, which is presented in the Supplementary Material.

When necessary, ANOVAs were corrected using the Greenhouse-Geisser adjustment and t-tests were corrected using the Welch-Satterthwaite adjustment. We report Wilcoxon signed rank test using uppercase T when the Shapiro-Wilk normality test failed, and Student test using lowercase t otherwise.

RESULTS

Main task

Spatial orienting of attention

To determine the role of the cue, we analysed the effect of condition on the latency and precision of the response distributions (fig. 2, see Analyses for details regarding the calculation of latency and precision). A repeated-measures ANOVA with latency as dependent variable, and condition as independent variable revealed a main effect of condition on latency ($F(1.51,28.76) = 194.10$, $MSE=1698.53$, $p<0.001$). Bonferroni corrected t-tests (alpha=0.05/3, corrected for 3 tests) confirmed that latency was lower for the pre-cue condition than both the exogenous condition ($t(19)=-6.28$, $p<0.001$) and endogenous condition ($t(19) = -15.28$, $p<0.001$), and that the latency in the exogenous condition was lower than in the endogenous condition ($t(19) = -14.98$, $p<0.001$). The latency profile was consistent with results from the spatial attention literature showing a faster orienting for exogenous/peripheral cues compared to endogenous/central cues (fig. 2B, replicating a study with the same paradigm: Carlson et al., 2006; and other paradigms: see Carrasco et al, 2011 for a review).

To investigate the effect of cue on the precision of the response, a second ANOVA with concentration as dependent variable and the same independent variables showed a main effect of condition ($F(1.91,36.23) = 4.04$, $MSE=0.43$, $p=0.03$). Bonferroni-corrected (alpha=0.05, corrected for 3 tests), t-tests showed no significant difference between pre-cue and exogenous condition ($t(19) = -2.19$, $p=0.041$), between pre-cue and endogenous condition ($t(19) = 0.41$, $p=0.688$) or between exogenous and endogenous condition ($t(19)=2.44$, $p=0.025$).

Importantly, the profile of the concentration parameter in each condition suggested that all conditions led to roughly similar performance, with a slight - albeit non-significant - gain for exogenous attention (fig. 2C). The cue was therefore affecting the moment of attentional orienting, but not the quality of the resulting perceptual decision, at least between pre-cue and endogenous condition.

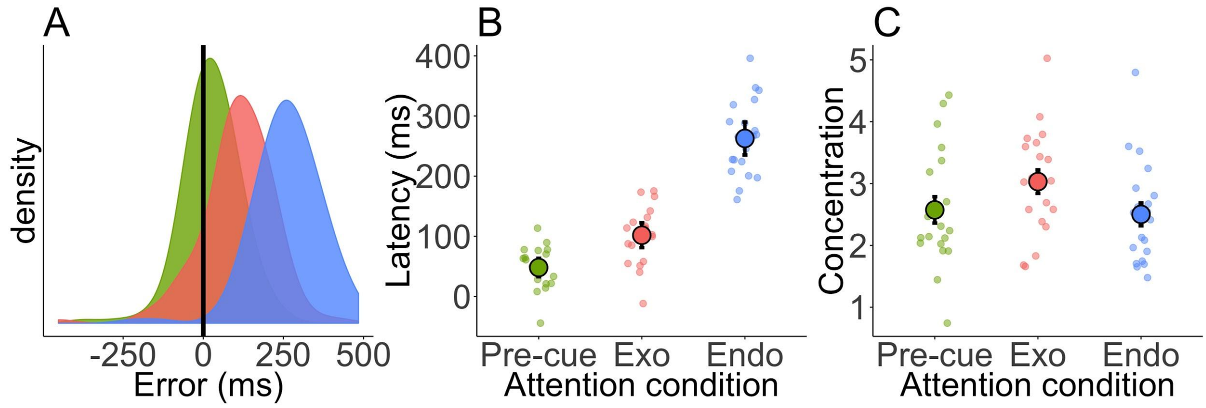

Figure 2. Latency and precision of attentional orienting. (A) Distributions of errors for a representative participant. The distribution of responses represents the angular error (objective phase minus reported phase) converted in ms. Von Mises distributions were fitted to estimate the latency (location parameter) and the precision (concentration parameter) of attentional selection. The pre-cue condition is pictured in green, exogenous and endogenous conditions are represented in red and blue, respectively. (B) The average latency for each condition. (C) The average concentration, a measure of precision, for each condition. Coloured dots correspond to individual participants in the given condition. Large dots with a black outline represent the mean across participants. Error bars represent across participants ±1 SEM.

Metacognition of attentional effects

First, as a measure of metacognitive ability, we calculated the difference in precision and latency for the "high confidence" trials compared to the "low confidence" trials. Location and concentration parameters were estimated separately for the high confidence trials and the low confidence trials (fig 3A) and then compared (fig 3B and 3C). A first ANOVA considered latency (location) as dependent variable and confidence x condition as independent variables. No significant effect of the latency on confidence ($F(1,19) = 0.99$, $MSE=341.19$, $p=0.33$) and no interaction was found ($F(1.73,32.92)=1.01$, $MSE=248.64$, $p=0.37$). the absence of a main effect of confidence suggests that confidence was oblivious to the delays induced by spatial orienting of attention.

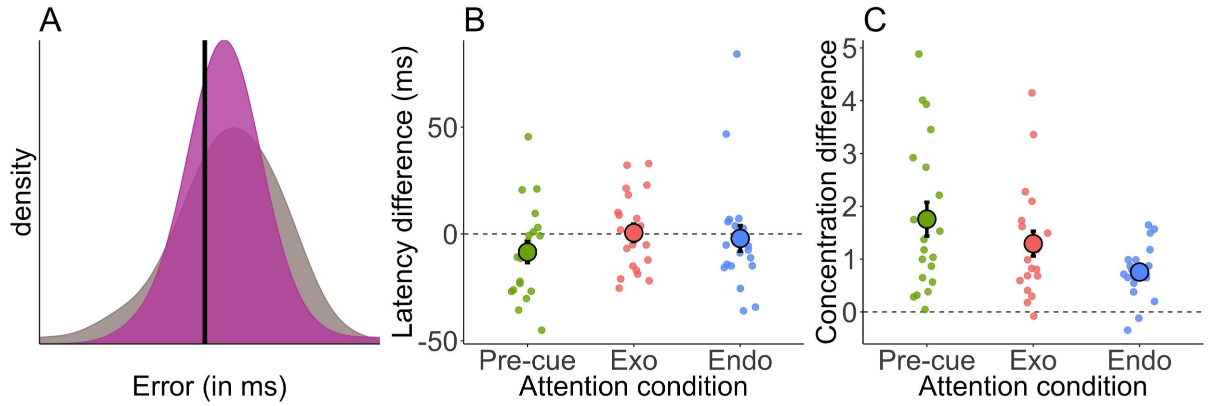

Figure 3. Metacognition of latency and precision. (A) Latency and concentration parameters were estimated by fitting Von Mises distribution to the High (purple) and Low (grey) confidence trials for each participant/condition. The parameter differences between High and Low confidence give an estimate of metacognitive access to these two dimensions. The figure plots the two distributions for a representative participant. (B) The average difference in latency between Low and High confidence trials for each condition. The absence of a significant difference suggests that confidence is oblivious to attentional latency. (C) The average difference in precision between High and Low confidence trials. Positive values suggest that confidence has access to the variability of response precision for all conditions, and even despite the significant metacognitive cost for endogenous orienting. Error bars represent across participants ±1 SEM.

A second ANOVA considered concentration (precision) as a dependent variable and confidence and condition as independent variables. A significant effect of confidence on the concentration parameter ($F(1,19)=48.21$, $MSE=1.0$, $p<0.001$), and an interaction between condition and confidence ($F(1.87,35.52)=7.06$, $MSE=0.38$, $p=0.003$) were observed. These results demonstrate that trials labelled with high confidence were associated with higher precision than low confidence trials. Bonferroni-corrected (alpha=0.05/3, corrected for 3 tests), t-tests showed a significant difference between the pre-cue and the endogenous condition ($T(19) = 179$, $p=0.004$), but not between pre-cue and exogenous ($t(19) = 1.99$, $p = 0.060$) or between exogenous and endogenous condition ($t(19) = 2.17$, $p = 0.043$). Together these results show that confidence is indeed able to access the magnitude of errors. Furthermore, they show that metacognitive ability is significantly greater for the pre-cue compared to the endogenous condition, pointing to a potential interaction between the orienting of voluntary attention and metacognitive ability.

Trial-by-trial metacognitive ability

The previous analyses suggested that participants were oblivious to latency, but were accurately monitoring the precision of their response across trials. Furthermore, orienting endogenous attention appeared to induce a decrease in metacognition of precision. These analyses however give us a broad picture of attention orientation across conditions, leaving open the question of error estimation on a trial level. A straightforward approach to this question is to assume that confidence has access to some form of evidence for each trial, which can be from the experimenter side related to the difference in absolute error between trial A and trial B in each pair (that is, the relative error magnitude). Given that confidence was not able to access the delay in phase report induced by attention (fig. 3B), and that we are here interested the response precision and not in the average bias (i.e., latency) , average latency in the considered condition/participant was systematically subtracted from the absolute error in each trial in the pair. For each pair, the difference in error (Δε) was calculated with the following formula:

$$\Delta\varepsilon = |\varepsilon_A - \mu| - |\varepsilon_B - \mu|$$

Where ε_A and ε_B are the error in first and second trial in the pair, and μ is the average error (or latency) for the considered participant/condition. A negative value of $\Delta\varepsilon$ would indicate a greater error for trial B, and a positive value a greater error for trial A. However, confidence calculations may be more complex, for example they may be taking into account the overall error amplitude (that is, the sum of the errors). The rational is that if the two errors in the pair are big, the ability to discriminate between them might be different from a situation where both errors are low, even when the difference between the two errors remains unchanged. This form of scaling is observed for first order decision (Shepard, 1987), and it has been proposed in the literature that confidence could follow a like scaling (Peirce & Jastrow, 1884; van den Berg, Yoo, & Ma, 2017). We therefore tested an alternative “scaling” model, where Δε was divided by the sum of the errors in the pair.

$$\frac{\Delta\varepsilon}{\Sigma\varepsilon} = \frac{|\varepsilon_A - \mu| - |\varepsilon_B - \mu|}{|\varepsilon_A - \mu| + |\varepsilon_B - \mu|}$$

For both the pure subtraction and scaling models, we used the relative error respective formula as a predictor of confidence in a logistic regression model (logit), estimated per participant and condition separately. Then we were able to compare the goodness-of-fit of each model using a Likelihood Ratio Test, because both models shared the same number of parameters. The scaling model significantly outperformed the pure subtraction model ($\chi^2(0) = 95.49$, $p<0.001$). We therefore selected the scaling model for all subsequent analyses. The slope (β) of the model gives an estimate of metacognitive ability, and the intercept provides an estimate of metacognitive bias in favour of the first trial in the pair (independently of actual performance). After Bonferroni correction (alpha = 0.05 / 3), the beta was significant for all three conditions, pre-cue ($t(19) = 7.40$, $p<0.001$), exogenous ($t(19) = 5.60$, $p<0.001$) and endogenous condition ($T(19) = 185$, $p = 0.001$), showing above chance metacognitive ability for all conditions at the group level. The bias, on the other hand, was significant for the pre-cue condition ($t(19) = -2.80$, $p = 0.011$) but not for exogenous ($t(19) = -0.94$, $p = 0.357$) and endogenous conditions ($t(19) = -0.97$, $p = 0.346$).

A repeated-measures ANOVA with metacognitive ability (β) as a dependent variable and condition as a independent variable showed a significant effect of condition ($F(1.95,37.13) = 4.11$, $MSE=0.35$, $p = 0.03$). After Bonferroni-correction (alpha=0.05/3, corrected for 3 tests), we found a significant difference between the pre-cue and the endogenous conditions ($T(19) = 179$, $p = 0.004$), but the differences between pre-cue and exogenous conditions ($t(19) = 1.46$, $p=0.160$) and between exogenous and endogenous conditions ($t(19) = 1.34$, $p=0.197$) failed to reach significance (fig. 4).

A second ANOVA with the bias (the intercept in the model) as a dependent variable and the same independent variables too showed a significant effect of condition ($F(1.96,37.20) = 8.65$, $MSE = 0.09$, $p<0.001$). Bonferroni-corrected t-tests confirmed a significantly greater bias for the pre-cue condition compared to the endogenous ($t(19) = -3.81$, $p=0.001$) and exogenous condition ($t(19) = -3.64$, $p=0.012$), but not between exogenous and endogenous conditions ($t(19) = -0.08$, $p = 0.937$).

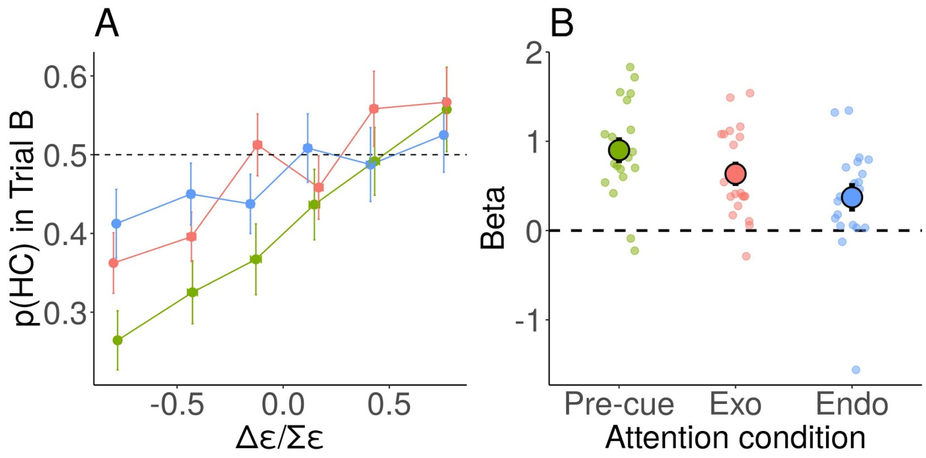

Figure 4. Orienting endogenous attention weakens metacognition. (A) The probability of High confidence for trial B in the pair, as a function of the scaled error difference between the two trials ($\frac{\Delta\varepsilon}{\Sigma\varepsilon}$) at the group level. For illustrative purposes, errors have been grouped by quantiles at the individual level, and the group average is represented with error bars for each quantile. The significant decrease in slope between the pre-cue (green) and endogenous condition (blue) confirms a metacognitive cost during voluntary orienting of spatial attention. (B) The average correlation coefficient (or slope) for each condition, used as a measure of metacognitive ability. Error bars represent across participants ±1 SEM.

RT TASK

The RT task was designed to probe the metacognitive ability in respect to the time of a distinct, motor process. We thus did not expect an effect of condition. This task included only the pre-cue and exogenous conditions. 2 participants were excluded from the analyses because of technical error (N = 18). In the following analyses, we used the median response time instead of the mean because response times are known to be non-Gaussian.

MAIN ANALYSIS

A repeated-measures ANOVA with median response time as a dependent variable and condition and confidence as independent variables showed an effect of confidence ($F(1,17) = 37.73$, $MSE=0$, $p<0.001$) but no effect of condition ($F(1,17) = 0.65$, $MSE=0$, $p = 0.43$) and no interaction ($F(1,17) = 1.80$, $MSE = 0.0$, $p=0.2$). Participants were therefore able to discriminate between fast and slow response times, and, as expected, condition did not significantly affect this ability (fig. 5A). For all subsequent analyses, we therefore combined both conditions together.

Just as we had for the main task, we evaluated how Type-2 comparison judgments (here, RT comparisons) could be predicted from the difference in Type-1 performance between the two trials (here, the difference in RTs), using a logistic regression. We then compared the pure subtraction model (ΔRT) to the scaling model ($\frac{\Delta RT}{\Sigma RT}$). Again, we found strong evidence in favour of the scaling model ($\chi^2(0)$ = 5405.7, p<0.001). We therefore used the scaling model in all subsequent analyses. Metacognitive ability (the β in the model) was significant at the group level (t(17) = 6.74, p<0.001), and we found no significant evidence for a metacognitive bias (the intercept in the model, t(17) = -1.75, p=0.098). These results demonstrate that participants are able to compare their reaction times on a trial-by-trial basis, and estimate the magnitude of the difference between the two trials.

Metacognition of internal timing

To find out whether our results for metacognition in the main attention task can be partially explained by metacognition of sensorimotor uncertainty (i.e., metacognition the RT task), we looked at whether participants having high or low sensorimotor uncertainty was a marker of the level of metacognitive ability. Recent work showed a specific role of motor preparation in shaping confidence (Fleming et al., 2015; Gajdos et al., 2019). Therefore, metacognition of RT taps into a specific kind of sensorimotor uncertainty by requiring the participants to evaluate their own response time following a simple stimulus onset event. A lack of correlation here would suggest that metacognitive ability in our main task cannot be reduced to sensorimotor uncertainty, and would thus be measuring metacognition of another process, probably linked to spatial attention. This analysis was conducted using the baseline pre-cue condition of the main task. We observed no significant correlation between metacognitive ability in each task (r = -0.189, 95% CI = [-0.304 0.603]; t(16) = -0.77, p=0.451). Our results thus suggest that metacognition of errors in the first task cannot be reduced to pure introspection of sensorimotor timing uncertainty.

Discussion

Our results shed light on three essential elements of attention orientation that we will discuss in turn. The first is the importance of the orienting process itself. The second element is the seemingly inconsistent oversight of delay: metacognition appeared to be strongly bound to attention, to the point of making confidence blind to the very limits of attention. Yet, because of this bound, metacognition still discriminated between different levels of response precision. Finally, the stark weakening of metacognitive ability during orienting of voluntary attention highlights the propensity of spatial attention to affect metacognition differently during and after orienting, possibly through top-down interactions.

Timing voluntary and involuntary attention

Our data fit well with the results from both time perception and attention literature. First, our results replicated previous studies that have found both exogenous and endogenous attention to be modulated by the perceived phase of moving clocks (Carlson et al., 2006; Chakravarthi & VanRullen, 2011; Hogendoorn et al., 2010). These results are also consistent with the observation that spatial attention modulates temporal resolution (Yeshurun & Levy, 2003) and that the reported time of visual events is directly affected by their relative distance from the attentional locus (Jovanovic & Mamassian, 2019).

Many studies however use a paradigm that involves a pre-cue in order to induce orientation of attention to a given location, and then present a target following a delay known to maximise attentional effects (e.g., 300ms). These paradigms however can overlook the variability of the orienting process from trial to trial: sometimes attention is allocated earlier, sometimes later. In our paradigm, the orientation of attention is expected to occur either at the beginning of the trial (pre-cue condition) or at the very moment the observer needs to register the phase of the clock (exogenous and endogenous condition). The current experimental design elicited a bias for the exogenous and endogenous condition compared to the pre-cue condition: on average, the reported phase was delayed, in accordance with the known latency for exogenous and endogenous attention. Importantly, even if our paradigm led to reasonable delay estimations for exogenous (~101ms) and endogenous (~262ms) attention, the absolute value of these mean errors in milliseconds is not necessary directly interpretable, as the use of temporally autocorrelated stimuli is known to affect perceived lag compared to decorrelated ones (Callahan-Flintoft, Holcombe, & Wyble, 2019; Sheth, Nijhawan, & Shimojo, 2000). It is mainly for this reason that they should be interpreted relatively to the pre-cue condition, where attention is pre-allocated at the right location.

On the other hand, we found no evidence for a difference between conditions regarding the precision of the response. Notably, average precision in the pre-cue and endogenous condition was matched, which allowed for a systematic analysis of confidence with equated performance across attention conditions. This equated performance was also robust to changes made to the underlying descriptive model (Von Mises distributions). Specifically, adding a guess parameter or allowing precision to fluctuate from trial-to-trial did not alter the original pattern (see Supplementary Material).

Finally, orienting endogenous attention did also elicit specific pupillary response profile compared to pre-allocated attention (see Supplementary Material). The phasic response following endogenous orienting was increased compared to pre-allocated attention, in a time window starting 660ms after cue onset and lasting until the end of the stimuli presentation. Notably, it was also possible to predict error magnitude from the trial-by-trial variability in pupil size from 0 to 364ms post-cue. This later result suggests that largely before the peak of the cue-locked pupillary response, pupil size carried meaningful information about task performance during voluntary orienting of spatial attention. This is interesting given that error magnitude was not decodable anymore at later stage, even at the peak of the task-evoked pupillary response. The temporal profile of the pupillary response was markedly different from the one of the pre-cue condition (the pre-cue condition showing no significant decoding period at all), despite sharing a voluntary or endogenous component.

Metacognition ignores attentional latency

We found that the average latency of attention was not accessible to confidence judgments. Participants appeared fully oblivious to the delay of both exogenous and endogenous attention (fig. 3B). This inability to monitor the delay of spatial attention mirrors what has been recently found for temporal attention (Recht et al., 2019). The majority of other studies that address this issue uses dual-task paradigms, showing that, metacognition ignores the delay in response times induced by the Psychological Refractory Period, in which the decision process for a second task is postponed until the decision process of a first task has been completed (Corallo et al., 2008; Marti et al., 2010). Our results show however that, without any need for a dual-task interaction, the very latency of both exogenous and endogenous attention remains concealed to metacognition.

The inability to monitor the average timing of cognitive processes need not preclude a fine-grained introspection of other aspects of processing, like the deviation from average latency (or relative error magnitude) or the nature of the inferential steps. Specifically, observers have been shown to be metacognitive aware of some of the processing stages during visual search and implicit spatial shifts of attention (Reyes & Sackur, 2014, 2017). In a similar vein, participants in our study were able to discriminate between error magnitudes, giving higher confidence to more precise trials for all conditions (fig. 3C). Furthermore, they were even able to accurately estimate error magnitude within each trial and use it in their confidence judgments (as shown by the positive slopes in fig. 4A). The reason for such metacognitive sensitivity to error variance rather than error mean can be explained by a simple generative model which would consider the internal evidence signal in a given trial as better approximated by a circular Gaussian (Von Mises) distribution over clock's phases. In this case, the distribution of errors observed in our data would be the result of the sampling process from this internal model. On average, the evidence will be greater for the phase corresponding to the mean parameter of the distribution. Making the assumption that confidence is a read-out of this evidence, confidence will be biased toward the mean of the distribution, but will still be lower on average for reported phases departing from the mean, exactly as observed in our data.

Allocating attention in space weakens metacognition

Attention takes time to be allocated, and requires cognitive control to be maintained (Carrasco, 2011). Our results demonstrate a direct cost of attention orientation in space for metacognition: during orienting, metacognitive processing of errors is altered compared to a condition where attention is already pre-allocated to the correct location. This metacognitive cost is observed despite perceptual report precision remaining unaffected by attentional orienting. Therefore, there is a bifurcation at some stage between the evidence used for perceptual report and the evidence used for metacognitive judgment. This relationship between Type-1 (i.e., phase reproduction) and Type-2 (i.e., confidence judgment) decisions is the subject of ongoing debates: the account of confidence using only the first-order decision evidence (e.g., Kiani & Shadlen, 2009) is challenged by numerous dissociations between confidence and accuracy, the existence of change of mind, and the empirical observation that confidence is on average stronger for correct choices than errors (for a review, see Fleming & Daw, 2017; Yeung & Summerfield, 2012). Our results show an interaction between the readout of Type-1 evidence by confidence and attentional allocation process. We propose that this interaction might reflect a post-decisional disruption of metacognitive evidence by the attentional system. The current finding that the process of allocating endogenous attention elicited greater metacognitive impairment might suggest that the top-down, frontal mechanisms needed for both voluntary attention orienting and metacognition could share certain central resources. For example, the neuroanatomical and functional bases of visual attention have been located within a large fronto-parietal network involving, amongst other areas, the frontal-eye-field (Buschman & Kastner, 2015), while the neural bases of visual metacognition are proposed to be mostly residing within the dorsolateral and anterior parts of the prefrontal cortex (Fleming, Ryu, Golfinos, & Blackmon, 2014; Fleming, Van Der Putten, & Daw, 2018; Fleming & Dolan, 2012; Shekhar & Rahnev, 2018). All of these regions have a strong implication in top-down cognitive control, biasing incoming signal from early visual cortices and monitoring perceptual selection and decision-making (Gilbert & Li, 2013; Rahnev, 2017). Further work will be needed to address how attention and metacognition interact at the functional level, to better understand the neural underpinnings of the metacognitive cost observed in the present study.

Conclusion

Metacognition allows individuals to reflect on the quality of their perceptual decisions. Yet, our results demonstrate that metacognition can be oblivious to the latency of spatial attention, an important modulator of perceptual accuracy. Furthermore, this experiment taps into the computational limitations of metacognition: the very process of orienting attention in space was found to weaken metacognitive ability. Together, our results provide invaluable information to our understanding of metacognition and its relationship with spatial attention.

Data availability

The data for the experiment is freely available via Open Science Framework.

Acknowledgements

This research was supported by the Agence Nationale de la Recherche (ANR-16-CE28-0002 and ANR-16-ASTR-0014 to VdG, ANR-17-EURE-0017 and ANR-18-CE28-0015 to PM) and by PSL University (doctoral scholarship to SR).

Contributions

SR, VdG and PM designed the experiment. SR conducted the experiment. SR, VdG and PM analyzed the data and wrote the manuscript.

Competing interests

The authors declare no competing interests.

REFERENCES

Boldt, A., Blundell, C., & De Martino, B. (2019). Confidence modulates exploration and exploitation in value-based learning. Neuroscience of Consciousness. https://doi.org/10.1093/nc/niz004

Buschman, T. J., & Kastner, S. (2015). From Behavior to Neural Dynamics: An Integrated Theory of Attention. Neuron, 88(1), 127–144. https://doi.org/10.1016/j.neuron.2015.09.017

Callahan-Flintoft, C., Holcombe, A. O., & Wyble, B. (2019). A delay in sampling information from temporally autocorrelated visual stimuli. BioRxiv, 656850. https://doi.org/10.1101/656850

Carlson, T. A., Hogendoorn, H., & Verstraten, F. A. J. (2006). The speed of visual attention: What time is it? Journal of Vision, 6, 1406–1411. https://doi.org/10.1167/6.12.6

Carrasco, M. (2011). Visual attention: The past 25 years. Vision Research. https://doi.org/10.1016/j.visres.2011.04.012

Chakravarthi, R., & VanRullen, R. (2011). Bullet trains and steam engines: Exogenous attention zips but endogenous attention chugs along. Journal of Vision. https://doi.org/10.1167/11.4.1

Corallo, G., Sackur, J., Dehaene, S., & Sigman, M. (2008). Limits on Introspection. Psychological Science. https://doi.org/10.1111/j.1467-9280.2008.02211.x

Coull, J. T., & Nobre, A. C. (1998). Where and when to pay attention: the neural systems for directing attention to spatial locations and to time intervals as revealed by both PET and fMRI. Journal of Neuroscience, 18(18), 7426–7435. https://doi.org/0270-6474/98/187426-10$05.00/0

Denison, R. N., Adler, W. T., Carrasco, M., & Ma, W. J. (2018). Humans incorporate attention-dependent uncertainty into perceptual decisions and confidence. Proceedings of the National Academy of Sciences, 201717720. https://doi.org/10.1073/pnas.1717720115

Desender, K., Boldt, A., & Yeung, N. (2018). Subjective Confidence Predicts Information Seeking in Decision Making. Psychological Science. https://doi.org/10.1177/0956797617744771

Fleming, S. M., Van Der Putten, E. J., & Daw, N. D. (2018). Neural mediators of changes of mind about perceptual decisions. Nature neuroscience, 21(4), 617.

Fleming, S. M., & Daw, N. D. (2017). Self-evaluation of decision performance: A general Bayesian framework for metacognitive computation. Psychological Review, 124(1), 1–59. https://doi.org/10.1002/smll.))

Fleming, S. M., Maniscalco, B., Ko, Y., Amendi, N., Ro, T., & Lau, H. (2015). Action-Specific Disruption of Perceptual Confidence. Psychological Science. https://doi.org/10.1177/0956797614557697

Fleming, S. M., Ryu, J., Golfinos, J. G., & Blackmon, K. E. (2014). Domain-specific impairment in metacognitive accuracy following anterior prefrontal lesions. Brain, 137(10), 2811-2822.

Fleming, S. M., & Dolan, R. J. (2012). The neural basis of metacognitive ability. Philosophical Transactions of the Royal Society B: Biological Sciences, 367(1594), 1338-1349.

Gajdos, T., Fleming, S. M., Saez Garcia, M., Weindel, G., & Davranche, K. (2019). Revealing subthreshold motor contributions to perceptual confidence. Neuroscience of Consciousness. https://doi.org/10.1093/nc/niz001

Guggenmos, M., Wilbertz, G., Hebart, M. N., & Sterzer, P. (2016). Mesolimbic confidence signals guide perceptual learning in the absence of external feedback. ELife. https://doi.org/10.7554/eLife.13388

Hainguerlot, M., Vergnaud, J. C., & De Gardelle, V. (2018). Metacognitive ability predicts learning cue-stimulus associations in the absence of external feedback. Scientific Reports. https://doi.org/10.1038/s41598-018-23936-9

Hogendoorn, H., Carlson, T. A., VanRullen, R., & Verstraten, F. A. J. (2010). Timing divided attention. Attention, Perception, and Psychophysics, 72(8), 2059–2068. https://doi.org/10.3758/APP.72.8.2059

Jovanovic, L., & Mamassian, P. (2019). When an Event Is Perceived Depends on Where We Attend. I-Perception. https://doi.org/10.1177/2041669519858096

Kiani, R., & Shadlen, M. N. (2009). Representation of confidence associated with a decision by neurons in the parietal cortex. Science (New York, N.Y.), 324(5928), 759–764. https://doi.org/10.1126/science.1169405

Mamassian, P. (2016). Visual Confidence. Annual Review of Vision Science, 2(1), 459–481. https://doi.org/10.1146/annurev-vision-111815-114630

Marti, S., Sackur, J., Sigman, M., & Dehaene, S. (2010). Mapping introspection's blind spot: Reconstruction of dual-task phenomenology using quantified introspection. Cognition. https://doi.org/10.1016/j.cognition.2010.01.003

Peirce, C., & Jastrow, J. (1884). On small differences in sensation. Retrieved from https://philarchive.org/archive/PEIOSD

Peirce, J. W. (2007). PsychoPy-Psychophysics software in Python. Journal of Neuroscience Methods. https://doi.org/10.1016/j.jneumeth.2006.11.017

Rahnev, D., Maniscalco, B., Graves, T., Huang, E., de Lange, F. P., & Lau, H. (2011). Attention induces conservative subjective biases in visual perception. Nature Neuroscience, 14(12), 1513–1515. https://doi.org/10.1038/nn.2948

Recht, S., Mamassian, P., & de Gardelle, V. (2019). Temporal attention causes systematic biases in visual confidence. Scientific Reports, 9(1), 11622. https://doi.org/10.1038/s41598-019-48063-x

Reyes, G., & Sackur, J. (2014). Introspection during visual search. Consciousness and Cognition. https://doi.org/10.1016/j.concog.2014.08.009

Reyes, G., & Sackur, J. (2017). Introspective access to implicit shifts of attention. Consciousness and Cognition, 48, 11–20. https://doi.org/10.1016/j.concog.2016.10.003

Shekhar, M., & Rahnev, D. (2018). Distinguishing the roles of dorsolateral and anterior PFC in visual metacognition. Journal of Neuroscience, 38(22), 5078-5087.

Shepard, R. N. (1987). Toward a universal law of generalization for psychological science. Science. https://doi.org/10.1126/science.3629243

Sheth, B. R., Nijhawan, R., & Shimojo, S. (2000). Changing objects lead briefly flashed ones. Nature Neuroscience. https://doi.org/10.1038/74865

van den Berg, R., Yoo, A. H., & Ma, W. J. (2017). Fechner's law in metacognition: A quantitative model of visual working memory confidence. Psychological Review. https://doi.org/10.1037/rev0000060

van den Berg, R., Zylberberg, A., Kiani, R., Shadlen, M. N., & Wolpert, D. M. (2016). Confidence Is the Bridge between Multi-stage Decisions. Current Biology. https://doi.org/10.1016/j.cub.2016.10.021

Wilimzig, C., Tsuchiya, N., Fahle, M., Einhäuser, W., & Koch, C. (2008). Spatial attention increases performance but not subjective confidence in a discrimination task. Journal of Vision, 8(5), 7. https://doi.org/10.1167/8.5.7

Yeshurun, Y., & Levy, L. (2003). Transient spatial attention degrades temporal resolution. Psychological Science. https://doi.org/10.1111/1467-9280.02436

Yeung, N., & Summerfield, C. (2012). Metacognition in human decision-making: confidence and error monitoring. Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences, 367(1594), 1310–1321. https://doi.org/10.1098/rstb.2011.0416

Zizlsperger, L., Sauvigny, T., & Haarmeier, T. (2012). Selective attention increases choice certainty in human decision making. PLoS ONE. https://doi.org/10.1371/journal.pone.0041136

Zizlsperger, L., Sauvigny, T., Händel, B., & Haarmeier, T. (2014). Cortical representations of confidence in a visual perceptual decision. Nature Communications. https://doi.org/10.1038/ncomms4940

Zylberberg, A., Wolpert, D. M., & Shadlen, M. N. (2018). Counterfactual Reasoning Underlies the Learning of Priors in Decision Making. Neuron. https://doi.org/10.1016/j.neuron.2018.07.035

CHAPTER 4 | SUPPLEMENTARY MATERIAL

Orienting spatial attention weakens metacognition

Supplementary Material

Samuel Recht (1)*, Pascal Mamassian (1)+, Vincent de Gardelle (2)+

(1) Laboratoire des systèmes perceptifs, Département d'études cognitives, École normale supérieure, PSL University, CNRS, 75005 Paris, France

(2) CNRS and Paris School of Economics, Paris, France

* Corresponding author: samuel.recht@gmail.com

+ these authors share senior authorship

TIMING SPATIAL ATTENTION

We assume that the distribution of errors in a given condition follows a circular normal distribution (also referred to as a Von Mises distribution):

$$f(x \mid \mu, k) \; = \; \frac{e^{k\, cos(x-\mu)}}{2\pi I_0(k)}$$

Where $I_0$is the modified Bessel function of the first kind of order 0, μ is the location parameter (equivalent to the mean in a normal distribution), k is the concentration parameter ($1/\,k$ is analogous to the variance in a normal distribution) and x the angular error in a given trial.

It has been suggested that the distribution of errors following the encoding of a stimulus into working memory can be modelled as a mixture of a Von Mises and a uniform distribution, the later accounting for guesses, which represent no encoding at all (Bays, Catalao, & Husain, 2009; Zhang & Luck, 2008). We therefore checked that a model involving a mixture of the two was better at explaining our data. The mixture model was defined as follows:

$$f(x \mid \mu, k, g) \ = g \ \frac{1}{2\pi} + (1 - g) \ \frac{e^{k \ cos(x-\mu)}}{2\pi I_0(k)}$$

Where I_0 is the modified Bessel function of the first kind of order 0, μ and k are the location and concentration parameters, respectively, g is the guess rate and x the angular error in a given trial. The guess rate however could be shared across attentional conditions or not, we therefore had two variants of the Von Mises + guess model: one with a shared guess rate across conditions ('VM+FG', 7 parameters) and one with a specific guess rate for each condition ('VM+G', 9 parameters).

A second line of thought in the working memory literature is that encoding from trial to trial is of variable precision rather than constant (Fougnie, Suchow, & Alvarez, 2012; Van Den Berg, Shin, Chou, George, & Ma, 2012). In this case, errors are coming from a mixture of Von Mises distributions with their concentration following a higher order distribution (often a Gamma distribution). We therefore tested a third, variable-precision model (adapted from Van Den Berg et al., 2012). Contrary to Van Den Berg and colleagues, we didn't used the Fisher's information (J) as the measure of precision, but we directly used the concentration parameter (k) instead. The Fisher's information being monotonically related to k, we kept the later to make it comparable to our main model ('pure VM').

$$f(x \mid \mu, \bar{k}, \tau) \ = \int_0^\infty \ \frac{e^{k \ cos(x-\mu)}}{2\pi I_0(k)} \ Gamma(k; \frac{\bar{k}}{\tau}, \tau) \ dk$$

Where I_0 is the modified Bessel function of the first kind of order 0, μ is the location parameter of the Von Mises distributions, $\frac{\bar{k}}{\tau}$ is the shape parameter (with $\bar{k}$ as the mean concentration) and τ the scale parameter of the gamma distribution, x the angular error in a given trial. The variable-precision model ('VP') is fitted separately for each condition, and therefore has 9 parameters. We also tested three other variants: one with a fixed shape, but variable scale parameter across conditions ('VP-FSh', 7 parameters), one with fixed scale but variable shape parameter ('VP-FSc', 7 parameters) and finally, one with both shape and scale parameters fixed across all attentional conditions ('VP-F', 5 parameters).

All of the tested models involved fitting a specific location parameter (μ) for each condition, in light of the strong and systematic difference in average latency observed between attentional conditions (Carlson, Hogendoorn, & Verstraten, 2006; Chakravarthi & VanRullen, 2011; Hogendoorn, Carlson, VanRullen, & Verstraten, 2010). Note that our model comparison approach was not meant to be fully exhaustive, but rather to check that our results hold when considering possible alternatives to our definition of precision.

Models were fitted using Maximum Likelihood Estimation (MLE). All analyses were carried out using R programming language. BIC and AIC were estimated for each model, and the difference between the pure VM model and the other models for each estimator is denoted ΔBIC and ΔAIC. A negative value suggests a better fit for the pure VM model. BIC is known to penalize more heavily the number of parameters than AIC.

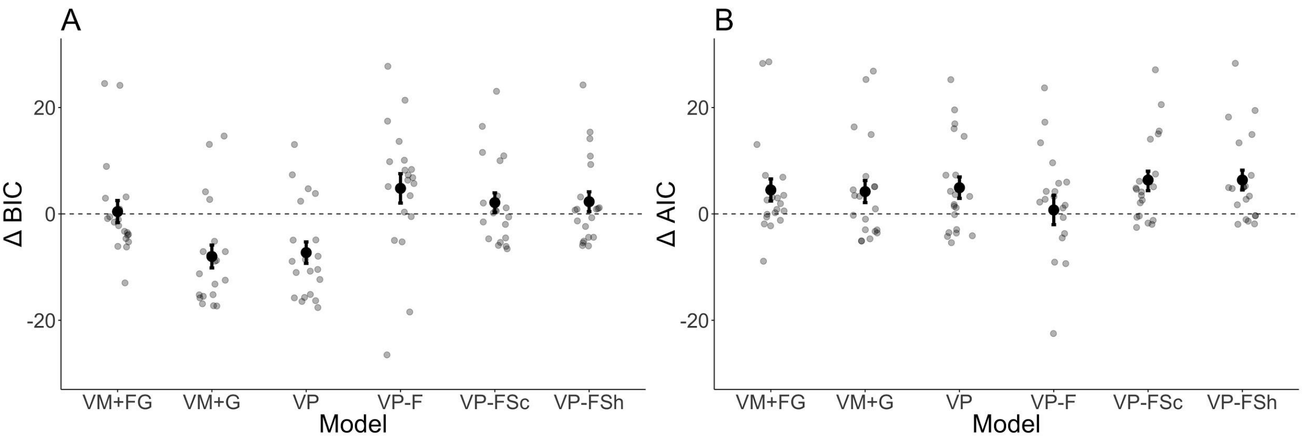

Figure S1. BIC and AIC comparison. (A) The difference in the Bayesian Information Criterion (BIC) between the pure Von Mises and each of the other models. A negative value suggests evidence in favour of the pure Von Mises model. (B) Same measure but using the Akaike Information Criterion (AIC). Low alpha dots correspond to individual participants in the given model. Black dots represent the mean across participants. Error bars represent across participants SEM.

When considering the Von Mises + guess family models, a first important observation is that the VM+G model, which supposes a variable guess rate between conditions, was significantly worse than the pure Von Mises, according to ΔBIC (T(19) = 27, p=0.002) and not significantly different according to ΔAIC (T(19) = 146, p=0.133). Importantly, it also performed significantly worse than the model with shared guess rate across conditions relative to BIC (T(19) = 201, p=0.002), the difference in AIC between these two models was not significant (T(19) = 145, p = 0.143). It is therefore highly unlikely that a change in guess rate between attentional conditions would explain the difference in metacognition observed in our data. The benefit of adding a stable guess rate across condition (VM+FG) was unclear, with only the AIC favouring this model (T(19) = 166, p=0.021), but not the BIC (T(19) = 74, p = 0.261).

When we look at the variable-precision models, the worst model was the full VP, which fitted a specific set of shape and scale parameters to each condition. This model's BIC was significantly worse than the pure VM (T(19)=26, p=0.002) and there was no significant difference in AIC (T(19) = 154, p=0.069). The VP-F, which fixes the parameters across conditions, was not significantly better than the pure VM for BIC (t(19)=1.74, p=0.098) nor AIC (t(19) = 0.27, p=0.793). When fixing one parameter of the VP, we found no significant difference in BIC (for VP-FSh: T(19)=118, p=0.647; for VP-FSc: T(19) = 119, p=0.622), but a lower AIC for both models (for VP-FSh: T(19) = 187, p=0.002; for VP-FSc: T(19) = 188, p=0.001). The average ΔAIC was 6.19 for the model with fixed shape and 6.37 for the fixed scale model. Therefore, both models were accounting equally well for the data, but the evidence favouring these models over the pure Von Mises was fairly low, particularly when using BIC.

Robustness of precision estimates

Despite the low evidence in favour of a guess rate, we nevertheless checked that concentration remained stable across conditions when accounting for guesses. We analysed the effect of condition on the precision parameters of the Von Mises + Fixed Guess model which was the only plausible candidate given our model comparison. The estimated values are shown in figures S2A and B. The latency parameter was not expected to vary at all from one model to the other, but the concentration could change - in theory uniformly - because of the added guess rate parameter. A repeated-measures ANOVA with latency as dependent variable and condition as independent variable revealed a main effect of condition on latency ($F(1.52,28.82) = 201.84$, $MSE=1651.17$, $p<0.001$). Bonferroni corrected paired t-tests confirmed that latency was lower in the pre-cue condition than for the exogenous ($t(19) = -6.51$, $p<0.001$) and endogenous ($t(19) = -15.64$, $p<0.001$) conditions, and that the latency in the exogenous condition was lower than in the endogenous condition ($t(19) = -15.12$, $p<0.001$).

A second ANOVA with concentration as dependent variable showed a main effect of condition ($F(1.95,37.06) = 3.95$, $MSE=0.94$, $p=0.03$). Bonferroni-corrected (alpha=0.05, corrected for 3 tests), paired t-tests showed no significant difference between the pre-cue and exogenous conditions ($t(19) = -1.03$, $p=0.315$) nor between the pre-cue and endogenous conditions ($t(19) = 1.69$, $p=0.108$), but a significant difference between the exogenous and endogenous conditions ($t(19)=3.00$, $p=0.007$).

The Von Mises + Fixed Guess model was therefore leading to the exact same conclusions as the pure Von Mises regarding both latency and concentration.

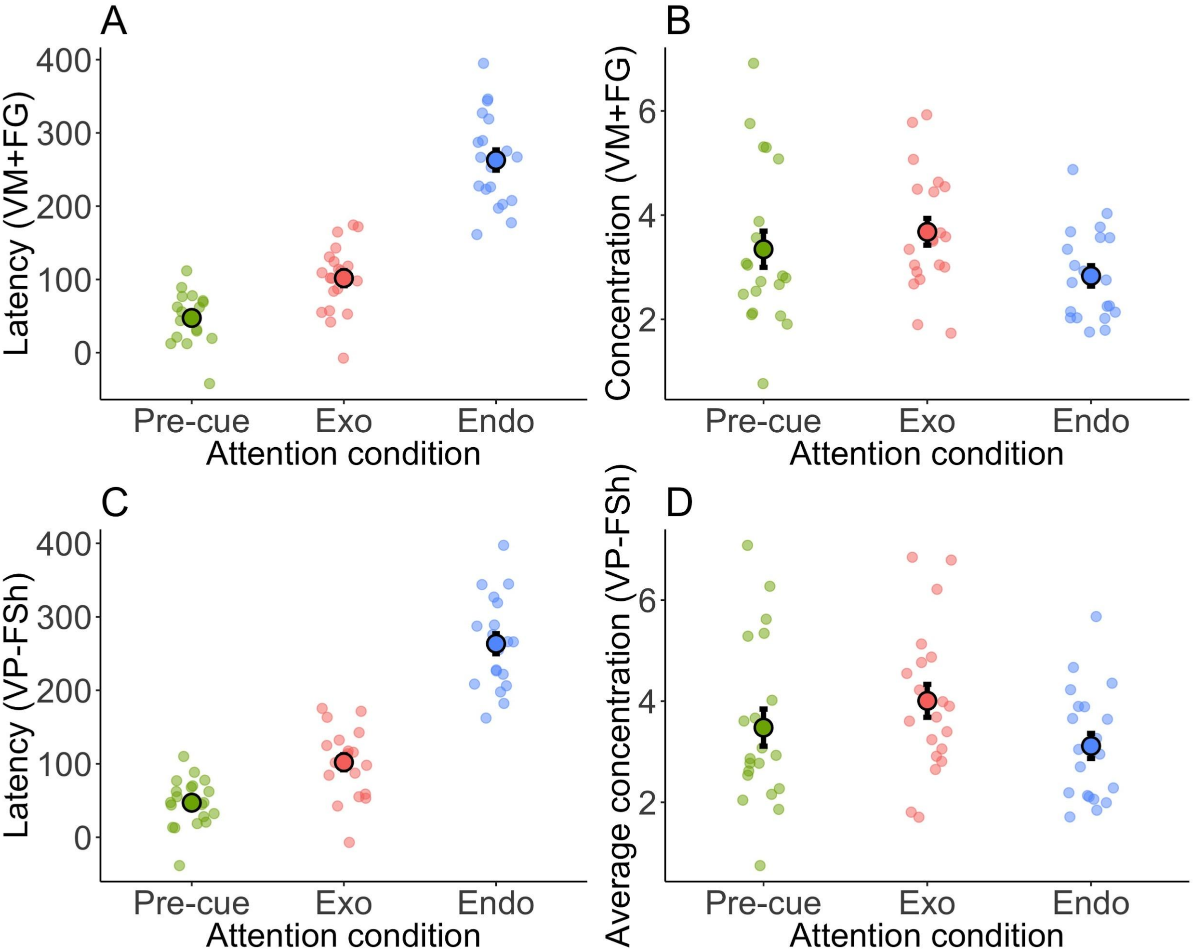

Figure S2. Estimated attention parameters for the two best alternative models. (A) The latency for each attentional orienting condition in the Von Mises + Fixed Guess model. (B) The precision for each condition in the Von Mises + Fixed Guess model. There is no significant difference between the pre-cue and exogenous/endogenous conditions. (C) The latency for each attentional orienting condition in the Variable-precision with fixed shape parameter model. (D) The average concentration for each attentional orienting condition in the Variable-precision with fixed shape parameter model. There is no significant difference between the pre-cue and exogenous/endogenous conditions. Coloured dots correspond to individual participants in the given condition. Black-outlined dots represent the mean across participants. Error bars represent across participants SEM.

The absence of difference in guess rate between conditions, and the stability of the concentration parameter pattern between the pure Von Mises and the Von Mises + Fixed Guess rate models rule out a strong effect of guess rate during attentional orienting. Furthermore, it has been suggested that models involving guess rates (like the Zhang & Luck model used in the present analysis) should be interpreted with caution given the risk of inflated guess rate estimates. In particular, this risk has been shown to exist when the true generative process is a variable precision model involving zero guess rate (Ma, 2018), or when the error space is non-linearly related to the stimulus space (Schurgin, Wixted, & Brady, 2018).

In a second control analysis, we considered the possibility of a variable precision across trials. To check whether a strong difference in average precision between the different attentional conditions when the VP model was used, we selected the Variable-precision with fixed shape model (figure S2, C and D). A repeated-measure ANOVA was applied resulting parameters values with average concentration as a dependent variable, and condition as an independent variable. The effect of condition on the average concentration was significant ($F(1.98, 37.69) = 4.06$, $MSE=1.00$, $p=0.03$), but this effect was driven by a higher average precision in the exogenous compared to endogenous condition ($t(19) = 2.78$, $p=0.012$). The difference between the pre-cue and endogenous/exogenous conditions was not significant (all $p>0.117$, Bonferroni-corrected with $\alpha=0.05/3$). We then run an ANOVA on latency. It confirmed the effect of condition on latency ($F(1.51, 28.74) = 203.46$, $MSE=1642.46$, $p<0.001$). The difference between the pre-cue and exogenous conditions ($t(19) = -6.62$, $p<0.001$), the pre-cue and endogenous conditions ($t(19) = -15.7$, $p<0.001$) and between the exogenous and endogenous conditions ($t(19) = -15.12$, $p<0.001$) were all significant after Bonferroni-correction ($\alpha=0.05/3$). These results are all fully consistent with what was observed using the pure VM model and confirm the robustness of this model in analysing our experimental results.

Together these results, in tandem with the low evidence for a strictly better performing model over the pure Von Mises, suggest that our attentional manipulation strongly affected average latency (μ) but not precision (k) of the response distributions. Importantly, this was true regardless of the metric used (fixed or variable concentration). Moreover, adding a guess rate parameter was only weakly beneficial when the guess rate was fixed across conditions. Adding this stable guess rate did not alter the original pattern (figure S2).

Pupillary analysis

Rational

Pupil size has been shown to correlate with attention, amongst other cognitive states such as alertness or cognitive load (Hoeks & Levelt, 1993; Wierda, van Rijn, Taatgen, & Martens, 2012). For overt spatial attention, when a saccadic shift is initiated towards a location of interest, the cortical processing of the landing location is known to start prior to saccade onset (Kowler, 2011). This preparatory process has been proposed to directly affect pupillary response via the intermediary layers of the Superior Colliculus (Wang & Munoz, 2018). Furthermore, orienting covert spatial attention to a particular location has been shown to increase pupillary response to local luminance at attended compared to unattended locations (Binda & Murray, 2015; Mathôt, Dalmaijer, Grainger, & Van der Stigchel, 2014). This specific pupillometric signature of attention has been recently used to measure the size of the attentional windows in space (Tkacz-Domb & Yeshurun, 2018; Yeshurun, 2019). Pupil dilation is not only observed during changes of local or global luminance, but also when an observer performs a cognitive task, a phenomenon coined task-evoked pupillary response (Beatty, 1982). Previous studies have demonstrated that pupil size is modulated by the neuromodulatory activity of the brainstem and is a known behavioural marker of central arousal (McGinley et al., 2015).

Pupil size and confidence have also been the subject of recent studies (Colizoli, de Gee, Urai, & Donner, 2018; Lempert, Chen, & Fleming, 2015). However, the relation between pupillary response, the timing of spatial attention, and metacognition has not, to our knowledge, been addressed yet. We hypothesised that given the specific time course of the pre-cue, exogenous and endogenous conditions, we might expect differences in the pupillary response following cue onset. Pupil size has been shown to vary with uncertainty and confidence, therefore we were also expecting a possible interaction between pupil dilation and confidence judgments.

Preprocessing & analyses

Pupillary data were collected using a 250 Hz sampling rate, Eyelink 1000+ (SR Research). Eyeblinks were detected using the Eyelink detection algorithm, and pupil area was linearly interpolated for each eyeblink period. For each trial, we epoched pupillary data locked to cue onset, the event of interest. The post-cue values were baseline-corrected by subtracting the average pupil size of a 400ms window before cue onset.

An effect was considered to be significant when $t > 2$ for 200ms or more, equivalent to a threshold of $p<0.05$ (Mathôt et al., 2014; Tkacz-Domb & Yeshurun, 2018).

Results

Attention: Overall pupillary response

To investigate the effect of condition on the pupillary response, we calculated the grand average for each participant and condition (fig. S3A). To test the effect of attentional condition on pupil area, we used a mixed effects model comparison approach, with fixed effects of condition and participants as random intercepts. This was done for each time bin. We found no significant difference between the pre-cue and exogenous conditions during the whole period (from 0 after to 2000ms post-cue). However, we found a significant difference in pupillary response between the pre-cue and endogenous conditions starting 660ms after cue onset and lasting until the end of the analysis window (2000ms).

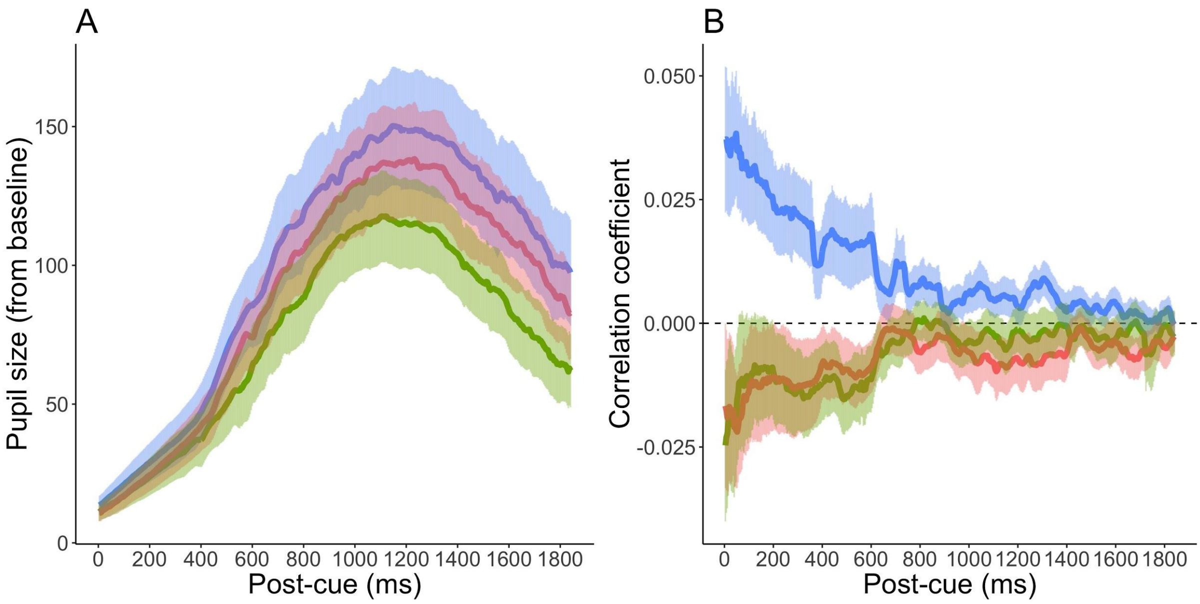

Figure S3. Pupillometry correlates of attentional orienting. (A) Overall pupillary response. The average, cue-locked and baseline-corrected pupil size as a function of time and attentional condition. 0 represents cue onset. Pre-cue: green curve; exogenous: red curve; endogenous: blue curve. (B) Trial-by-trial analysis. The per-condition correlation coefficient between error magnitude and pupil size for each time bin following cue onset. A positive value indicates that pupil size is positively correlated with error magnitude on a trial-by-trial basis. Light shading represents across participants SEM.

ATTENTION: TRIAL-BY-TRIAL ANALYSIS

Next, we considered the relation between pupil size and raw absolute error. We used the pupil size at each time point, separately for each participant/condition, as a regressor to predict error magnitude (fig. S3B). More precisely, the vector of per-trial pupil size for a given participant, condition and time bin was used to predict the vector of per-trial error magnitude. This analysis permitted us to investigate the task-related information embedded into pupillary data with high temporal resolution.

We started by assessing the correlation between pupil size and error for each time bin. our analysis revealed that pupil size significantly predicted error magnitude for the endogenous condition, from 0 to 364ms after cue onset. but not for the pre-cue and exogenous conditions at any time point. Then, we tested the effect between conditions using a mixed-effects model with condition as fixed effect and participants as random intercepts, applied to the distribution of correlation coefficients separately for each time bin. The difference between the pre-cue and endogenous conditions was significant from 0 to 624ms after cue onset, but there was no difference between pre-cue and exogenous conditions at any time point. This result highlights the specific role arousal states played during the process of endogenous orienting.

METACOGNITION: OVERALL PUPILLARY RESPONSE

We further investigated the pupillary response for high and low confidence trial, independently of condition. We calculated the average for each participant and confidence level (fig. S4A). To test the effect of confidence on pupil area, we used a mixed effects model comparison approach, with fixed effect of confidence and participants as random intercepts. This was done for each time bin. Our analysis revealed no significant differences in pupillary response between high and low confidence trials.

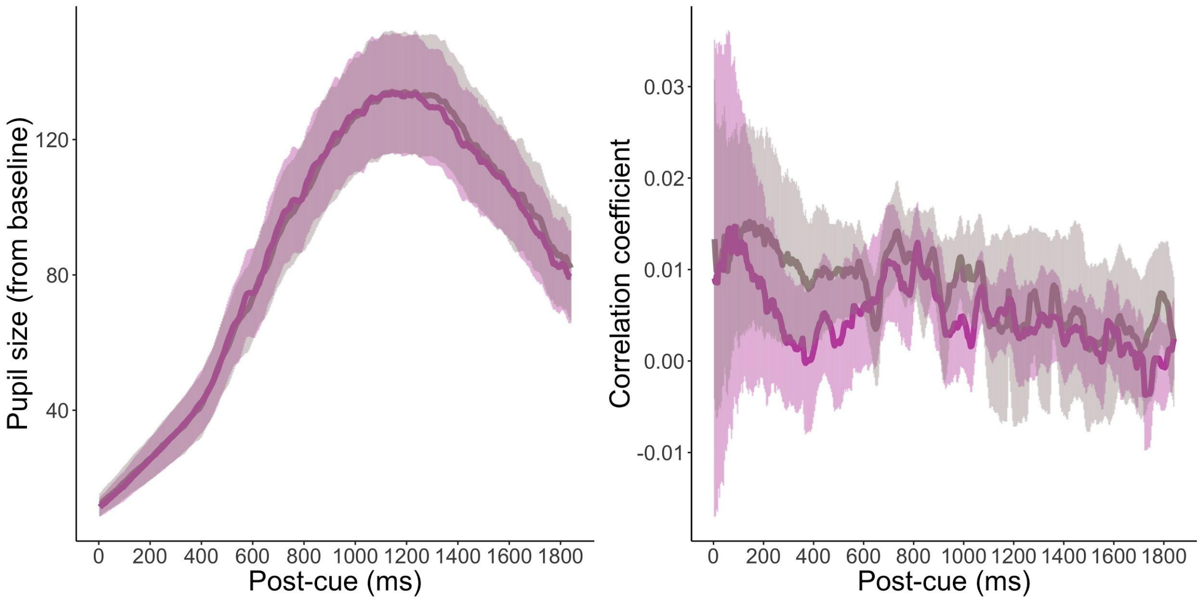

Figure S4. Pupillometry correlates of confidence. (A) Overall pupillary response. The average, cue-locked and baseline-corrected pupil size as a function of time and confidence. Purple line represents high confidence trials. 0 represents cue onset. (B) Trial-by-trial analysis. The correlation coefficient between error magnitude and pupil size for each time bin following cue onset. Purple line represents high confidence trials, and grey line low confidence trials. A positive value suggests error magnitude is positively correlated with pupil size. Light shading represents across participants SEM.

Confidence: trial-by-trial analysis

As we had for attentional orienting, we investigated the relation between pupil size and error magnitude, but this time as a function of confidence. We used the pupil size at each time point, separately for each participant/confidence level, as a regressor to predict error magnitude (fig. S4B). We found no significant correlation either for high or for low confidence, and no significant difference between confidence levels.

Metacognition: trial-by-trial analysis

Finally, to determine the relation between pupil size and metacognition in a trial-by-trial analysis, we examined the difference in pupil size between the high and low confidence trials in each pair, and how this difference correlates with our measure of metacognitive evidence ($\Delta\varepsilon/\Sigma\varepsilon$). We found no significant time points where difference in pupil size predicted metacognitive evidence, and no significant difference between conditions.

Specific pupillary signature when deploying voluntary attention

Together, our results point to the implication of the central arousal system during the initiation of an endogenous attentional episode. First, orienting of endogenous attention is marked by an increase in phasic pupillary response compared to sustained attention (endogenous vs pre-cue, fig. S3A). Secondly, the analysis of the fine temporal structure of the correlation between errors and pupil size showed an early time window during which the spontaneous fluctuation of the central arousal system was determinant in shaping trial-by-trial error. From 0 to 364ms, the magnitude of error in a given trial can be predicted from pupillary data. This is not reducible to a cue-initiated pupillary response. The cue-initiated pupillary response took longer to unfold (fig S3A). This effect is thus best interpreted as the state of the central arousal system at the very moment of cue onset. Importantly, the pupillary pattern of voluntary, endogenous orienting was markedly different from a situation where an observer has pre-allocated endogenous attention to the cued location (pre-cue condition in green on fig S3B) or when automatic, exogenous orienting attention was involved (exogenous condition in red on fig. S3B). These effects are observed despite matched response precision between pre-cue and endo conditions (see Main Paper), and as such cannot be reduced to a difference in task difficulty.

In sum, we found no evidence for a specific pupillary signature of confidence in the post-cue period. There was no evidence for high/low confidence trials in shaping this profile (fig. S4A), nor for determining confidence using per-trial pupil size (fig. S4B), only the pupillary instantaneous state in the endogenous condition appeared to carry information about error magnitude (fig. S3B). Interestingly, this condition corresponds to one in which metacognitive ability is lowest. Finally, regarding metacognition, we didn't find any effect of difference in pupil size. These null results should nonetheless be interpreted with caution because they were done on a half the data points, and may thus be due to noise from low statistical power.

References

Bays, P. M., Catalao, R. F. G., & Husain, M. (2009). The precision of visual working memory is set by allocation of a shared resource. Journal of Vision. https://doi.org/10.1167/9.10.7

Beatty, J. (1982). Task-evoked pupillary responses, processing load, and the structure of processing resources. Psychological Bulletin. https://doi.org/10.1037/0033-2909.91.2.276

Binda, P., & Murray, S. O. (2015). Spatial attention increases the pupillary response to light changes. Journal of Vision. https://doi.org/10.1167/15.2.1

Carlson, T. A., Hogendoorn, H., & Verstraten, F. A. J. (2006). The speed of visual attention: What time is it? Journal of Vision, 6, 1406–1411. https://doi.org/10.1167/6.12.6

Chakravarthi, R., & VanRullen, R. (2011). Bullet trains and steam engines: Exogenous attention zips but endogenous attention chugs along. Journal of Vision. https://doi.org/10.1167/11.4.1

Colizoli, O., de Gee, J. W., Urai, A. E., & Donner, T. H. (2018). Task-evoked pupil responses reflect internal belief states. Scientific Reports. https://doi.org/10.1038/s41598-018-31985-3

Fougnie, D., Suchow, J. W., & Alvarez, G. A. (2012). Variability in the quality of visual working memory. Nature Communications. https://doi.org/10.1038/ncomms2237

Hoeks, B., & Levelt, W. J. M. (1993). Pupillary dilation as a measure of attention: a quantitative system analysis. Behavior Research Methods, Instruments, & Computers. https://doi.org/10.3758/BF03204445

Hogendoorn, H., Carlson, T. A., VanRullen, R., & Verstraten, F. A. J. (2010). Timing divided attention. Attention, Perception, and Psychophysics, 72(8), 2059–2068. https://doi.org/10.3758/APP.72.8.2059

Kowler, E. (2011). Eye movements: The past 25 years. Vision Research. https://doi.org/10.1016/j.visres.2010.12.014

Lempert, K. M., Chen, Y. L., & Fleming, S. M. (2015). Relating Pupil Dilation and Metacognitive Confidence during Auditory Decision-Making. Plos One, 10(5), e0126588. https://doi.org/10.1371/journal.pone.0126588

Ma, W. J. (2018). Problematic usage of the Zhang and Luck mixture model. BioRxiv. https://doi.org/10.1101/268961

Mathôt, S., Dalmaijer, E., Grainger, J., & Van der Stigchel, S. (2014). The pupillary light response reflects exogenous attention and inhibition of return. Journal of Vision. https://doi.org/10.1167/14.14.7

McGinley, M. J., Vinck, M., Reimer, J., Batista-Brito, R., Zagha, E., Cadwell, C. R., … McCormick, D. A. (2015). Waking State: Rapid Variations Modulate Neural and Behavioral Responses. Neuron. https://doi.org/10.1016/j.neuron.2015.09.012

Schurgin, M. W., Wixted, J. T., & Brady, T. F. (2018). Psychophysical Scaling Reveals a Unified Theory of Visual Memory Strength. BioRxiv. https://doi.org/10.1101/325472

Tkacz-Domb, S., & Yeshurun, Y. (2018). The size of the attentional window when measured by the pupillary response to light. Scientific Reports. https://doi.org/10.1038/s41598-018-30343-7

Van Den Berg, R., Shin, H., Chou, W. C., George, R., & Ma, W. J. (2012). Variability in encoding precision accounts for visual short-term memory limitations. Proceedings of the National Academy of Sciences of the United States of America. https://doi.org/10.1073/pnas.1117465109

Wang, C. A., & Munoz, D. P. (2018). Neural basis of location-specific pupil luminance modulation. Proceedings of the National Academy of Sciences of the United States of America. https://doi.org/10.1073/pnas.1809668115

Wierda, S. M., van Rijn, H., Taatgen, N. a., & Martens, S. (2012). Pupil dilation deconvolution reveals the dynamics of attention at high temporal resolution. Proceedings of the National Academy of Sciences, 109(22), 8456–8460. https://doi.org/10.1073/pnas.1201858109

Yeshurun, Y. (2019). The spatial distribution of attention. Current Opinion in Psychology. https://doi.org/10.1016/j.copsyc.2018.12.008

Zhang, W., & Luck, S. J. (2008). Discrete fixed-resolution representations in visual working memory. Nature. https://doi.org/10.1038/nature06860

GENERAL DISCUSSION

The confidence we have in what we perceive guides us in deciding how to act upon the world. The blurriness, the instantaneity and the uncertainty, these characteristics of our perception all need to be accounted for in our metacognitive reasoning about the world. Yet, what if a part of this uncertainty remained concealed from our judgment? A vast number of studies show that we are not aware of the myriad of sensory and cognitive processes determining our everyday interactions with the environment. As Helmholtz pointed out in his *Treatise on physiological optic*s (1825): "Judgments, including those involving conceptions that are undoubtedly acquired by experience, are also determined directly by physiological agencies in characteristic fashion, and may come to consciousness as something bestowed immediately, complete and obligatory". In this sense, even though we need to reflect on the quality of our perception, attention, the mechanism by which we access and select such percepts may sometimes fail in its process, spinning its own story to our mind, a little white lie. In this dissertation, we have shown that confidence can remain oblivious to these white lies. Confidence overlooks the time it takes for attention to unfold, even when such ignorance is detrimental for the task at hand. In this discussion, we present an overview of our results and expand upon and explore fine-grained details and implications.

1. METACOGNITIVE SENSITIVITY, BIAS, AND PSYCHOPHYSICAL SCALES

In this dissertation, we grounded our studies in the fundamental distinction between bias and sensitivity, and built upon this foundation, drawing conclusions about the temporal structure of attention and confidence. Here, we will set forth the intrinsic interests and some limitations of the methods used in the dissertation (1.1). We will then dive into a closer look at one of our findings that points to what the nature of Type 1 evidence used in Type 2 decisions is like (1.2).

1.1 On the methods used in the present dissertation

A pure measure of how raw confidence and attention correlate does not provide much information as to the nature of the evidence used during Type 2 decision-making because both metacognitive sensitivity and bias are conflated. We saw in the General introduction (Section 3.2.2), the notable absence of true Type 2 analyses distinguishing metacognitive sensitivity from bias in the joint study of attention and confidence. In our studies, we used a number of analyses and methods to tackle this problem. In the first chapter, meta-d' and group averages were used to measure how confidence and metacognitive sensitivity were affected during exogenous cueing. In chapter 3, we considered the distribution of average confidence during a selection episode, as well as the difference in accuracy between high and low confidence trials, a simple approach which allowed us to identify most of the dissociations found during temporal orienting of attention as Type 2 'bias'. We will go into future detail about this finding further in the discussion (Section 2.1 of this discussion). Finally, in chapter 2 and 4, the method used was a combination of continuous report (i.e., a reproduction task) and confidence 2AFC, to predict confidence based on trial-by-trial difference in error. The blend of a Type 1 reproduction task with confidence 2AFC had an intrinsic advantage over the other methods we used: it allows for a model-free investigation of the relationship between confidence and performance, and reduces bias compared to other measures. Importantly, it does not assume, in its essential form, a determined distribution or source for Type 2 evidence. In chapters 2 and 4, we were then able to run a simple correlation analysis on the confidence 2AFC data to evaluate an observer's metacognitive ability.

How do these methodological choices affect, if at all, the conclusions about attention and confidence? The literature presents three broad relationships between attention and confidence (General introduction, Section 3.1): (1) confidence does not take into account attention; (2) confidence decreases with attention; (3) or confidence increases with attention. In our studies, we could not find any evidence for confidence being oblivious to the increase in accuracy induced by exogenous attention (1). The involuntary nature of attentional orienting (Chapter 1) did not change much to the correlation between confidence and accuracy, contrary to what has been previously claimed (Kurtz et al., 2017). Confidence was also frequently updated in accordance to the variability in the attentional state as measured by both metacognitive sensitivity and bias (Chapters 1-4). Moreover, we did not find a consistent decrease (2) or increase in confidence (3). Indeed, our results shed light on a crucial but overlooked aspect: the effect of attention on confidence depends on time. Notably, we observed a decrease in metacognitive sensitivity when attention was misplaced or delayed. This was caused by confidence overlooking the latency of both spatial (Chapter 4) and temporal attention (Chapter 3). Moreover, confidence was also lower at the attentional locus, when multiple stimuli shared the same attentional episode, a phenomenon mostly resulting from metacognitive bias (Chapter 3). In contrast, when attention was not delayed, metacognitive ability was greater at the peak of the attentional episode, and progressively decreased as attention was disengaging (Chapter 2).

At first glance, these results might seem contradictory, favouring both the views that attention increases and decreases confidence, depending on the stimuli, the moment, and the metric (bias versus sensitivity). Yet, these results can be accounted for in one integrated account. In a next section (1.3), we thus propose an integrated account of these seemingly disparate effects, by re-examining the concept of an attentional episode.

1.2 Psychophysical scaling in metacognition

We will first expand upon an unanticipated finding concerning the potential nature of the evidence used during Type-2 judgments. We found that in two different implementations of the confidence 2AFC paradigm, and with two stimulus types (e.g., in Chapter 2, the stimulus is static but not in Chapter 4), the difference in error between two Type 1 decisions strongly predicted confidence judgments. Moreover, the correlation between error difference and confidence was significantly enhanced when taking into account the overall error magnitude within the Type 1 decisions pair (fig. 1, A and B of the General discussion). This improvement was found on the individual participant level in Chapter 2, and at the group level across all tasks in Chapter 4. These findings, and their consistency across tasks, demonstrate that the acuity of confidence in comparing the precision of two responses was affected by the cumulated precision of those responses, an influence which may found its source in the nature of Type 1 evidence made available to confidence, as we shall see below.

A canonical principle in perception, purportedly shared between humans and animals, is the 'universal law of generalization' (Shepard, 1987): "A psychological space is established for any set of stimuli by determining metric distances between the stimuli such that the probability that a response learned to any stimulus will generalize to any other is an invariant monotonic function of the distance between them. To a good approximation, this probability of generalization (i) decays exponentially with this distance, and (ii) does so in accordance with one of two metrics, depending on the relation between the dimensions along which the stimuli vary." (Shepard, 1987). According to Shepard, the greater the perceptual distance between two stimuli, the lower the probability of the two stimuli being members of the same perceptual category. A fundamental implication of the generalisation principle is exponential decay, which is observed with stimuli distance: when two stimuli are both far from the central feature dimension, the greater the possibility to confuse the two from this target category perspective. This implication has recently been used to explain how items are encoded into working memory, offering a much simpler model than existent explanations in literature (Schurgin, Wixted, & Brady, 2019). Large errors made on a stimulus dimension (e.g., orientation) are more equally distributed than small errors because the perceptual difference between them is smaller. This observation stems from the distance in internal feature space: observers represent the stimuli within a psychological continuum that is, contrary to the physical feature space, intrinsically non-linear. In light of this non-linear representation of the word, Type 2 decisions is unlikely to have access to a linearized version of the non-linear evidence used for Type 1 decisions. Thus, our finding of empirical non-linearity observed in metacognition could be partially due to the nature of Type 1 psychological space itself.

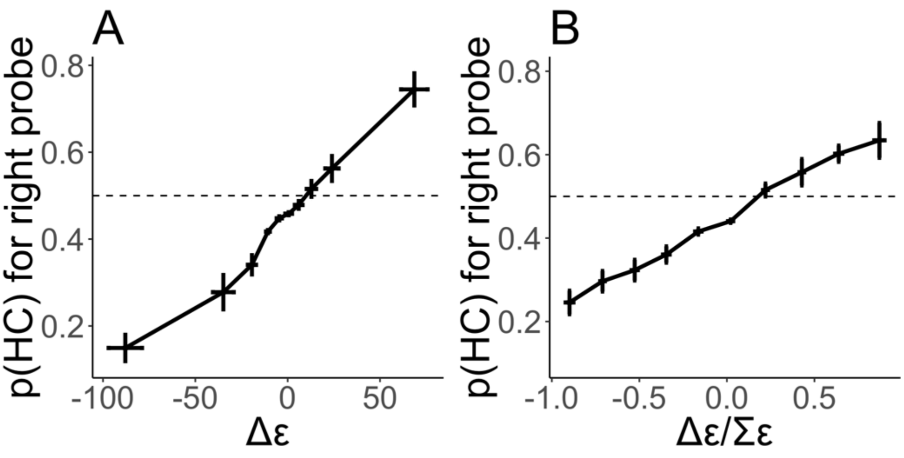

Figure 1. Psychophysical scaling in metacognition. Figures from the experiment in Chapter 2. (A) The probability of selecting one of the two Type 1 decisions (here, the decision about the right probe) during confidence judgments, as a function of the error difference between the two probes. (B) Same as (A), but this time as a function of the scaled error difference between the two probes. Non-linear scaling brings the relationship between confidence and error to near linear shape. (A) and (B) Negative error represents greater error for the probe on the right. For illustration, the errors have been grouped in 10 quantiles. The error bars are within-participant ±1 SEM.

We thus propose that the Type 1 evidence that confidence has access to is non-linear. For instance, let us consider a task, such as the one in Chapter 4, in which an observer has to report the phase of a clock at cue onset (see Chapter 4). Observers are presented with two consecutive trials of the same type, and then have to pick which of the two trials they are the most confident about. The Type 1 decision is of circular Gaussian shape more or less centred on the true phase within each attentional episode (Chapter 4). We assume, for brevity, that the peak of the attentional episode has no delay and is centred on the true phase. The observer's internal generative model thus obeys a circular normal, with greater evidence for the correct phase, and decreasing evidence for phases at larger distance from it, much like the prediction of the generalisation law. The phase to be reported by the participant will be the one with maximum evidence, a phase likely in the vicinity of the correct phase. However, random noise fluctuation applied on each phase candidate along the perceptual space could allow for distant phases to win, though less frequently – and therefore, with less evidence – than phases immediately surrounding the true phase position [1]. There is a direct side-effect of this model: the difference in average evidence between two neighbouring errors decreases the further they are from the distribution peak, making the discrimination between them harder [1]. The confidence 2AFC judgment, in which participants select the Type 1 response they are most confident in, should be, ideally, based on the difference in evidence between those two responses. A psychophysical scaling account of Type 1 evidence space precisely predicts lower metacognitive sensitivity when both the errors in the pair are large, something we observed in all of our confidence 2AFC datasets (fig. 1A versus 1B). This pattern suggests that confidence neatly tracks Type 1 evidence, down to its non-linearity.

The scaled model of confidence is not in itself new. Peirce & Jastrow posited that confidence can be mapped as the log of the Type 1 evidence (1884; see Section 1.2.1 of the General introduction). More recently, van den Berg & Ma proposed a similar mapping for confidence and working memory, where confidence was log-related to the precision of memory encoding (van den Berg et al., 2017). However, this form of Weber-Fechner scaling of confidence in log form is different from our account of generalisation: while these authors proposed that confidence is a log-transform of Type 1 evidence, we suggest that Type 1 evidence distribution is enough to produce non-linearities in certain Type 2 decision spaces. For the moment, this account is only valid for the confidence 2AFC tasks presented in the current thesis, in which two Type 1 decisions have to be compared. Generalising this approach to confidence ratings might prove to be more challenging, but ought to be considered in future experimental and modelling work.

---

[1] An implementation of such a model for stimuli in the time dimension rather than the space dimension (i.e., RSVP), inspired from the Attentional Gating Model (Reeves & Sperling, 1986), is proposed in the Supplementary Material of the Chapter 3.

2. An episodic account of attention and confidence

Intuitively, better accuracy should lead to greater confidence: if attention increases accuracy, confidence should increase as well. Nonetheless, the literature is still divided on this point. As we saw earlier, a potential reason for this division is the absence of a viable account of time in attentional manipulation. We instead defined the temporal aspect of the relationship between attention and confidence as a bound which affects both the bias and the sensitivity of a confidence judgment. This bound is based on the notion of selective episodes from the temporal attention literature (see Section 2.2.2 of the General introduction): orienting attention to a given location initiates a Gaussian-like selection episode, spread over time. Attention is thus considered to have a time-line, with different moments or states: (a) the orienting process, or engagement, during which attention is allocated; (b) the selection, with better encoding quality at the peak; (c) the disengagement, during which attention ceases to be effective. In the following two sections, we will detail how the peak of the selection episode affects confidence differently than its 'boundaries' (i.e., the orienting and disengagement processes, respectively). We will show that a simple understanding of confidence as being constrained by an attentional episode can explain many of the findings reported in the present dissertation.

2.1 When confidence chooses attention rather than accuracy

We will begin our interpretation of how time affects the relationship between attention and confidence, by considering confidence at the peak of the attentional episode. When we consider the peak of this episode, both exogenous and endogenous attention had a positive impact on confidence. In Chapter 1, when cue-target onset asynchrony was used to maximise the effect of exogenous attention, metacognitive ability did not suffer: the early increase in accuracy was followed by an increase in confidence, and the metacognitive ratio (meta-d'/d') remained stable over time. This result confirms that confidence is able to adjust to the effect of involuntary capture of spatial attention, and shows that when attention is oriented exogenously, metacognitive ability is not impaired. The same

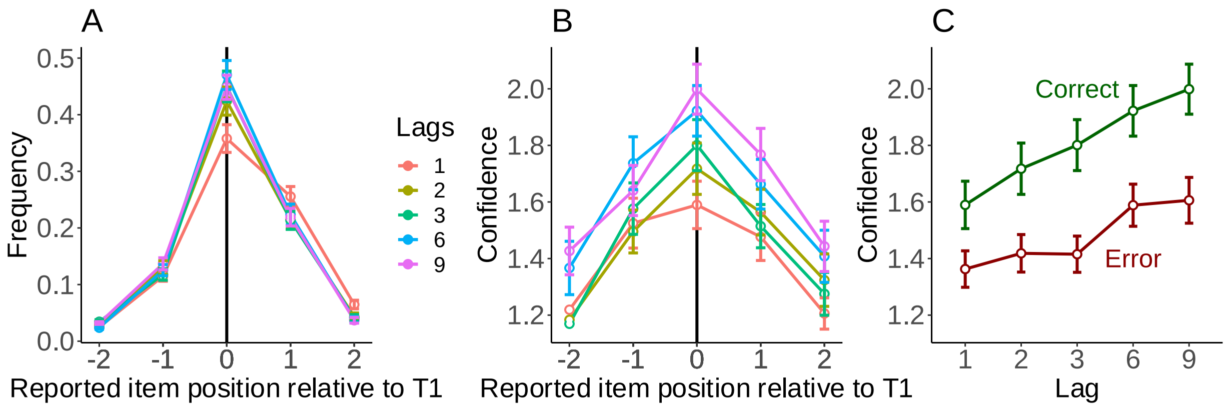

Figure 2. Confidence parallels selection episode in time. Figure reproduced from Chapter 3. (A) The frequency of reports for items around target true position, separately for each lag, when no selection delay occurred. (B) 'Position-based' metacognition. The average confidence per position, for each lag. Confidence decreased monotonically with distance from the target position. This is also true in the case (not depicted here) in which the selection peak is misplaced. (C) 'Error-based' metacognition. The average confidence level for correct responses and errors, which provides an estimate of metacognition. Error bars represent standard error of the mean across participants.

---

was true for endogenous spatial attention: in Chapter 2, confidence was greater in the immediate aftermath of the selection peak, and decreased monotonically after. When we consider the time-line of an attentional episode, the selection peak determines the relationship in three ways: (a) the under-confidence for a second target which is sharing the same 'peak' as a previous target; (b) the strong correlation between the probability of report and average confidence; (c) and the effect of delayed selection on confidence. The peak is thus positive for confidence, and dissecting the attentional time-line allows for an in-depth comprehension of how confidence tracks performance during the stronger period of an attentional episode.

First, we will address the second item under-confidence bias. Despite a very strong bias favouring lower confidence ratings when two targets shared the same attentional episode, we still found greater confidence for correct responses. Therefore, sharing an attentional episode between distinct targets might be a source of confusion mostly at the decision stage, but does not remove metacognitive ability. The observed under-confidence thus points out to the importance of the selection episode in determining confidence bias.

Second, we turn to the strong correlation between the probability of selecting a stimulus (note that this is different from accuracy, which is the probability of selecting the target stimulus) and confidence (fig. 2, A and B). Confidence paralleled the probability of selecting a given item over time, and that bound was also observed when the selection peak was misplaced (fig. 4 of Chapter 3): even when the item at the attentional episode's peak was not the target, confidence was still greater for that item, and decreased for items further away from the attentional episode's peak (even when the participant finally selected the correct, target item, see Chapter 3). This shows that confidence was less sensitive to selection of the correct item (or accuracy) than to the probability of selection (or the effect attention had on the stimuli).

Rather than using an umbrella term of metacognition, we could instead distinguish two sorts of metacognitive abilities that allows for the strong correlation between probability of selection and confidence: error-based metacognition, which is the difference in confidence for errors and correct responses (fig. 2C), and position-based metacognition, which is the significant increase in confidence for the most frequently selected item and decrease for errors further away (fig. 2B). This dichotomy leads us to our last point: the effect of delayed selection on confidence. Even if error-based metacognition remains the only, truly objective measure of metacognitive ability in the present context, it should be noted that position-based metacognition is nevertheless informative. It highlights the strong link metacognition has with temporal attention: when attention was misplaced, confidence gave systematically different weights to errors as if the misplacement never occurred. This inability of confidence to account for delay in temporal attention has also been observed for spatial attention (Chapter 4). For both exogenous and endogenous spatial orienting, confidence fully ignored the latency of attention. Attentional selection is thus major determinant of confidence, ahead of accuracy itself: when attention is efficiently allocated to the correct point in space and time, it does mirror accuracy, but when attention is misplaced, this mirroring will cease to exist, leading to a decrease in metacognitive sensitivity. Looking at the time line thus allows us to dissect the differential effects of the evidence used for Type 2 judgments.

2.2 The 'boundaries' of attentional selection distort metacognition

The positive relationship we found between confidence and the centre of mass of the attentional episode predicts that confidence should be lower at the 'boundaries' of the episode (see fig. 2b), but it does not necessary tell us much about the pattern metacognitive sensitivity ought to follow. In Chapter 2, we used a dual-task paradigm to induce an endogenous attentional episode during a first task, and to probe the residual spatial effect of this episode on a following task. We termed this residual facilitation 'disengagement', as the participant had to (at least partially) disengage from the initial location to succeed in the second task. In this experiment, we observed a decrease in confidence, mirroring the decrease in attentional effect. Importantly, metacognitive ability also decreased for probes further away from the initial target, demonstrating that the attentional episode plays a role in overall metacognitive ability. Thus, it seems that average confidence and metacognitive ability both act in similar ways at the boundaries.

In Chapter 4, we further investigate metacognitive ability by stepping beyond disengagement. On the other end of the spectrum, engagement, or the mechanism which initiates the attentional episode (i.e., the *orienting* of attention). Despite different selection delays in accordance with the time course of attention, response *precision* was not differently affected by the early ('pre-cue' condition) or late ('endogenous' condition) orienting of attention. Supplemental, model-based analyses confirmed the robustness of those precision estimates (see the Supplementary Material in Chapter 4). The equated level of performance in these two conditions allowed us to specifically investigate metacognitive ability while Type 1 precision remained stable. We found late orienting of attention, situations in which the response cue was also the trigger for endogenous attention to orient, to weaken metacognitive ability. This condition, with the onset of attention allocation anchored to phase to be reported, allowed us to probe perceptual decision right after the orienting process occurred. The initial orienting of endogenous attention in space disrupted metacognition to a greater extent than the period of sustained attention that would follow (i.e., the 'pre-cue' condition), showing that *orienting* attention alters metacognition. Once again, this early interaction between attention and confidence provides evidence for the crucial role of attentional episode in shaping metacognitive ability.

2.3 Does metacognitive ability need attention?

In the previous section, we saw that perceptual episodes affected the Type 2 decision process via bias shifts and also moulded metacognition, the very ability of confidence to reflect Type 1 performance. Using the framework of selective perceptual episodes, it seems possible to explain the diversity of empirical results in this dissertation. This 'dynamic' understanding of attention and confidence arose from the direct investigate of systematic temporal manipulations of attention through our paradigms, something that was lacking in the existent literature. Yet, by focusing our work on attentional episodes, we did not employ conditions in which attention was 'truly absent': attention was misallocated (as in Chapter 3), delayed (Chapter 4) and sometimes lowered via invalid location orientation (Chapters 1 and 2). This choice was however not a simple oversight.

Delayed attention could be considered as inattention, at first glance. However, this was not the case. In the present dissertation, we showed that confidence can still discriminate between errors when attention is misallocated, but this attention-based metacognition cannot be measured using standard accuracy-based descriptors (Chapters 3 and 4, and Section 2.1 of the present discussion). Thus, the precise meaning of what an experimenter selects as an accuracy metric for Type 1 is crucial: observing greater confidence for missed targets is one thing, but understanding *why* such Type 2 inconsistency occurs is an important step toward understanding confidence. The ability of confidence to monitor the probability of report suggests that it is likely not a case of complete inattention. The difference however between delayed attention and complete inattention remains to be investigated, thus does not provide information about what happens to metacognition in the near absence of attention.

Though the aspect of inattention was not the subject of the current dissertation, it remains an important aspect to consider. One might be tempted to apply the attentional episode account of confidence to the understanding of inattention, by considering that the 'boundaries' of the episode approximately equivalent to the absence of attention. Such an account would predict a sharp drop of metacognitive ability during inattention. However, it would not reflect inattention *per se* but rather a state of lower, or 'diffused' attention. To probe metacognition during inattention, a paradigm needs to be carefully designed to prevent multiple cofounding factors, outlined in the previous sections (see Section 2.3.1 and 2.3.2 of the General introduction).

In this section, we proposed an episodic account of attention and, notably, of confidence, in which metacognitive ability is moulded by attentional episodes. This view makes attention a crucial purveyor of contrasted Type 1 evidence for Type 2 judgments and underlines the necessity to systematically control for potential attentional effects in confidence experiments.

3. Perspectives and further work

In this discussion, we have put forth an integrative account of the relationship between confidence and performance, by considering time in a way akin to the attention episode. This has allowed us to account for the disarming variety of results found in our studies: confidence following performance, confidence being oblivious to performance, and so on. Yet, this thesis, as any empirical work, entails many subsequent experimental and theoretical questions. In the following and last section of the discussion, we will propose some new avenues to be explored.

3.1 Attentional rhythms and confidence

In recent years, a body of literature investigating cyclicity in perception has re-emerged at the cornerstone of psychophysics and neuroimaging (Harter, 1967; Valera, Toro, Roy John, & Schwartz, 1981). This literature goes beyond purely relying on correlational analyses between behaviour and neural oscillatory patterns, by investigating behavioural oscillations directly (for a review, see VanRullen, 2016). Most of the paradigms in the study of behavioural oscillations capitalise on the effect a salient event, or self-generated motor command, has on perception: it is thought to 'reset' the phase of ongoing brain rhythms, permitting an event-locked phase alignment across trials, conditions and participants. By probing accuracy at different moments after reset, authors have been able to observe oscillatory patterns in performance, with phases of low and high accuracy levels. This method has been successfully applied to unveil oscillations in spatial attention, which has been shown to sample the environment at about 4-7 Hz. In other words, the discrete 'spotlight' of attention would switch from one location to another at a certain speed (Dugué et al., 2015, 2016; Fiebelkorn et al., 2013; Landau & Fries, 2012; Senoussi et al., 2019; VanRullen, Carlson, & Cavanagh, 2007). For example, when endogenous attention has to reorient to a new location following an invalid cue, it still periodically comes back to the initial location, even if the previous location has ceased to be relevant (Dugué et al., 2016; Senoussi et al., 2019). Whether confidence does track those oscillatory patterns is a question we are currently investigating, in a follow-up study based on the paradigm presented in Chapter 2. In this experiment, we will apply spectral and model-based analyses to look for periodicity in both attention and confidence. The results of these investigations will be interesting given the tight bond observed in the previous chapters between attention and confidence: if both are found to oscillate at the same frequency, it would strengthen the view of an episodic account of confidence and attention even further.

3.2 Model-based analysis

The work presented in this thesis was primarily experimental, the aim being to collect empirical data in order to develop a better understanding of the confidence-attention tandem in time. In most of the experiments, we found a tight link between Type 1 and Type 2 decisions, with dissociations between the two occurring for the most part when attention was misallocated. Our data at first glance seem compatible with the view of confidence as grounded in Type 1 evidence and noise. However, most of the modelling work presented in this dissertation was meant to either present how Type 1 evidence fuelled confidence while remaining agnostic to attentional mechanisms (Chapter 3) or to test the precision of the attentional episode without considering confidence (Chapter 4 and its Supplementary Material). There is thus a need for an integrative model-based approach to both attention and confidence, something which was beyond the scope of the current thesis. We hope that the temporal account we have drawn can lay the foundation to a full computational approach.

3.3 Functional and physiological markers

An integrative model of attention and confidence would notably benefit from a better understanding of the joint analysis of their neural correlates. More specifically, the shape and boundaries of attentional episodes make them an interesting candidate for time-resolved analysis in electro and magneto-encephalography. In the scope of this thesis, we looked at a simpler physiological marker, pupillary response, know to reflect central arousal and attentional states (e.g., Binda & Murray, 2015; Hoeks & Levelt, 1993; Mathôt & Van der Stigchel, 2015; Tkacz-Domb & Yeshurun, 2018; Wierda, van Rijn, Taatgen, & Martens, 2012; Yeshurun, 2019). In Chapter 4, we found a specific and very early marker of variability in endogenous attention allocation (fig. 3, B), too early to be attributed to stimulus-evoked pupillary response (fig. 3, A). One interpretation for this early pupillary correlate is that of spontaneous fluctuations in the central arousal state, which would determine attention's ability to engage. Interestingly, this effect of spontaneous activity was not present for sustained attention or exogenous orienting of attention, pointing to the role of the orientation process as a catalyser of the relationship between pupil and error. We, however, did not find a similar signature for confidence, indicating no early potent marker of confidence construction around cue onset (see Supplementary Material of Chapter 4). However, this effect is also consistent with a task-related noise: replacing confidence 2AFC with confidence ratings could potentially provide a continuous Type 2 variable to which the pupillary signal can be correlated. Further work may use this method to tackle confidence more specifically, perhaps revealing an interesting dissociation, or association.

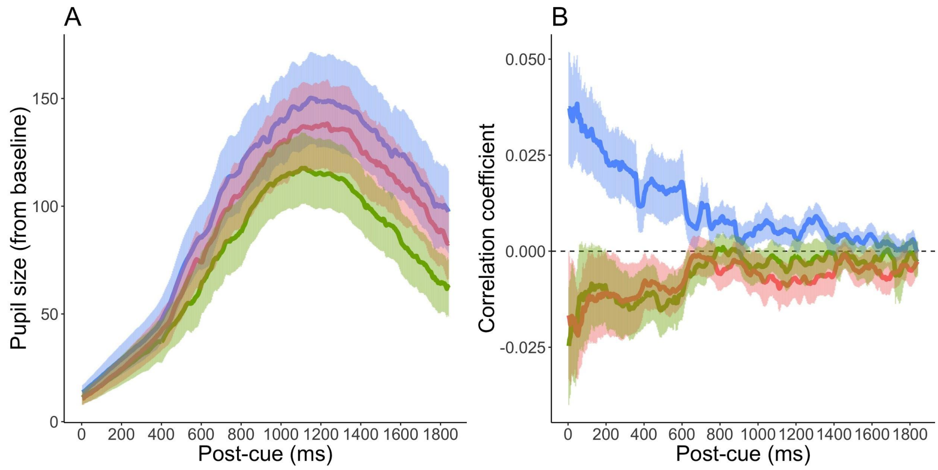

Figure 3. Pupillometry correlates of attentional orienting. Reproduced from the Supplementary Material of Chapter 4. (A) Overall pupillary response. The average, cue-locked and baseline-corrected pupil size as a function of time and attentional condition. Zero on the x-axis represents cue onset. Pre-cue: green curve; exogenous: red curve; endogenous: blue curve. (B) Trial-by-trial analysis. The per-condition correlation coefficient between error magnitude and pupil size for each time bin following cue onset. A positive value indicates that pupil size is positively correlated with error magnitude on a trial-by-trial basis. Light shading represents across participants SEM.

3.4 Confidence in attention or attention in confidence?

Before concluding, we would like to ponder the potential meaning – or the behavioural relevance – of the results presented in this thesis. We showed that metacognition often needs attention to unfold, their tight bound making metacognition incapable of tracking the limits of the attentional system. However, is it necessarily a bad thing? Ignoring discrepancies at such a milliseconds timescale may be beneficial in a real-world setting. Moreover, the implicit (and often explicit) assumption in the present work was that confidence is determined by attention. There is, however, no need for this relation to be unidirectional: the level of perceptual confidence produced across time at one location might well determine the pattern of attention in the near future. For example, the exploration-exploitation trade-off developed initially in the reinforcement learning literature applied to attention (e.g., Ehinger, Kaufhold, & König, 2018; Gottlieb, Oudeyer, Lopes, & Baranes, 2013; Manohar & Husain, 2013) suggests that confidence may well modulate this trade-off in value-based learning (Boldt et al., 2019). For these reasons, the hypothesis of confidence as a determiner of attentional foraging patterns would thus be an interesting perspective for a better understanding of the confidence-attention tandem.

4. CONCLUSION

In this thesis, we capitalised on manipulating the temporal structure of selective attention to study the relation between confidence and accuracy in perceptual decision-making. We found confidence to be highly responsive to the temporal dynamics of selective attention, to the point of dissociating from performance in certain cases. The empirical work presented in this dissertation therefore highlights the importance of selective attention in the construction of visual confidence and contributes to the understanding of the exact nature of the evidence signal used during metacognitive judgments.

****

References

Akyürek, E. G., Eshuis, S. A. H., Nieuwenstein, M. R., Saija, J. D., Başkent, D., & Hommel, B. (2012). Temporal target integration underlies performance at lag 1 in the attentional blink. *Journal of Experimental Psychology: Human Perception and Performance*, *38*(6), 1448–1464. https://doi.org/10.1037/a0027610

Akyürek, E. G., & Hommel, B. (2005). Short-term memory and the attentional blink: Capacity versus content. *Memory and Cognition*. https://doi.org/10.3758/BF03195332

Anderson, B. (2011). There is no such thing as attention. *Frontiers in Psychology*. https://doi.org/10.3389/fpsyg.2011.00246

Anton-Erxleben, K., Abrams, J., & Carrasco, M. (2010). Evaluating comparative and equality judgments in contrast perception: Attention alters appearance. *Journal of Vision*. https://doi.org/10.1167/10.11.6

Anton-Erxleben, K., Abrams, J., & Carrasco, M. (2011). Equality judgments cannot distinguish between attention effects on appearance and criterion: A reply to Schneider (2011). *Journal of Vision*. https://doi.org/10.1167/11.13.8

Baguley, T. (2012). Calculating and graphing within-subject confidence intervals for ANOVA. *Behavior Research Methods*. https://doi.org/10.3758/s13428-011-0123-7

Baldassi, S., Megna, N., & Burr, D. C. (2006). Visual Clutter Causes High-Magnitude Errors. *PLoS Biology*. https://doi.org/10.1371/journal.pbio.0040056

Balsdon, T., & Azzopardi, P. (2015). Absolute and relative blindsight. *Consciousness and Cognition*. https://doi.org/10.1016/j.concog.2014.09.010

Balsdon, T., & Clifford, C. W. G. (2018). Visual processing: Conscious until proven otherwise. *Royal Society Open Science*. https://doi.org/10.1098/rsos.171783

Bang, J. W., Shekhar, M., & Rahnev, D. (2019). Sensory noise increases metacognitive efficiency. *Journal of Experimental Psychology: General*. https://doi.org/10.1037/xge0000511

Barrett, A. B., Dienes, Z., & Seth, A. K. (2013). Measures of metacognition on signal-detection theoretic models. *Psychological Methods*, *18*(4), 535–552. https://doi.org/10.1037/a0033268

Barthelme, S., & Mamassian, P. (2010). Flexible mechanisms underlie the evaluation of visual confidence. *Proceedings of the National Academy of Sciences*, *107*(48), 20834–20839. https://doi.org/10.1073/pnas.1007704107

Barthelmé, S., & Mamassian, P. (2009). Evaluation of Objective Uncertainty in the Visual System. *PLoS Computational Biology*, *5*(9), e1000504. https://doi.org/10.1371/journal.pcbi.1000504

Barttfeld, P., Wicker, B., McAleer, P., Belin, P., Cojan, Y., Graziano, M., … Sigman, M. (2013). Distinct patterns of functional brain connectivity correlate with objective performance and subjective beliefs. *Proceedings of the National Academy of Sciences of the United States of America*. https://doi.org/10.1073/pnas.1301353110

Binda, P., & Murray, S. O. (2015). Keeping a large-pupilled eye on high-level visual processing. *Trends in Cognitive Sciences*. https://doi.org/10.1016/j.tics.2014.11.002

Bogacz, R., Brown, E., Moehlis, J., Holmes, P., & Cohen, J. D. (2006). The physics of optimal decision making: A formal analysis of models of performance in two-alternative forced-choice tasks. *Psychological Review*, *113*(4), 700–765. https://doi.org/10.1037/0033-295X.113.4.700

Boldt, A., Blundell, C., & De Martino, B. (2019). Confidence modulates exploration and exploitation in value-based learning. *Neuroscience of Consciousness*. https://doi.org/10.1093/nc/niz004

Boyer, J. L., Harrison, S., & Ro, T. (2005). Unconscious processing of orientation and color without primary visual cortex. *Proceedings of the National Academy of Sciences of the United States of America*. https://doi.org/10.1073/pnas.0505332102

Brascamp, J. W., Van Boxtel, J. J. A., Knapen, T. H. J., & Blake, R. (2010). A dissociation of attention and awareness in phase-sensitive but not phase-insensitive visual channels. *Journal of Cognitive Neuroscience*. https://doi.org/10.1162/jocn.2009.21397

Breitmeyer, B., & Ogmen, H. (2010). *Visual Masking: Time Slices Through Conscious and Unconscious Vision*. *Visual Masking: Time Slices Through Conscious and Unconscious Vision*. https://doi.org/10.1093/acprof:oso/9780198530671.001.0001

Broadbent, D. E., & Broadbent, M. H. P. (1987). From detection to identification: Response to multiple targets in rapid serial visual presentation. *Perception & Psychophysics*, *42*(2), 105–113. https://doi.org/10.3758/BF03210498

Buschman, T. J., & Kastner, S. (2015). From Behavior to Neural Dynamics: An Integrated Theory of Attention. *Neuron*, *88*(1), 127–144. https://doi.org/10.1016/j.neuron.2015.09.017

Buschman, T. J., & Miller, E. K. (2007). Top-down versus bottom-up control of attention in the prefrontal and posterior parietal cortices. *Science*. https://doi.org/10.1126/science.1138071

Busse, L., Katzner, S., & Treue, S. (2008). Temporal dynamics of neuronal modulation during exogenous and endogenous shifts of visual attention in macaque area MT. *Proceedings of the National Academy of Sciences of the United States of America*. https://doi.org/10.1073/pnas.0707369105

Callahan-Flintoft, C., Holcombe, A. O., & Wyble, B. (2019). A delay in sampling information from temporally autocorrelated visual stimuli. *BioRxiv*, 656850. https://doi.org/10.1101/656850

Carlson, T. A., Hogendoorn, H., & Verstraten, F. A. J. (2006). The speed of visual attention: What time is it? *Journal of Vision*, *6*, 1406–1411. https://doi.org/10.1167/6.12.6

Carrasco, Marisa. (2011). Visual attention: The past 25 years. *Vision Research*. https://doi.org/10.1016/j.visres.2011.04.012

Carrasco, Marisa, & Barbot, A. (2019). Spatial attention alters visual appearance. *Current Opinion in Psychology*. https://doi.org/10.1016/j.copsyc.2018.10.010

Carrasco, Marisa, Ling, S., & Read, S. (2004). Attention alters appearance. *Nature Neuroscience*, *7*(3), 308–313. https://doi.org/10.1167/12.9.1387

Carrasco, Marisa, Penpeci-Talgar, C., & Eckstein, M. (2000). Spatial covert attention increases contrast sensitivity across the CSF: Support for signal enhancement. In *Vision Research*. https://doi.org/10.1016/S0042-6989(00)00024-9

Carrasco, Marissa, Williams, P. E., & Yeshurun, Y. (2002). Covert attention increases spatial resolution with or without masks: Support for signal enhancement. *Journal of Vision*. https://doi.org/10.1167/2.6.4

Certeau, M. de. (2005). *La possession de Loudun*. (Gallimard, Ed.), *La possession de Loudun* (Folio Hist). Paris. https://doi.org/10.14375/np.9782070319138

Chakravarthi, R., & VanRullen, R. (2011). Bullet trains and steam engines: Exogenous attention zips but endogenous attention chugs along. *Journal of Vision*. https://doi.org/10.1167/11.4.1

Chica, A. B., & Lupiáñez, J. (2009). Effects of endogenous and exogenous attention on visual processing: An Inhibition of Return study. *Brain Research*. https://doi.org/10.1016/j.brainres.2009.04.011

Chun, M M. (1997). Temporal binding errors are redistributed by the attentional blink. *Perception & Psychophysics*, *59*(8), 1191–1199. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9401454

Chun, Marvin M., & Potter, M. C. (1995). A Two-Stage Model for Multiple Target Detection in Rapid Serial Visual Presentation. *Journal of Experimental Psychology: Human Perception and Performance*. https://doi.org/10.1037/0096-1523.21.1.109

Clarke, F. R., Birdsall, T. G., & Tanner, W. P. (1959). Two types of roc curves and definitions of parameters. *Journal of the Acoustical Society of America*. https://doi.org/10.1121/1.1907764

Clarke, Frank R. (1960). Confidence Ratings, Second-Choice Responses, and Confusion Matrices in Intelligibility Tests. *The*

Journal of the Acoustical Society of America. https://doi.org/10.1121/1.1907873

Corallo, G., Sackur, J., Dehaene, S., & Sigman, M. (2008). Limits on Introspection. *Psychological Science*. https://doi.org/10.1111/j.1467-9280.2008.02211.x

Corbetta, M., & Shulman, G. L. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nature Reviews Neuroscience*. https://doi.org/10.1038/nrn755

Coull, J. T., Frith, C. D., Büchel, C., & Nobre, A. C. (2000). Orienting attention in time: Behavioural and neuroanatomical distinction between exogenous and endogenous shifts. *Neuropsychologia*. https://doi.org/10.1016/S0028-3932(99)00132-3

Coull, J. T., & Nobre, A. C. (1998). Where and when to pay attention: the neural systems for directing attention to spatial locations and to time intervals as revealed by both PET and fMRI. *Journal of Neuroscience*, *18*(18), 7426–7435. https://doi.org/0270-6474/98/187426-10$05.00/0

Cousineau, D. (2005). Confidence intervals in within-subject designs: A simpler solution to Loftus and Masson's method. *Tutorials in Quantitative Methods for Psychology*. https://doi.org/10.20982/tqmp.01.1.p042

Craston, P., Wyble, B., Chennu, S., & Bowman, H. (2009). The attentional blink reveals serial working memory encoding: Evidence from virtual and human event-related potentials. *Journal of Cognitive Neuroscience*. https://doi.org/10.1162/jocn.2009.21036

Crick, F., & Koch, C. (2003). A framework for consciousness. *Nature Neuroscience*. https://doi.org/10.1038/nn0203-119

Damasio, A. R., Damasio, H., & Chui, H. C. (1980). Neglect following damage to frontal lobe or basal ganglia. *Neuropsychologia*. https://doi.org/10.1016/0028-3932(80)90058-5

Dayan, P., Hinton, G. E., Neal, R. M., & Zemel, R. S. (1995). The Helmholtz machine. *Neural Computation*. https://doi.org/10.1162/neco.1995.7.5.889

de Gardelle, V., Charles, L., & Kouider, S. (2011). Perceptual awareness and categorical representation of faces: Evidence from masked priming. *Consciousness and Cognition*. https://doi.org/10.1016/j.concog.2011.02.001

de Gardelle, V., Kouider, S., & Sackur, J. (2010). An oblique illusion modulated by visibility: Non-monotonic sensory integration in orientation processing. *Journal of Vision*. https://doi.org/10.1167/10.10.6

de Gardelle, V., Le Corre, F., & Mamassian, P. (2016). Confidence as a Common Currency between Vision and Audition. *PloS One*, *11*(1), e0147901. https://doi.org/10.1371/journal.pone.0147901

de Gardelle, V., & Mamassian, P. (2014). Does Confidence Use a Common Currency Across Two Visual Tasks? *Psychological Science*, *25*(6), 1286–1288. https://doi.org/10.1177/0956797614528956

Dell'Acqua, R., Doro, M., Dux, P. E., Losier, T., & Jolicoeur, P. (2016). Enhanced frontal activation underlies sparing from the attentional blink: Evidence from human electrophysiology. *Psychophysiology*. https://doi.org/10.1111/psyp.12618

Denison, R. N., Adler, W. T., Carrasco, M., & Ma, W. J. (2018). Humans incorporate attention-dependent uncertainty into perceptual decisions and confidence. *Proceedings of the National Academy of Sciences*, 201717720. https://doi.org/10.1073/pnas.1717720115

Denison, R. N., Heeger, D. J., & Carrasco, M. (2017). Attention flexibly trades off across points in time. *Psychonomic Bulletin & Review*, *24*(4), 1142–1151. https://doi.org/10.3758/s13423-016-1216-1

Denison, R. N., Yuval-Greenberg, S., & Carrasco, M. (2019). Directing voluntary temporal attention increases fixational stability. *Journal of Neuroscience*. https://doi.org/10.1523/JNEUROSCI.1926-18.2018

Desender, K., Boldt, A., & Yeung, N. (2018). Subjective Confidence Predicts Information Seeking in Decision Making. *Psychological Science*. https://doi.org/10.1177/0956797617744771

Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual Review of Neuroscience*, *18*, 193–222. https://doi.org/10.1146/annurev.ne.18.030195.001205

Dugué, L., McLelland, D., Lajous, M., & VanRullen, R. (2015). Attention searches nonuniformly in space and in time. *Proceedings of the National Academy of Sciences*, *112*(49), 15214–15219. https://doi.org/10.1073/pnas.1511331112

Dugué, L., Merriam, E. P., Heeger, D. J., & Carrasco, M. (2018). Endogenous and exogenous attention distinctly modulate fMRI activity in visual cortex. *BioRxiv*.

Dugué, L., Roberts, M., & Carrasco, M. (2016). Attention reorients periodically. Current Biology, 26(12), 1595-1601. https://doi.org/10.1016/j.cub.2016.04.046

Duncan, J., Ward, R., & Shapiro, K. (1994). Direct measurement of attentional dwell time in human vision. *Nature*. https://doi.org/10.1038/369313a0

Dux, P. E., & Marois, R. (2009). The attentional blink: a review of data and theory. *Atten Percept Psychophys*, *71*(8), 1683–1700. https://doi.org/10.3758/APP.71.8.1683

Eckstein, M. P., Thomas, J. P., Palmer, J., & Shimozaki, S. S. (2000). A signal detection model predicts the effects of set size on visual search accuracy for feature, conjunction, triple conjunction, and disjunction displays. *Perception and Psychophysics*. https://doi.org/10.3758/BF03212096

Ehinger, B. V., Kaufhold, L., & König, P. (2018). Probing the temporal dynamics of the exploration- exploitation dilemma of eye movements. *Journal of Vision*. https://doi.org/10.1167/18.3.6

Engbert, R., & Kliegl, R. (2003). Microsaccades uncover the orientation of covert attention. *Vision Research*. https://doi.org/10.1016/S0042-6989(03)00084-1

Eriksen, B. A., & Eriksen, C. W. (1974). Effects of noise letters upon the identification of a target letter in a nonsearch task. *Perception & Psychophysics*. https://doi.org/10.3758/BF03203267

Eriksen, C. W., & Murphy, T. D. (1987). Movement of attentional focus across the visual field: A critical look at the evidence. *Perception & Psychophysics*. https://doi.org/10.3758/BF03203082

Eriksen, C. W., & St. James, J. D. (1986). Visual attention within and around the field of focal attention: A zoom lens model. *Perception & Psychophysics*. https://doi.org/10.3758/BF03211502

Eriksen, C. W., & Yeh, Y. Y. (1985). Allocation of Attention in the Visual Field. *Journal of Experimental Psychology: Human Perception and Performance*. https://doi.org/10.1037/0096-1523.11.5.583

Fiebelkorn, I. C., Saalmann, Y. B., & Kastner, S. (2013). Rhythmic sampling within and between objects despite sustained attention at a cued location. *Current Biology*, *23*(24), 2553–2558. https://doi.org/10.1016/j.cub.2013.10.063

Fleming, S. M., Ryu, J., Golfinos, J. G., & Blackmon, K. E. (2014). Domain-specific impairment in metacognitive accuracy following anterior prefrontal lesions. *Brain*, awu221-. https://doi.org/10.1093/brain/awu221

Fleming, Stephen M., & Lau, H. C. (2014). How to measure metacognition. *Frontiers in Human Neuroscience*, *8*(July), 1–9. https://doi.org/10.3389/fnhum.2014.00443

Fleming, Stephen M., Maniscalco, B., Ko, Y., Amendi, N., Ro, T., & Lau, H. (2015). Action-Specific Disruption of Perceptual Confidence. *Psychological Science*. https://doi.org/10.1177/0956797614557697

Fleming, Stephen M., Ryu, J., Golfinos, J. G., & Blackmon, K. E. (2014). Domain-specific impairment in metacognitive accuracy following anterior prefrontal lesions. *Brain*. https://doi.org/10.1093/brain/awu221

Fleming, Stephen M., Van Der Putten, E. J., & Daw, N. D. (2018). Neural mediators of changes of mind about perceptual decisions. *Nature Neuroscience*. https://doi.org/10.1038/s41593-018-0104-6

Fleming, Stephen M, & Daw, N. D. (2017). Self-evaluation of decision performance: A general Bayesian framework for metacognitive computation. *Psychological Review*, *124*(1), 1–59. https://doi.org/10.1002/smll.))

Fleming, Stephen M, & Dolan, R. J. (2012). The neural basis of metacognitive ability. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *367*(1594), 1338–1349. https://doi.org/10.1098/rstb.2011.0417

Fleming, Stephen M, Weil, R. S., Nagy, Z., Dolan, R. J., & Rees, G. (2010). Relating introspective accuracy to individual differences in brain structure. *Science (New York, N.Y.)*, *329*(5998), 1541–1543. https://doi.org/10.1126/science.1191883

Fries, P. (2015). Rhythms for Cognition: Communication through Coherence. *Neuron*, *88*(1), 220–235. https://doi.org/10.1016/j.neuron.2015.09.034

Gajdos, T., Fleming, S. M., Saez Garcia, M., Weindel, G., & Davranche, K. (2019). Revealing subthreshold motor contributions to perceptual confidence. *Neuroscience of Consciousness*. https://doi.org/10.1093/nc/niz001

Galvin, S. J., Podd, J. V, Drga, V., & Whitmore, J. (2003). Type 2 tasks in the theory of signal detectability: discrimination between correct and incorrect decisions. *Psychonomic Bulletin & Review*, *10*(4), 843–876. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/15000533

Giordano, A. M., McElree, B., & Carrasco, M. (2009). On the automaticity and flexibility of covert attention: A speed-accuracy trade-off analysis. *Journal of Vision*. https://doi.org/10.1167/9.3.30

Gobell, J., & Carrasco, M. (2005). Attention alters the appearance of spatial frequency and gap size. *Psychological Science*, *16*(8), 644–651. https://doi.org/10.1111/j.1467-9280.2005.01588.x

Golla, H., Ignashchenkova, A., Haarmeier, T., & Thier, P. (2004). Improvement of visual acuity by spatial cueing: A comparative study in human and non-human primates. *Vision Research*. https://doi.org/10.1016/j.visres.2004.01.009

Goodbourn, P. T., Martini, P., Barnett-Cowan, M., Harris, I. M., Livesey, E. J., & Holcombe, A. O. (2016). Reconsidering Temporal Selection in the Attentional Blink. *Psychological Science*, *27*(8), 1146–1156. https://doi.org/10.1177/0956797616654131

Gorea, A., & Sagi, D. (2000). Failure to handle more than one internal representation in visual detection tasks. *Proceedings of the National Academy of Sciences of the United States of America*. https://doi.org/10.1073/pnas.97.22.12380

Gottlieb, J., Oudeyer, P. Y., Lopes, M., & Baranes, A. (2013). Information-seeking, curiosity, and attention: Computational and neural mechanisms. *Trends in Cognitive Sciences*. https://doi.org/10.1016/j.tics.2013.09.001

Graziano, M., & Sigman, M. (2009). The Spatial and Temporal Construction of Confidence in the Visual Scene. *PLoS ONE*, *4*(3), e4909. https://doi.org/10.1371/journal.pone.0004909

Green, D. G., & Swets, J. A. (1966). *Signal detection theory and psychophysics. Wiley & Sons, Inc.* https://doi.org/10.1901/jeab.1969.12-475

Griffin, I. C., & Nobre, A. C. (2003). Orienting Attention to Locations in Internal Representations. *Journal of Cognitive Neuroscience*. https://doi.org/10.1162/089892903322598139

Guggenmos, M., Wilbertz, G., Hebart, M. N., & Sterzer, P. (2016). Mesolimbic confidence signals guide perceptual learning in the absence of external feedback. *ELife*. https://doi.org/10.7554/eLife.13388

Hafed, Z. M., & Ignashchenkova, A. (2013). On the dissociation between microsaccade rate and direction after peripheral cues: microsaccadic inhibition revisited. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *33*(41), 16220–16235. https://doi.org/10.1523/JNEUROSCI.2240-13.2013

Hainguerlot, M., Vergnaud, J. C., & De Gardelle, V. (2018). Metacognitive ability predicts learning cue-stimulus associations in the absence of external feedback. *Scientific Reports*. https://doi.org/10.1038/s41598-018-23936-9

HARTER, M. R. (1967). EXCITABILITY CYCLES AND CORTICAL SCANNING: A REVIEW OF TWO HYPOTHESES OF CENTRAL INTERMITTENCY IN PERCEPTION. *Psychological Bulletin*. https://doi.org/10.1037/h0024725

Heilman, K. M., & Valenstein, E. (1979). Mechanisms underlying hemispatial neglect. *Annals of Neurology*, *5*(2), 166–170. https://doi.org/10.1002/ana.410050210

von Helmholtz, H. (1825). Treatise on physiological optics: The perception of vision.

Herrmann, K., Montaser-Kouhsari, L., Carrasco, M., & Heeger, D. J. (2010). When size matters: Attention affects performance by contrast or response gain. *Nature Neuroscience*. https://doi.org/10.1038/nn.2669

Hoeks, B., & Levelt, W. J. M. (1993). Pupillary dilation as a measure of attention: a quantitative system analysis. *Behavior Research Methods, Instruments, & Computers*. https://doi.org/10.3758/BF03204445

Hogendoorn, H., Carlson, T. A., VanRullen, R., & Verstraten, F. A. J. (2010). Timing divided attention. *Attention, Perception, and Psychophysics*, *72*(8), 2059–2068. https://doi.org/10.3758/APP.72.8.2059

Hommel, B., & Akyürek, E. G. (2005). Lag-1 sparing in the attentional blink: Benefits and costs of integrating two events into a single episode. *Quarterly Journal of Experimental Psychology Section A: Human Experimental Psychology*, *58*(8), 1415–1433. https://doi.org/10.1080/02724980443000647

Hopfinger, J. B., & West, V. M. (2006). Interactions between endogenous and exogenous attention on cortical visual processing. *NeuroImage*. https://doi.org/10.1016/j.neuroimage.2005.12.049

James, W. (1890). The principles of psychology (Vols. 1 & 2). *New York Holt*.

Jannati, A., & Di Lollo, V. (2012). Relative blindsight arises from a criterion confound in metacontrast masking: Implications for theories of consciousness. *Consciousness and Cognition*. https://doi.org/10.1016/j.concog.2011.10.003

Jovanovic, L., & Mamassian, P. (2019). When an Event Is Perceived Depends on Where We Attend. *I-Perception*. https://doi.org/10.1177/2041669519858096

Kanai, R., Tsuchiya, N., & Verstraten, F. A. J. (2006). The Scope and Limits of Top-Down Attention in Unconscious Visual Processing. *Current Biology*. https://doi.org/10.1016/j.cub.2006.10.001

Karnath, H. O., Ferber, S., & Himmelbach, M. (2001). Spatial awareness is a function of the temporal not the posterior parietal lobe. *Nature*. https://doi.org/10.1038/35082075

Kastner, S., & Ungerleider, L. G. (2000). Mechanisms of visual attention in the human cortex. *Annual Review of Neuroscience*, *23*, 315–341. https://doi.org/10.1146/annurev.neuro.23.1.315

Kevin O'Regan, J., Deubel, H., Clark, J. J., & Rensink, R. A. (2000). Picture changes during blinks: Looking without seeing and seeing without looking. *Visual Cognition*. https://doi.org/10.1080/135062800394766

Kiani, R., Corthell, L., Shadlen, M. N., Adelson, E. H., Bergen, J. R., Audley, R. J., … Sigman, M. (2014). Choice Certainty Is Informed by Both Evidence and Decision Time. *Neuron*, *84*(6), 1329–1342. https://doi.org/10.1016/j.neuron.2014.12.015

Kiani, R., & Shadlen, M. N. (2009). Representation of confidence associated with a decision by neurons in the parietal cortex. *Science (New York, N.Y.)*, *324*(5928), 759–764. https://doi.org/10.1126/science.1169405

Klein, R. M., & MacInnes, W. J. (1999). Inhibition of return is a foraging facilitator in visual search. *Psychological Science*. https://doi.org/10.1111/1467-9280.00166

Koizumi, A., Maniscalco, B., & Lau, H. (2015). Does perceptual confidence facilitate cognitive control? *Attention, Perception, and Psychophysics*. https://doi.org/10.3758/s13414-015-0843-3

Komura, Y., Nikkuni, A., Hirashima, N., Uetake, T., & Miyamoto, A. (2013). Responses of pulvinar neurons reflect a subject's confidence in visual categorization. *Nature Neuroscience*. https://doi.org/10.1038/nn.3393

Kurtz, P., Shapcott, K. A., Kaiser, J., Schmiedt, J. T., & Schmid, M. C. (2017). The Influence of Endogenous and Exogenous Spatial Attention on Decision Confidence. *Scientific Reports*, *7*(1). https://doi.org/10.1038/s41598-017-06715-w

Landau, A. N., & Fries, P. (2012). Attention samples stimuli rhythmically. *Current Biology*, *22*(11), 1000–1004. https://doi.org/10.1016/j.cub.2012.03.054

Landau, A. N., Schreyer, H. M., Pelt, S. Van, Fries, P., Landau, A. N., Schreyer, H. M., … Fries, P. (2015). Distributed attention is implemented through theta-rhythmic gamma modulation Distributed attention is implemented through theta-rhythmic gamma modulation.

Lau, H. C., & Passingham, R. E. (2006). Relative blindsight in normal observers and the neural correlate of visual consciousness. *Proceedings of the National Academy of Sciences*, *103*(49), 18763–18768. https://doi.org/10.1073/pnas.0607716103

Lawrence, M. A., & Klein, R. M. (2013). Isolating exogenous and endogenous modes of temporal attention. *Journal of Experimental Psychology: General*. https://doi.org/10.1037/a0029023

Leahey, T. H. (1981). A History of Psychology: Main Currents in Psychological Thought. *The American Journal of Psychology*. https://doi.org/10.2307/1422357

Leh, S. E., Johansen-Berg, H., & Ptito, A. (2006). Unconscious vision: New insights into the neuronal correlate of blindsight using diffusion tractography. *Brain*. https://doi.org/10.1093/brain/awl111

Leibnitz, G. W. (1765). *Nouveaux Essais sur l'entendement humain*.

Liu, T., Abrams, J., & Carrasco, M. (2009). Voluntary attention enhances contrast appearance. *Psychological Science*. https://doi.org/10.1111/j.1467-9280.2009.02300.x

Lovejoy, L. P., & Krauzlis, R. J. (2010). Inactivation of primate superior colliculus impairs covert selection of signals for perceptual judgments. *Nature Neuroscience*. https://doi.org/10.1038/nn.2470

Lu, Z. L., & Dosher, B. A. (1998). External noise distinguishes attention mechanisms. *Vision Research*. https://doi.org/10.1016/S0042-6989(97)00273-3

Lu, Z. L., Lesmes, L. A., & Dosher, B. A. (2002). Spatial attention excludes external noise at the target location. *Journal of Vision*. https://doi.org/10.1167/2.4.4

Lupiáñez, J., Klein, R. M., & Bartolomeo, P. (2006). Inhibition of return: Twenty years after. *Cognitive Neuropsychology*. https://doi.org/10.1080/02643290600588095

Macmillan, N. a, & Creelman, C. D. (2005). *Detection Theory: A User's Guide*. *Detection theory: A user's guide (2nd ed.).*

Mamassian, P. (2016). Visual Confidence. *Annual Review of Vision Science*, *2*(1), 459–481. https://doi.org/10.1146/annurev-vision-111815-114630

Maniscalco, B., & Lau, H. (2012). A signal detection theoretic approach for estimating metacognitive sensitivity from confidence ratings. *Consciousness and Cognition*, *21*(1), 422–430. https://doi.org/10.1016/j.concog.2011.09.021

Maniscalco, B., Peters, M. A. K., & Lau, H. (2016). Heuristic use of perceptual evidence leads to dissociation between performance and metacognitive sensitivity. *Attention, Perception, and Psychophysics*. https://doi.org/10.3758/s13414-016-1059-x

Manohar, S. G., & Husain, M. (2013). Attention as foraging for information and value. *Frontiers in Human Neuroscience*. https://doi.org/10.3389/fnhum.2013.00711

Martens, S., & Wyble, B. (2010). The attentional blink: Past, present, and future of a blind spot in perceptual awareness. *Neuroscience and Biobehavioral Reviews*. https://doi.org/10.1016/j.neubiorev.2009.12.005

Marti, S., Sackur, J., Sigman, M., & Dehaene, S. (2010). Mapping introspection's blind spot: Reconstruction of dual-task phenomenology using quantified introspection. *Cognition*. https://doi.org/10.1016/j.cognition.2010.01.003

Marti, S., Sigman, M., & Dehaene, S. (2012). A shared cortical bottleneck underlying attentional blink and psychological refractory period. *NeuroImage*. https://doi.org/10.1016/j.neuroimage.2011.09.063

Martini, P. (2012). Sources of bias and uncertainty in a visual temporal individuation task. *Attention, Perception, and Psychophysics*, *75*(1), 168–181. https://doi.org/10.3758/s13414-012-0384-y

Masson, M. E. J., & Rotello, C. M. (2009). Sources of bias in the Goodman–Kruskal gamma coefficient measure of association: Implications for studies of metacognitive processes. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*(2), 509–527. https://doi.org/10.1037/a0014876

Massoni, S., Gajdos, T., & Vergnaud, J.-C. (2014). Confidence measurement in the light of signal detection theory. *Frontiers in*

Psychology, *5*, 1455. https://doi.org/10.3389/fpsyg.2014.01455

Mathôt, S., & Van der Stigchel, S. (2015). New Light on the Mind's Eye: The Pupillary Light Response as Active Vision. *Current Directions in Psychological Science*. https://doi.org/10.1177/0963721415593725

Maunsell, J. H. R., & Treue, S. (2006). Feature-based attention in visual cortex. *Trends in Neurosciences*. https://doi.org/10.1016/j.tins.2006.04.001

McAdams, C. J., & Maunsell, J. H. (1999). Effects of attention on orientation-tuning functions of single neurons in macaque cortical area V4. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *19*(1), 431–441. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9870971

McConkie, G. W., & Currie, C. B. (1996). Visual Stability Across Saccades while Viewing Complex Pictures. *Journal of Experimental Psychology: Human Perception and Performance*. https://doi.org/10.1037/0096-1523.22.3.563

McCormick, C. R., Redden, R. S., Lawrence, M. A., & Klein, R. M. (2018). The independence of endogenous and exogenous temporal attention. *Attention, Perception, and Psychophysics*. https://doi.org/10.3758/s13414-018-1575-y

Meyniel, F., Sigman, M., & Mainen, Z. F. (2015). Confidence as Bayesian Probability: From Neural Origins to Behavior. *Neuron*. https://doi.org/10.1016/j.neuron.2015.09.039

Montagna, B., & Carrasco, M. (2006). Transient covert attention and the perceived rate of flicker. *Journal of Vision*. https://doi.org/10.1167/6.9.8

Montagna, B., Pestilli, F., & Carrasco, M. (2009). Attention trades off spatial acuity. *Vision Research*, *49*(7), 735–745. https://doi.org/10.1016/j.visres.2009.02.001

Moran, J., & Desimone, R. (1985). Selective attention gates visual processing in the extrastriate cortex. *Science*. https://doi.org/10.1126/science.4023713

Morey, R. D. (2008). Confidence Intervals from Normalized Data: A correction to Cousineau (2005). *Tutorials in Quantitative Methods for Psychology*. https://doi.org/10.20982/tqmp.04.2.p061

Morgan, M. J., Ward, R. M., & Castet, E. (1998). Visual Search for a Tilted Target: Tests of Spatial Uncertainty Models. *Quarterly Journal of Experimental Psychology Section A: Human Experimental Psychology*. https://doi.org/10.1080/713755766

Morillon, B., Schroeder, C. E., & Wyart, V. (2014). Motor contributions to the temporal precision of auditory attention. *Nature Communications*. https://doi.org/10.1038/ncomms6255

Morrone, M. C., Denti, V., & Spinelli, D. (2002). Color and luminance contrasts attract independent attention. *Current Biology*. https://doi.org/10.1016/S0960-9822(02)00921-1

Mort, D. J., Malhotra, P., Mannan, S. K., Rorden, C., Pambakian, A., Kennard, C., & Husain, M. (2003). The anatomy of visual neglect. *Brain*. https://doi.org/10.1093/brain/awg200

Mulckhuyse, M., & Theeuwes, J. (2010). Unconscious attentional orienting to exogenous cues: A review of the literature. *Acta Psychologica*. https://doi.org/10.1016/j.actpsy.2010.03.002

Muller, T., & Nobre, A. C. (2014). Perceiving the passage of time: Neural possibilities. *Annals of the New York Academy of Sciences*. https://doi.org/10.1111/nyas.12545

Naccache, L., Blandin, E., & Dehaene, S. (2002). Unconscious masked priming depends on temporal attention. *Psychological Science*. https://doi.org/10.1111/1467-9280.00474

Nakayama, K., & Martini, P. (2011). Situating visual search. *Vision Research*. https://doi.org/10.1016/j.visres.2010.09.003

Nelson, T. O. (1984). A comparison of current measures of the accuracy of feeling-of-knowing predictions. *Psychological Bulletin*, *95*(1), 109–133. https://doi.org/10.1037/0033-2909.95.1.109

Newby, E. A., & Rock, I. (1998). Inattentional blindness as a function of proximity to the focus of attention. *Perception*.

https://doi.org/10.1068/p271025

Nobre, Anna C., & van Ede, F. (2017). Anticipated moments: temporal structure in attention. *Nature Reviews Neuroscience*. https://doi.org/10.1038/nrn.2017.141

Nobre, Anna Christina, & Rohenkohl, G. (2014). Time for the Fourth Dimension in Attention. *The Oxford Handbook of Attention*. https://doi.org/10.1093/oxfordhb/9780199675111.001.0001

O'Regan, J. K., Rensink, R. A., & Clark, J. J. (1999). Change-blindness as a result of "mudsplashes." *Nature*. https://doi.org/10.1038/17953

Peelen, M. V., Heslenfeld, D. J., & Theeuwes, J. (2004). Endogenous and exogenous attention shifts are mediated by the same large-scale neural network. *NeuroImage*. https://doi.org/10.1016/j.neuroimage.2004.01.044

Peirce, C., & Jastrow, J. (1884). On small differences in sensation. Retrieved from https://philarchive.org/archive/PEIOSD

Peirce, J. W. (2007). PsychoPy-Psychophysics software in Python. *Journal of Neuroscience Methods*, *162*(1–2), 8–13. https://doi.org/10.1016/j.jneumeth.2006.11.017

Persaud, N., McLeod, P., & Cowey, A. (2007). Post-decision wagering objectively measures awareness. *Nature Neuroscience*, *10*(2), 257–261. https://doi.org/10.1038/nn1840

Pestilli, F., & Carrasco, M. (2005). Attention enhances contrast sensitivity at cued and impairs it at uncued locations. *Vision Research*. https://doi.org/10.1016/j.visres.2005.01.019

Peters, M. A. K., Thesen, T., Ko, Y. D., Maniscalco, B., Carlson, C., Davidson, M., … Lau, H. (2017). Perceptual confidence neglects decision-incongruent evidence in the brain. *Nature Human Behaviour*. https://doi.org/10.1038/s41562-017-0139

Pincham, H. L., Bowman, H., & Szucs, D. (2016). The experiential blink: Mapping the cost of working memory encoding onto conscious perception in the attentional blink. *Cortex*, *81*, 35–49. https://doi.org/10.1016/J.CORTEX.2016.04.007

Pleskac, T. J., & Busemeyer, J. R. (2010). Two-stage dynamic signal detection: A theory of choice, decision time, and confidence. *Psychological Review*. https://doi.org/10.1037/a0019737

Posner, M. I. (1980). Orienting of attention. *Quarterly Journal of Experimental Psychology*. https://doi.org/10.1080/00335558008248231

Posner, M. I., Rafal, R. D., Choate, L. S., & Vaughan, J. (1985). Inhibition of Return: Neural Basis and Function. *Cognitive Neuropsychology*. https://doi.org/10.1080/02643298508252866

Pouget, A., Drugowitsch, J., & Kepecs, A. (2016). Confidence and certainty: Distinct probabilistic quantities for different goals. *Nature Neuroscience*. https://doi.org/10.1038/nn.4240

Prinzmetal, W., McCool, C., & Park, S. (2005). Attention: Reaction time and accuracy reveal different mechanisms. *Journal of Experimental Psychology: General*. https://doi.org/10.1037/0096-3445.134.1.73

Ptak, R. (2012). The frontoparietal attention network of the human brain: Action, saliency, and a priority map of the environment. *Neuroscientist*. https://doi.org/10.1177/1073858411409051

Raab, D. H. (1962). DIVISION OF PSYCHOLOGY: STATISTICAL FACILITATION OF SIMPLE REACTION TIMES*. *Transactions of the New York Academy of Sciences*, *24*(5 Series II), 574–590. https://doi.org/10.1111/j.2164-0947.1962.tb01433.x

Rahnev, D., Bahdo, L., de Lange, F. P., & Lau, H. (2012). Prestimulus hemodynamic activity in dorsal attention network is negatively associated with decision confidence in visual perception. *Journal of Neurophysiology*, *108*(5), 1529–1536. https://doi.org/10.1152/jn.00184.2012

Rahnev, D., Koizumi, A., McCurdy, L. Y., D'Esposito, M., & Lau, H. (2015). Confidence Leak in Perceptual Decision Making. *Psychological Science*. https://doi.org/10.1177/0956797615595037

Rahnev, D., Maniscalco, B., Graves, T., Huang, E., de Lange, F. P., & Lau, H. (2011). Attention induces conservative subjective biases in visual perception. *Nature Neuroscience*, *14*(12), 1513–1515. https://doi.org/10.1038/nn.2948

Rahnev, D., Maniscalco, B., Luber, B., Lau, H., & Lisanby, S. H. (2012). Direct injection of noise to the visual cortex decreases accuracy but increases decision confidence. *Journal of Neurophysiology*. https://doi.org/10.1152/jn.00985.2011

Rahnev, Dobromir, & Denison, R. N. (2018). Suboptimality in perceptual decision making. *Behavioral and Brain Sciences*, *41*, e223. https://doi.org/10.1017/S0140525X18000936

Raymond, J. E., Shapiro, K. L., & Arnell, K. M. (1992). Temporary Suppression of Visual Processing in an RSVP Task: An Attentional Blink? *Journal of Experimental Psychology: Human Perception and Performance*, *18*(3), 849–860. Retrieved from https://psych.hanover.edu/classes/cognition/papers/raymond et al 1992 attn blink.pdf

Recht, S., Mamassian, P., & de Gardelle, V. (submitted). Orienting spatial attention weakens metacognition.

Recht, Samuel, de Gardelle, V., & Mamassian, P. (2017). Exogenous cues and visual confidence. *Journal of Vision*, *17*(10), 670. https://doi.org/10.1167/17.10.670

Recht, Samuel, Mamassian, P., & de Gardelle, V. (2019). Temporal attention causes systematic biases in visual confidence. *Scientific Reports*, *9*(1), 11622. https://doi.org/10.1038/s41598-019-48063-x

Reeves, A., & Sperling, G. (1986). Attention gating in short-term visual memory. *Psychological Review*. https://doi.org/10.1037/0033-295X.93.2.180

Rensink, R. A., O'Regan, J. K., & Clark, J. J. (1997). To see or not to see: The Need for Attention to Perceive Changes in Scenes. *Psychological Science*. https://doi.org/10.1111/j.1467-9280.1997.tb00427.x

Rensink, R. A., O'Regan, J. K., & Clark, J. J. (2000). On the failure to detect changes in scenes across brief interruptions. *Visual Cognition*. https://doi.org/10.1080/135062800394720

Reyes, G., & Sackur, J. (2014). Introspection during visual search. *Consciousness and Cognition*. https://doi.org/10.1016/j.concog.2014.08.009

Reyes, G., & Sackur, J. (2017). Introspective access to implicit shifts of attention. *Consciousness and Cognition*, *48*, 11–20. https://doi.org/10.1016/j.concog.2016.10.003

Reynolds, J. H., & Heeger, D. J. (2009). The Normalization Model of Attention. *Neuron*. https://doi.org/10.1016/j.neuron.2009.01.002

Reynolds, J. H., Pasternak, T., & Desimone, R. (2000). Attention Increases Sensitivity of V4 Neurons. *Neuron*, *26*(3), 703–714. https://doi.org/10.1016/S0896-6273(00)81206-4

Rohenkohl, G., Coull, J. T., & Nobre, A. C. (2011). Behavioural dissociation between exogenous and endogenous temporal orienting of attention. *PLoS ONE*, *6*(1). https://doi.org/10.1371/journal.pone.0014620

Rosenthal, D. (2018). Consciousness and confidence. *Neuropsychologia*. https://doi.org/10.1016/j.neuropsychologia.2018.01.018

Ruff, C. C., Kristjánsson, Á., & Driver, J. (2007). Readout from iconic memory and selective spatial attention involve similar neural processes. *Psychological Science*. https://doi.org/10.1111/j.1467-9280.2007.01998.x

Sara, S. J., & Bouret, S. (2012). Orienting and Reorienting: The Locus Coeruleus Mediates Cognition through Arousal. *Neuron*. https://doi.org/10.1016/j.neuron.2012.09.011

Schneider, K. A. (2011). Attention alters decision criteria but not appearance: A reanalysis of Anton-Erxleben, Abrams, and Carrasco (2010). *Journal of Vision*. https://doi.org/10.1167/11.13.7

Schneider, K. A., & Komlos, M. (2008). Attention biases decisions but does not alter appearance. *Journal of Vision*. https://doi.org/10.1167/8.15.3

Schoenherr, J. R., Leth-Steensen, C., & Petrusic, W. M. (2010). Selective attention and subjective confidence calibration. *Attention, Perception, and Psychophysics*. https://doi.org/10.3758/APP.72.2.353

Scholl, B. J. (2001). Objects and attention: The state of the art. *Cognition*. https://doi.org/10.1016/S0010-0277(00)00152-9

Schubotz, R. I. (2007). Prediction of external events with our motor system: towards a new framework. *Trends in Cognitive Sciences*. https://doi.org/10.1016/j.tics.2007.02.006

Schurgin, M. W., Wixted, J. T., & Brady, T. F. (2019). Psychophysical Scaling Reveals a Unified Theory of Visual Memory Strength. *BioRxiv*, 325472. https://doi.org/10.1101/325472

Senoussi, M., Moreland, J. C., Busch, N. A., & Dugué, L. (2019). Attention explores space periodically at the theta frequency. *Journal of Vision*. https://doi.org/10.1167/19.5.22

Sergent, C., & Dehaene, S. (2004). Is Consciousness a Gradual Phenomenon?: Evidence for an All-or-None Bifurcation During the Attentional Blink. *Psychological Science*, *15*(11), 720–728. https://doi.org/10.1111/j.0956-7976.2004.00748.x

Sergent, C., Wyart, V., Babo-Rebelo, M., Cohen, L., Naccache, L., & Tallon-Baudry, C. (2013). Cueing attention after the stimulus is gone can retrospectively trigger conscious perception. *Current Biology*. https://doi.org/10.1016/j.cub.2012.11.047

Sergent, C, Ruff, C., & Barbot, A. (2011). Top–down modulation of human early visual cortex after stimulus offset supports successful postcued report. *Journal of Cognitive* …. Retrieved from http://www.mitpressjournals.org/doi/abs/10.1162/jocn.2010.21553

Sergent, Claire. (2018). The offline stream of conscious representations. *Philosophical Transactions of the Royal Society B: Biological Sciences*. https://doi.org/10.1098/rstb.2017.0349

Shekhar, M., & Rahnev, D. (2018). Distinguishing the roles of dorsolateral and anterior PFC in visual metacognition. *Journal of Neuroscience*. https://doi.org/10.1523/JNEUROSCI.3484-17.2018

Shepard, R. N. (1987). Toward a universal law of generalization for psychological science. *Science*. https://doi.org/10.1126/science.3629243

Sheth, B. R., Nijhawan, R., & Shimojo, S. (2000). Changing objects lead briefly flashed ones. *Nature Neuroscience*. https://doi.org/10.1038/74865

Shiu, L. po, & Pashler, H. (1994). Negligible Effect of Spatial Precuing on Identification of Single Digits. *Journal of Experimental Psychology: Human Perception and Performance*. https://doi.org/10.1037/0096-1523.20.5.1037

Simione, L., Akyürek, E. G., Vastola, V., Raffone, A., & Bowman, H. (2017). Illusions of integration are subjectively impenetrable: Phenomenological experience of Lag 1 percepts during dual-target RSVP. *Consciousness and Cognition*, *51*, 181–192. https://doi.org/10.1016/j.concog.2017.03.004

Simons, D. J., & Chabris, C. F. (1999). Gorillas in our midst: Sustained inattentional blindness for dynamic events. *Perception*. https://doi.org/10.1068/p281059

Snir, G., & Yeshurun, Y. (2017). Perceptual episodes, temporal attention, and the role of cognitive control: Lessons from the attentional blink. In *Progress in Brain Research*. https://doi.org/10.1016/bs.pbr.2017.07.008

Solomon, J. A., & Morgan, M. J. (2018). Precues' elevation of sensitivity is not only preattentive, but largely monocular. *Attention, Perception, and Psychophysics*. https://doi.org/10.3758/s13414-018-1564-1

Solomon, Joshua A. (2004). The effect of spatial cues on visual sensitivity. In *Vision Research*. https://doi.org/10.1016/j.visres.2003.12.003

Spekreijse, H. (2000). Pre-attentive and attentive mechanisms in vision. Perceptual organization and dysfunction. *Vision Research*. https://doi.org/10.1016/S0042-6989(00)00060-2

Sperling, G. (1960). The information available in brief visual presentations. *Psychological Monographs: General and Applied*. https://doi.org/10.1037/h0093759

Sperling, G., & Melchner, M. J. (1978). The attention operating characteristic: Examples from visual search. *Science*.

https://doi.org/10.1126/science.694536

Spitzer, H., Desimone, R., & Moran, J. (1988). Increased attention enhances both behavioral and neuronal performance. *Science*. https://doi.org/10.1126/science.3353728

Thibault, L., Van Den Berg, R., Cavanagh, P., & Sergent, C. (2016). Retrospective attention gates discrete conscious access to past sensory stimuli. *PLoS ONE*. https://doi.org/10.1371/journal.pone.0148504

Tian, X., Yoshida, M., & Hafed, Z. M. (2016). A Microsaccadic Account of Attentional Capture and Inhibition of Return in Posner Cueing. *Frontiers in Systems Neuroscience*, *10*, 23. https://doi.org/10.3389/fnsys.2016.00023

Tkacz-Domb, S., & Yeshurun, Y. (2018). The size of the attentional window when measured by the pupillary response to light. *Scientific Reports*. https://doi.org/10.1038/s41598-018-30343-7

Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, *12*(1), 97–136. https://doi.org/10.1016/0010-0285(80)90005-5

Turatto, M., Vescovi, M., & Valsecchi, M. (2007). Attention makes moving objects be perceived to move faster. *Vision Research*. https://doi.org/10.1016/j.visres.2006.10.002

Valera, F. J., Toro, A., Roy John, E., & Schwartz, E. L. (1981). Perceptual framing and cortical alpha rhythm. *Neuropsychologia*. https://doi.org/10.1016/0028-3932(81)90005-1

Valsecchi, M., Vescovi, M., & Turatto, M. (2010). Are the effects of attention on speed judgments genuinely perceptual? *Attention, Perception, and Psychophysics*. https://doi.org/10.3758/APP.72.3.637

Van Den Berg, R., Shin, H., Chou, W. C., George, R., & Ma, W. J. (2012). Variability in encoding precision accounts for visual short-term memory limitations. *Proceedings of the National Academy of Sciences of the United States of America*. https://doi.org/10.1073/pnas.1117465109

van den Berg, R., Yoo, A. H., & Ma, W. J. (2017). Fechner's law in metacognition: A quantitative model of visual working memory confidence. *Psychological Review*. https://doi.org/10.1037/rev0000060

van den Berg, R., Zylberberg, A., Kiani, R., Shadlen, M. N., & Wolpert, D. M. (2016). Confidence Is the Bridge between Multi-stage Decisions. *Current Biology*. https://doi.org/10.1016/j.cub.2016.10.021

VanRullen, R. (2016). Perceptual Cycles. *Trends in Cognitive Sciences*. https://doi.org/10.1016/j.tics.2016.07.006

VanRullen, R., Carlson, T., & Cavanagh, P. (2007). The blinking spotlight of attention. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(49), 19204–19209. https://doi.org/10.1073/pnas.0707316104

Verghese, P. (2001). Visual search and attention: A signal detection theory approach. *Neuron*. https://doi.org/10.1016/S0896-6273(01)00392-0

Vickers, D., & Packer, J. (1982). Effects of alternating set for speed or accuracy on response time, accuracy and confidence in a unidimensional discrimination task. *Acta Psychologica*, *50*(2), 179–197. https://doi.org/10.1016/0001-6918(82)90006-3

Vul, E., Hanus, D., & Kanwisher, N. (2008). Delay of selective attention during the attentional blink. *Vision Research*, *48*(18), 1902–1909. https://doi.org/10.1016/j.visres.2008.06.009

Vul, E., Hanus, D., & Kanwisher, N. (2009). Attention as inference: Selection is probabilistic; responses are all-or-none samples. *Journal of Experimental Psychology: General*, *138*(4), 546–560. https://doi.org/10.1037/a0017352

Vul, E., Nieuwenstein, M., & Kanwisher, N. (2008). Temporal Selection is Suppressed, Delayed, and Diffused During the Attentional Blink. *Psychological Science*, *19*(1), 55–61. https://doi.org/10.1111/j.1467-9280.2008.02046.x

Wang, Z., & Klein, R. M. (2010). Searching for inhibition of return in visual search: A review. *Vision Research*. https://doi.org/10.1016/j.visres.2009.11.013

Weil, L. G., Fleming, S. M., Dumontheil, I., Kilford, E. J., Weil, R. S., Rees, G., … Blakemore, S. J. (2013). The development of metacognitive ability in adolescence. *Consciousness and Cognition*. https://doi.org/10.1016/j.concog.2013.01.004

Weiskrantz, L. (1996). Blindsight revisited. *Current Opinion in Neurobiology*. https://doi.org/10.1016/S0959-4388(96)80075-4

Wiener, M., Turkeltaub, P., & Coslett, H. B. (2010). The image of time: A voxel-wise meta-analysis. *NeuroImage*. https://doi.org/10.1016/j.neuroimage.2009.09.064

Wierda, S. M., van Rijn, H., Taatgen, N. a., & Martens, S. (2012). Pupil dilation deconvolution reveals the dynamics of attention at high temporal resolution. *Proceedings of the National Academy of Sciences*, *109*(22), 8456–8460. https://doi.org/10.1073/pnas.1201858109

Wilimzig, C., Tsuchiya, N., Fahle, M., Einhäuser, W., & Koch, C. (2008). Spatial attention increases performance but not subjective confidence in a discrimination task. *Journal of Vision*, *8*(5), 7. https://doi.org/10.1167/8.5.7

Williford, T., & Maunsell, J. H. R. (2006). Effects of spatial attention on contrast response functions in macaque area V4. *Journal of Neurophysiology*. https://doi.org/10.1152/jn.01207.2005

Wong, K. F. E. (2002). The relationship between attentional blink and psychological refractory period. *Journal of Experimental Psychology: Human Perception and Performance*. https://doi.org/10.1037/0096-1523.28.1.54

Wyart, V., Dehaene, S., & Tallon-Baudry, C. (2012). Early dissociation between neural signatures of endogenous spatial attention and perceptual awareness during visual masking. *Frontiers in Human Neuroscience*. https://doi.org/10.3389/fnhum.2012.00016

Wyart, V., & Tallon-Baudry, C. (2008). Neural dissociation between visual awareness and spatial attention. *Journal of Neuroscience*. https://doi.org/10.1523/JNEUROSCI.4748-07.2008

Wyble, B., Bowman, H., & Nieuwenstein, M. (2009). The Attentional Blink Provides Episodic Distinctiveness: Sparing at a Cost. *Journal of Experimental Psychology: Human Perception and Performance*, *35*(3), 787–807. https://doi.org/10.1037/a0013902

Wyble Brad, B., Potter, M. C., Bowman, H., & Nieuwenstein, M. (2011). Attentional episodes in visual perception. *Journal of Experimental Psychology: General*. https://doi.org/10.1037/a0023612

Yeshurun, Y. (2019). The spatial distribution of attention. *Current Opinion in Psychology*. https://doi.org/10.1016/j.copsyc.2018.12.008

Yeshurun, Y., & Carrasco, M. (1998). Attention improves or impairs visual performance by enhancing spatial resolution. *Nature*. https://doi.org/10.1038/23936

Yeshurun, Y., & Levy, L. (2003). Transient spatial attention degrades temporal resolution. *Psychological Science*. https://doi.org/10.1111/1467-9280.02436

Yeshurun, Y., & Rashal, E. (2010). Precueing attention to the target location diminishes crowding and reduces the critical distance. *Journal of Vision*. https://doi.org/10.1167/10.10.16

Yeung, N., & Summerfield, C. (2012). Metacognition in human decision-making: confidence and error monitoring. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *367*(1594), 1310–1321. https://doi.org/10.1098/rstb.2011.0416

Yuval-Greenberg, S., Merriam, E. P., & Heeger, D. J. (2014). Spontaneous microsaccades reflect shifts in covert attention. *Journal of Neuroscience*. https://doi.org/10.1523/JNEUROSCI.0582-14.2014

Zackon, D. H., Casson, E. J., Zafar, A., Stelmach, L., & Racette, L. (1999). The temporal order judgment paradigm: Subcortical attentional contribution under exogenous and endogenous cueing conditions. *Neuropsychologia*. https://doi.org/10.1016/S0028-3932(98)00134-1

Zhou, L. F., Buetti, S., Lu, S., & Cai, Y. C. (2018). Attentional effect on contrast appearance: From enhancement to attenuation. *Journal of Experimental Psychology: Human Perception and Performance*. https://doi.org/10.1037/xhp0000499

Zizlsperger, L., Sauvigny, T., & Haarmeier, T. (2012). Selective attention increases choice certainty in human decision making. *PLoS ONE*. https://doi.org/10.1371/journal.pone.0041136

Zizlsperger, L., Sauvigny, T., Händel, B., & Haarmeier, T. (2014). Cortical representations of confidence in a visual perceptual decision. *Nature Communications*. https://doi.org/10.1038/ncomms4940

Zylberberg, A., Wolpert, D. M., & Shadlen, M. N. (2018). Counterfactual Reasoning Underlies the Learning of Priors in Decision Making. *Neuron*. https://doi.org/10.1016/j.neuron.2018.07.035

RESUME EN FRANÇAIS

La dynamique du monde qui nous entoure nécessite sans cesse d'adapter nos décisions à son incertitude latente. Cette incertitude définit autant notre perception que le fonctionnement même de nos fonctions cognitives. La « métacognition » d'un individu – la manière dont il raisonne sur ses propres perceptions - peut être étudiée en comparant sa confiance à la qualité objective de ses décisions perceptives. Parce que l'attention sélective est une source importante de modulation sensorielle, une bonne métacognition des effets de l'attention sur la perception semble primordiale. La façon dont la confiance émerge du processus d'orientation de l'attention, et se développe ensuite dans l'espace et le temps, fait l'objet de cette thèse. Nous y décrivons notamment la solide dépendance que la confiance cultive à l'égard de l'attention visuelle, une dépendance qui subsiste à chaque étape du processus attentionnel. Les travaux expérimentaux présentés dans cette thèse suggèrent ainsi une dépendance si forte qu'une orientation erronée de l'attention passe souvent inaperçue au niveau métacognitif. Ces résultats témoignent de l'incapacité de la confiance à prendre en compte certaines des limites temporelles de l'attention sélective. Dans ce résumé en français, nous présenterons, après une introduction sommaire aux différents concepts scientifiques clés, un résumé des résultats expérimentaux obtenus durant cette thèse. Ces résultats seront enfin mis en perspective dans un résumé de la discussion présente à la fin du manuscrit.

INTRODUCTION GENERALE

Prendre une décision est une forme d'abandon, un renoncement selon les mots du romancier italien Italo Calvino. Cette renonciation est parfois un déchirement, parfois un soulagement, mais son issue est un monde de possibilités disparue à jamais. Ce monde des possibles, ce « coût d'opportunité », et la conception subjective que nous en avons, déterminent nos décisions quotidiennes. D'une manière générale, notre perception de l'environnement est un flux incessant de décisions. Une décision perceptuelle est une décision sur ce que l'on a effectivement perçu, et parfois, lorsqu'il y a peu de consensus sur la question, le cerveau peut avoir à faire des calculs et des hypothèses. De ce point de vue, la perception peut être elle-même comprise comme une décision, c'est-à-dire une tentative de réduire l'incertitude du monde qui nous entoure. Ce jeu avec l'incertain existe principalement pour une raison : le choix optimal n'existe jamais dans le monde réel, en raison de l'impossibilité pour l'esprit de reproduire l'infinité de probabilités propre aux événements extérieurs. Pour le physiologiste Hermann von Helmholtz et de nombreux neuroscientifiques contemporains, « le système perceptuel humain [est] un moteur d'inférence statistique dont la fonction est d'inférer les causes probables de l'information sensorielle » (Dayan, Hinton, Neal et Zemel, 1995). En l'absence de certitude arrêtée, il faut donc constamment choisir, c'est-à-dire abandonner une multitude de possibilités au profit d'une seule. Nous commencerons donc la première partie de l'introduction en nous concentrant sur deux aspects de la psychophysique moderne dans le contexte de l'étude de la prise de décision perceptuelle : le « comment » des décisions perceptuelles et le degré de compréhension subjective que nous portons sur ces décisions. Pour ce faire, nous assemblerons deux pièces du puzzle : comment le psychophysicien étudie (a) la décision perceptuelle elle-même, également appelée Type 1 (pour une décision de « premier ordre ») et (b) le sentiment de confiance connexe qui va avec, également connu sous le nom de Type 2 (pour « décision de second ordre »). Le but de cette introduction n'est pas de fournir au lecteur une vue exhaustive du domaine, mais plutôt de zoomer sur les aspects les plus pertinents de la question pour les chapitres à suivre.

La perception d'un objet dans une scène visuelle riche peut être définie comme le produit d'une fonction séparant le signal du bruit : par exemple, le cerveau doit déterminer les contours de l'objet, son appartenance à une famille connue d'objets et la sémantique qui lui est associée, afin d'en déduire enfin sa probable identité. Le défi de ce processus d'inférence est précisément ce qui rend le cerveau si fascinant et complexe à étudier. L'idée d'une fonction de filtrage du bruit transformant une entrée probabiliste en une réponse discrète a conduit, dans les années 60, à l'adaptation de la théorie de la détection au domaine de la psychologie (Green & Swets, 1966). La théorie de la détection du signal, en psychophysique, postule qu'une décision perceptuelle résulte de la combinaison d'une certaine sensibilité (ou « d' ») et d'un biais de réponse (ou « critère »), appliquée à une entrée donnée. Les distributions de probabilité du signal et du bruit sont supposées être normales (c.-à-d. Gaussiennes) et souvent de variance égale, fournissant une probabilité calculable pour chaque niveau d'information sensorielle dans un espace de décision donné.

Dans le cadre d'une expérience, un stimulus présenté à l'observateur peut, par exemple, être échantillonné de deux catégories distinctes : il peut être orienté dans le sens horaire (stimulus A) ou bien dans le sens antihoraire (stimulus B). Chacune de ces deux catégories est liée à une distribution de probabilités donnée (voir fig. 1a de la discussion générale). Les fonctions de vraisemblance de chacun des deux stimuli sont souvent supposées être de variance égale. La distance entre les deux moyennes de distribution (en unités d'information de Type 1) correspond à la sensibilité interne, c'est-à-dire à la distinction entre les deux catégories du point de vue de l'observateur. Plus la sensibilité est grande, meilleure est la discrimination. La présentation d'un stimulus à un observateur entraînera une certaine accumulation d'information : le point sur l'axe de cette quantité représentant un échantillon donné est appelé la variable de décision. Il reste une dernière étape avant de convertir efficacement l'information de Type 1 en une décision réelle. Pour répondre, l'observateur doit choisir le stimulus en plaçant un seuil, ou critère, le long de l'axe d'information de Type 1 : toute valeur en dessous de ce critère sera classée comme favorisant le stimulus A, et toute valeur au-dessus du critère comme favorisant le stimulus B. La TDS (« théorie de la détection du signal ») se concentre précisément sur cette différence entre le critère - ou biais - et la sensibilité réelle d'un observateur.

Quand est-il de la confiance ? Lorsqu'un observateur décide de l'orientation d'un stimulus, il peut avoir plus ou moins confiance dans la qualité de sa réponse. La TDS de type 1 ne semble pas produire une idée claire de ce que la confiance signifie empiriquement. Sans définition claire, les méthodes doivent laisser suffisamment de place pour que des différences potentielles apparaissent entre les indices utilisés dans les réponses de type 1 et les indices utilisés dans les jugements de confiance. Pour ce faire, la confiance peut être objectivement liée à sa propre forme de sensibilité. Cette sensibilité spécifique provient du constat suivant : le fait d'avoir une faible confiance dans une réponse de type 1 lorsque cette réponse est incorrecte ne doit pas être considéré comme aussi erroné que le fait d'être très confiant à ce sujet. De la même manière que pour le type 1, la confiance a sa propre sensibilité (c'est-à-dire, sa propre mesure de performance) et peut donc être considérée comme une véritable décision de type 2, une décision dont l'objet est une autre décision.

Paradoxalement, le champ de la psychologie et, récemment, des neurosciences cognitives manquent souvent d'une définition sans équivoque de l'attention et contournent ce dilemme en se concentrant sur ce qu'elle fait plutôt que sur ce qu'elle est (Anderson, 2011). La citation notoirement célèbre de James - « Tout le monde sait ce qu'est l'attention » - reste un diagnostic assez juste de la pathologie : en tant qu'êtres humains, la pratique quotidienne de notre attention suffirait à sa compréhension. James avait cependant proposé une définition de l'attention : « C'est la prise de possession par l'esprit, sous une forme claire et vivante, de l'un de ce qui semble être simultanément plusieurs objets ou courants de pensée possibles. (…) Cela implique le retrait de certaines choses afin de traiter efficacement les autres (…) » (p. 405, James, 1890). Bien entendu, cette définition ne satisfera pas un puriste de la méthode scientifique, encore moins un philosophe, qui questionnera inévitablement le sens de « possession », « clair », « objets » et « pensée ». Un psychologue contemporain pourra également affirmer que l'attention peut prendre possession de plusieurs objets simultanément. Néanmoins, cette définition n'a pas beaucoup changé depuis la publication de *The principles of psychology* il y a plus d'un siècle. Un aspect intéressant de l'attention est souvent négligé dans la définition de James : la notion de retrait. Cette interprétation de l'attention comme mécanisme qui sélectionne un stimulus tout en étant préjudiciable au traitement d'autres stimuli est encore aujourd'hui la pierre angulaire de notre définition de l'attention. Pour citer la psychophysicienne Marisa Carrasco, « c'est le mécanisme qui transforme le regard en vision. (…) L'attention nous permet de traiter de manière sélective la grande quantité d'informations auxquelles nous sommes confrontés, en priorisant certains aspects de l'information tout en ignorant les autres et en nous concentrant sur un certain emplacement ou aspect de la scène visuelle » (Carrasco, 2011). L'attention, c'est la sélection et la priorisation d'un stimulus jugé pertinent : les ressources limitées de tout organisme vivant nécessitent de facto un mécanisme de filtrage. Chez l'homme et de nombreux animaux, ce processus de sélection peut être ajusté dynamiquement dans l'espace et le temps pour privilégier les informations nécessaires au maintien de l'homéostasie. Afin de distinguer la nature des différents objets de l'attention sélective, une taxonomie riche - parfois redondante - a émergé au fil des années.

Par souci de simplicité, nous nous concentrerons ici sur deux aspects de l'attention dans le domaine de la vision, à savoir l'attention spatiale et temporelle. L'attention spatiale se rapporte à la priorisation et à l'amélioration du traitement d'un stimulus à un endroit particulier (Carrasco, 2011). L'attention temporelle, quant à elle, fait référence à l'amélioration d'un stimulus à un moment donné (Coull et Nobre, 1998; Nobre et van Ede, 2017). Le lecteur l'aura sûrement remarqué, la définition du stimulus visuel reste ici largement imprécise. Il est également envisageable de sélectionner un stimulus non pas par ses aspects spatiaux ou temporels, mais en fonction de caractéristiques plus intrinsèques (telles que la couleur ou la forme). Ce troisième type de sélectivité de l'attention visuelle a été baptisé « attention basée sur les caractéristiques » (« feature-based attention »), et fait l'objet d'une littérature importante (voir Maunsell & Treue, 2006; Carrasco, 2011 pour une revue de la littérature sur le sujet). Enfin, une dernière version quelque peu hybride se nomme attention basée sur les objets (« object-based attention », Scholl, 2001). Dans l'attention basée sur les objets, le processus de sélection est régi par la structure de l'objet lui-même (par exemple, concentrer l'attention sur une forme rectangulaire facilite le traitement des stimuli qui y sont contenus). Nous ne couvrirons pas l'attention basée sur les caractéristiques et celle basée sur les objets dans la présente thèse.

La différence de performance perceptuelle entre deux endroits – le premier où l'attention s'est posée, et l'autre en dehors du focus attentionnel - est à la base de l'étude psychophysique de l'attention. Initialement, l'attention avait été décrite comme un projecteur (Posner, 1980), mais l'obscurité censée baigner une partie de la scène visuelle est souvent relative : il subsiste la possibilité de traiter une partie de l'information en dehors du focus de l'attention, comme nous le verrons plus loin. Dans d'autres études, l'attention a été comparée à un zoom, parfois avec une granularité plus grossière à la périphérie de son foyer (Eriksen & St. James, 1986; Eriksen & Yeh, 1985). Le principe latent derrière ces métaphores est la sérialité du processus attentionnel. Par définition, il y a une limite théorique à la taille du focus de l'attention, elle ne peut pas englober toute la scène visuelle. Cette observation trouve son origine dans les résultats empiriques des tâches impliquant la recherche de conjonctions : lorsqu'une cible est intégrée parmi des distracteurs, le temps nécessaire pour identifier la cible est proportionnel au nombre de distracteurs, ce qui suggère que l'attention explore la scène visuelle d'une manière discrète et sérielle (Treisman & Gelade, 1980). De ce point de vue, la métaphore du projecteur a l'avantage d'évoquer une ressource unique mais modulable. Cependant, il faut être prudent : alors que la littérature suggère une certaine flexibilité dans la taille du focus attentionnel, l'élargissement de la zone sélectionnée a souvent un coût en termes de performances (Eriksen & Murphy, 1987; Eriksen & St. James, 1986). La métaphore du « zoom », avec différents degrés de résolution est un candidat intéressant pour illustrer le concept. Cependant, l'attention peut non seulement améliorer, mais également compromettre l'acuité visuelle à certains emplacements. En d'autres termes, faire attention à un endroit de la scène visuelle conduit à une plus grande résolution à cet emplacement, mais diminue également la résolution aux autres emplacements (Herrmann, Montaser-Kouhsari, Carrasco et Heeger, 2010; Pestilli et Carrasco, 2005). En laboratoire, la manipulation de l'attention visuelle est généralement réalisée à l'aide d'indices. Le principe général est de présenter un stimulus saillant peu de temps avant le début du stimulus d'intérêt, pour attirer l'attention vers l'emplacement prédéterminé et faciliter la discrimination. Ce paradigme classique est souvent appelé « paradigme de Posner », du nom de Michael Posner qui a opérationnalisé l'approche dans une étude historique des années 80 (Posner, 1980). Une expérience de repérage Posner typique implique deux localisations distinctes, de chaque côté d'une croix de fixation présentée au centre de l'écran. À chaque essai, le participant se voit présenter un indice central, indiquant deux scénarios possibles : soit l'indice est neutre, auquel cas la cible est susceptible d'apparaître aléatoirement sur l'un ou l'autre côté de l'écran, soit l'indice indique un seul emplacement, prédisant avec ~ 80% de chances où la cible apparaîtra. Du point de vue de l'expérimentateur, il y a trois conditions : valide, invalide et neutre. Il est alors possible de comparer le temps de réponse du participant dans la condition valide ou dans la condition invalide, face à la condition neutre. Par rapport à la condition neutre qui sert de référence, les participants sont généralement plus rapides dans les essais valides et plus lents dans les essais invalides (Posner, 1980). Au fil des ans, cette approche s'est révélée très robuste (fig. 2 de la discussion générale).

Approche experimentale

« Les sensations subjectives n'intéressent principalement que les investigateurs scientifiques. Si elles se font remarquer dans l'activité ordinaire des sens, elles ne sont proprement bonnes qu'à distraire l'attention. »

- Hermann von Helmholtz, Traité d'optique physiologique : les perceptions de la vision (1825).

Selon les mots de Helmholtz, la curiosité de l'expérimentateur peut souvent dépasser ce qu'est réellement la perception. Aujourd'hui, grâce à la notion de confiance perceptuelle, l'expérimentateur peut étudier les impressions subjectives en termes objectifs. L'attention est ainsi devenue un candidat viable pour sonder les limites de l'introspection.

Chapitre 1 : Metacognition & indices exogenes

Dans le premier chapitre de cette thèse, nous utilisons une implémentation canonique d'un paradigme de Posner pour étudier la relation entre l'attention et la confiance. Jusqu'à présent, la littérature sur la métacognition et l'attention n'avait pas étudié l'effet de la manipulation exogène de l'attention sur les jugements de confiance. Malgré le fameux « paradigme de Posner » exogène appliqué avec succès pour étudier de nombreux aspects de l'attention spatiale aux niveaux sensoriel et cognitif, pas une seule étude, à notre connaissance, ne l'a combiné directement avec des jugements de confiance (sans potentiel confusions, voir la section 3.1.1 de l'introduction générale). Pourtant, le rôle de l'attention exogène sur la perception consciente et les jugements de visibilité a été étudié à l'aide de nombreux paradigmes attentionnels, autant dans l'espace que dans le temps : par exemple, des indices exogènes ont été utilisés pour modifier la visibilité subjective (voir la section 2.3.2 de l'introduction générale). Dans le chapitre 1, nous étudions ainsi l'effet de l'attention spatiale exogène sur les jugements de confiance, via un paradigme hautement reproduit dans la littérature attentionnelle.

Nous y montrons que les jugements de confiance peuvent s'adapter à l'augmentation initiale de la sensibilité induite par l'attention. L'augmentation précoce de la sensibilité et de la confiance est de courte durée et disparaît pour des intervalles de temps plus longs, ce qui confirme l'aspect exogène de notre manipulation. La capacité métacognitive reste quant à elle stable sur les différentes périodes entre indice et cible. Ces résultats suggèrent que la confiance visuelle est capable de suivre les effets perceptifs d'indices exogènes imprévisibles.

Chapitre 2 : Metacognition et desengagement de l'attention

Dans le chapitre précédent, nous avons observé que la confiance était en mesure de suivre les premiers effets d'indices exogènes sur la sensibilité. Ce résultat suggère que malgré l'utilisation d'indices imprévisibles et non pertinents pour la tâche, le gain de précision induit par ces événements transitoires et non pertinents était toujours détectable dans les jugements de confiance. Cependant, l'orientation exogène n'est pas la seule situation dans laquelle les effets de l'attention spatiale peuvent être considérés comme non pertinents pour une tâche. Une autre situation de ce type se présente lorsque l'attention volontaire dans un lieu spécifique n'est plus pertinente pour la tâche à accomplir. L'attention spatiale devrait alors être désengagée de l'emplacement initial. Malgré le désengagement, lorsqu'un événement local se produit peu de temps après la fin de l'épisode d'attention endogène, cet événement pourrait tout de même bénéficier des effets résiduels de l'attention. Ici, nous définissons le « désengagement attentionnel » comme le processus de désallocation progressive de l'attention endogène d'un endroit donné avant de finalement la réorienter vers un autre endroit. En ce sens, le désengagement attentionnel est une phase de transition entre deux états attentionnels stables. Nous verrons dans ce chapitre que le désengagement attentionnel peut prendre plus de temps que la réorientation, qui peut se produire dans un laps de temps très court. L'expérience présentée a été initialement conçue pour tester deux aspects de l'attention sélective et de la confiance : (a) l'effet global du désengagement attentionnel sur les jugements de confiance et (b) la structure temporelle rythmique plus fine de l'attention sélective et ses effets sur la confiance. Nous ne présenterons que le premier aspect de ce travail dans le présent chapitre, le deuxième aspect nécessitant de grandes quantités de données. Dans ce chapitre, nous étudions donc comment la confiance suit les effets de la réorientation et du désengagement attentionnels après un épisode initial d'attention endogène. Le protocole expérimental a été spécifiquement conçu pour sonder la performance perceptuelle ainsi que la confiance des participants. Surtout, nous utilisons une mesure de confiance sans biais afin d'extrapoler la sensibilité métacognitive.

Dans ce chapitre, nous observons ainsi que la confiance est en mesure de suivre le désengagement progressif de l'attention endogène. En particulier, les jugements métacognitifs sont prédictifs de la fluctuation - observable essai par essai - de la différence d'erreur entre l'emplacement de la cible et celui du distracteur. La capacité métacognitive diminue également avec le désengagement, ce qui suggère un rôle spécifique de l'attention sélective sur la métacognition. Enfin, la confiance semble également s'adapter à la réorientation brusque de l'attention provoquée par des indices invalides, confirmant le lien étroit qui est susceptible d'exister entre la confiance et les mécanismes spatio-temporels de l'attention.

CHAPITRE 3 : L'ATTENTION TEMPORELLE CAUSE DES BIAIS SYSTEMATIQUES DANS LA CONFIANCE VISUELLE

Dans le chapitre précédent, nos résultats attestent du rôle puissant de la structure temporelle de l'attention spatiale dans la construction de la confiance perceptuelle. Pourtant, pour mieux comprendre cette influence, nous aurions besoin de manipuler le timing de l'attention indépendamment des exigences de la tâche, afin d'induire des conflits entre l'état de l'attention et la capacité à effectuer la tâche. Dans le présent chapitre, nous adaptons un paradigme classique de clignement attentionnel pour induire des discontinuités dans l'orientation de l'attention temporelle. Cette approche nous permet ainsi d'étudier comment la confiance réagit lorsque l'attention est poussée à ses limites, en sélectionnant le mauvais stimulus dans le temps.

Ici, notre objectif est d'évaluer comment la confiance et les performances des observateurs sont affectées lorsque l'attention temporelle est mise à l'épreuve, et de vérifier si la confiance est capable de suivre les limites de l'attention temporelle. L'attention temporelle améliore un stimulus à un moment donné (Coull et Nobre, 1998) et inhibe d'autres moments (Denison et al., 2017), tout comme l'attention spatiale dans l'espace (Carrasco, 2011). L'attention et la confiance sont toutes deux liées à la précision : l'attention augmente le rapport signal / bruit du stimulus, tandis que la confiance reflète idéalement cette augmentation. L'attention et la confiance ont déjà été étudiées ensemble dans le domaine spatial, conduisant à des résultats mitigés : certaines études ont observé une dissociation entre les deux (Rahnev et al., 2011; Schoenherr et al., 2010; Wilimzig et al., 2008) , tandis que d'autres ont suggéré que l'attention spatiale est bien intégrée à la confiance (Denison et al., 2018; Recht, de Gardelle et Mamassian, 2017; Zizlsperger et al., 2012, 2014). Dans le domaine temporel, ce lien entre attention temporelle et confiance reste largement inexploré. Cette question est particulièrement pertinente compte tenu de la possibilité que l'attention et la confiance puissent fonctionner à des échelles de temps distinctes (D. Rahnev et al., 2015).

Dans certaines circonstances, l'attention temporelle peut être supprimée, retardée ou déplacée. Une solide observation concernant les limites de l'attention temporelle est celle du «clignement attentionnel» (Broadbent & Broadbent, 1987; Raymond et al., 1992). Plus précisément, lorsque deux cibles sont intégrées dans un flux de présentation visuelle série rapide, la deuxième cible T2 est souvent manquée lorsqu'elle apparaît peu de temps (150-300 ms) après la première cible T1. Lorsque la sélection temporelle n'est pas simplement supprimée dans le cas de cibles T2 manquées, elle est retardée, de sorte qu'un distracteur suivant T2 serait signalé à sa place. Ces retards de sélection, parfois appelés « intrusions d'erreur post-cible » (Chun, 1997; Vul, Hanus, et al., 2008) sont une deuxième caractéristique du clignement attentionnel. Enfin, lorsque T2 est présentée immédiatement après T1 (60-100 ms), les deux cibles sont en moyenne rapportées correctement. Cet effet, baptisé «lag-1 sparing» (Hommel & Akyürek, 2005) est une troisième caractéristique du clignement attentionel. Ces trois caractéristiques peuvent être expliquées par une variété de modèles (Dux & Marois, 2009; Martens & Wyble, 2010). Cependant, la question de savoir si la confiance suit ces trois caractéristiques reste ouverte.

Pour répondre à cette question, nous avons utilisé un paradigme de clignement attentionnel en combinaison avec des jugements de confiance, afin d'évaluer si les jugements de confiance des participants au sujet des rapports de T2 refléteraient la diminution de la précision pendant le clignement attentionnel, la conservation de la précision au lag-1, et le retard dans la sélection temporelle qui suit le clignement attentionnel. Nous avons également recueilli des jugements de confiance pour T1 comme base de comparaison. Pour mesurer les erreurs et les retards dans la sélection temporelle, nous avons présenté aux participants un flux rapide de lettres et indiqué deux lettres dans le flux pour un rapport ultérieur. La position de chaque lettre dans le flux a fourni des informations essentielles sur le moment où l'attention a été déployée (Goodbourn et al., 2016; Martini, 2012; Vul, Nieuwenstein, et al., 2008).

Nous avons trouvé une forte corrélation entre la fréquence des rapports et la confiance lors de la sélection temporelle (T1), qui se maintient lorsque l'attention doit se réorienter vers un deuxième point dans le temps (T2), suggèrant que la décision et la confiance partagent principalement le même signal de preuve lors de l'orientation temporelle de l'attention. Ce couplage étroit pourrait empêcher la confiance d'accéder aux retards de sélection induits par le clignement attentionnel, comme le montre les résultats empiriques décris dans le présent chapitre. De plus, la confiance semble être affectée par une heuristique pénalisant une cible trop proche dans le temps d'un épisode attentionnel antérieur, pénalité qui expliquerait la sous-confiance observée durant le phénomène de « lag-1 sparing ». Ces multiples phénomènes suggèrent que la confiance n'évalue pas parfaitement l'état de l'attention temporelle dans des situations difficiles, probablement en raison d'un biais heuristique tardif et du fait que la confiance est, d'une certaine façon « attelée » à la dynamique de l'attention temporelle.

Chapitre 4 : Orienter l'attention spatial affaiblit la metacognition

Dans les chapitres précédents, nous avons observé que la confiance était capable de détecter des changements de précision lorsque l'attention était orientée vers le bon moment (chapitre 3, première cible) ou lorsqu'elle se désengageait progressivement du bon endroit (chapitre 2). En revanche, lorsque l'attention était orientée vers le mauvais moment (chapitre 3, deuxième cible), la confiance ne reflétait plus que la précision. Au contraire, elle a continué d'utiliser l'attention en tant que fournisseur fiable d'information, une approche qui semble responsable d'une baisse de la capacité métacognitive. S'il y a une telle dépendance entre la capacité métacognitive et l'attention, qu'en est-il du processus d'orientation lui-même ? Dans ce dernier chapitre, nous avons adapté un paradigme de l'horloge de Wundt pour étudier l'effet de la variabilité essai par essai de l'orientation attentionnelle sur la confiance. Wundt a décrit son paradigme original comme suit : « Laissons, par exemple, une aiguille se déplacer sur une échelle circulaire avec une vitesse uniforme et suffisamment lente, de sorte que les impressions qu'elle donne ne fusionnent pas, mais permettent à sa position à tout instant d'être distinctement perçue. Que le mécanisme d'horlogerie qui la fait tourner déclenche une cloche à chaque révolution, mais à un moment qui peut être varié, de sorte que l'observateur n'a jamais la possibilité de savoir à l'avance quand le coup de cloche aura lieu. (…) Le coup de cloche peut être perçu soit exactement au moment où l'aiguille pointe quand il sonne - dans ce cas, il n'y aura pas de décalage temporel ; soit nous pouvons le combiner avec une position ultérieure de l'aiguille - (…) c'est donc un délai -un décalage, comme nous l'appellerons (…) » (cité dans James, 1887, p. 415). Dans le paradigme présenté dans ce chapitre, nous avons simplement remplacé le son de la cloche par un bref stimulus visuel, et nous avons capitalisé sur l'effet que l'attention a sur le « décalage temporel positif », ou « délai » décrit par Wundt.

Il y a ainsi un aspect du déploiement attentionnel qui a été négligé dans la littérature jusqu'à présent : les observateurs peuvent-ils évaluer le temps qu'il faut pour déployer l'attention spatiale ? La structure temporelle de l'attention spatiale est généralement considérée au travers de ses différents types de traitement. La taxonomie classique dans la littérature différencie l'attention exogène de l'attention endogène. Exogène signifie une orientation involontaire, précoce et de courte durée de l'attention, tandis qu'endogène correspond à une allocation volontaire, tardive et durable (Carrasco, 2011). La nature d'un épisode attentionnel est donc définie principalement par le temps qu'il lui faut pour émerger, l'attention exogène prenant environ 100 ms pour être efficace alors qu'il faut environ 300 ms pour allouer une attention endogène. Par conséquent, le temps est un élément essentiel de l'attention, et pourtant on sait peu de choses sur la façon dont les fluctuations de la temporalité de l'attention affectent la confiance et la métacognition. Ici, nous avons adapté un paradigme de « horloges de Wundt » où les participants doivent reproduire la phase d'une horloge à l'affichage d'un indice. Fondamentalement, ce rapport continu est connu pour être affecté par l'attention et a été considéré comme un indicateur indirect du délai de l'attention (Carlson, Hogendoorn et Verstraten, 2006; Chakravarthi et VanRullen, 2011; Hogendoorn, Carlson, VanRullen et Verstraten, 2010). En ancrant les caractéristiques du stimulus à la temporalité de l'attention, ce paradoxe nous a permis d'enregistrer une signature de la fluctuation temporelle de l'attention spatiale et d'étudier son effet sur les jugements de confiance. Nous l'avons fait en demandant aux participants d'estimer (indirectement) comment le temps de traitement sensoriel était affecté par l'attention dans une tâche perceptuelle. Pour s'assurer que ce processus ne pouvait pas être expliqué par la métacognition de l'incertitude sensorimotrice, nous avons comparé ces résultats à une tâche de détection simple dans laquelle les participants devaient estimer leurs propres temps de réponse. Notre étude a révélé trois résultats majeurs. Premièrement, la confiance visuelle a ignoré la latence de l'attention à la fois exogène et endogène. Deuxièmement, la métacognition a été spécifiquement modifiée *pendant mais pas après* l'orientation endogène de l'attention vers un endroit particulier. Enfin, la capacité métacognitive dans la tâche principale n'était pas corrélée à la métacognition des temps de réponse, ce qui suggère que la métacognition de la variabilité temporelle dans la première tâche ne peut pas être réduite à la métacognition de l'incertitude sensorimotrice.

Discussion & mise en perspective

La confiance que nous avons en ce que nous percevons nous guide dans nos actions et influence de façon durable notre comportement. Le flou, l'instantanéité et l'incertitude, caractéristiques fondamentales de notre perception, devraient toutes être prises en compte dans notre raisonnement métacognitif. Et si une partie de cette incertitude restait inaccessible à notre jugement ? Un grand nombre d'études montrent que nous ne sommes pas conscients de la myriade de processus sensoriels et cognitifs qui déterminent nos interactions quotidiennes avec l'environnement. Comme Helmholtz l'a souligné dans son *Traité sur l'optique physiologique* (1825) : « Les jugements, y compris ceux impliquant des conceptions qui sont indubitablement acquises par l'expérience, sont également déterminés directement par les agences physiologiques de manière caractéristique, et peuvent émerger dans la conscience comme quelque chose de donné immédiatement, d'une façon complète et nécessaire ». En ce sens, même si nous devons réfléchir à la qualité de notre perception, l'attention - le mécanisme par lequel nous accédons et sélectionnons de tels perceptions - peut parfois échouer dans son processus, offrant ainsi à notre esprit un petit mensonge sur la réalité. Dans cette thèse, nous avons montré que la confiance peut rester inconsciente de l'existence de ces petits mensonges. La confiance néglige ainsi le temps nécessaire à l'attention pour se déployer efficacement, même lorsqu'une telle ignorance nuit à la tâche à accomplir.

Dans cette discussion, nous présentons un aperçu de nos résultats et nous en discutons la portée.

Dans cette thèse, nous avons fondé nos études sur la distinction fondamentale entre biais et sensibilité, et nous sommes appuyés sur ce fondement pour tirer des conclusions sur la structure temporelle de l'attention et de la confiance. Ici, nous exposerons les intérêts intrinsèques et certaines limites des méthodes utilisées dans la thèse (1.1). Nous examinerons ensuite de plus près l'une de nos constatations qui montre à quoi ressemble la nature des preuves de type 1 utilisées dans les décisions de type 2 (1.2).

Une mesure pure de la corrélation entre la confiance brute et l'attention ne fournit pas beaucoup d'informations sur la nature des indices utilisés lors de la prise de décision de Type 2 car la sensibilité métacognitive et le biais sont confondus. Nous avons vu dans l'introduction générale (section 3.2.2), l'absence notable de véritables analyses de Type 2 distinguant la sensibilité métacognitive du biais dans l'étude conjointe de l'attention et de la confiance. Dans nos études, nous avons utilisé un certain nombre d'analyses et de méthodes pour résoudre ce problème. Dans le premier chapitre, des « méta-d » et des moyennes de groupe ont été utilisés pour mesurer la façon dont la confiance et la sensibilité métacognitive étaient affectées pendant les repères exogènes. Dans le chapitre 3, nous avons considéré la distribution de la confiance moyenne au cours d'un épisode de sélection, ainsi que la différence de précision entre les essais de confiance élevée et faible, une approche simple qui nous a permis d'identifier la plupart des dissociations trouvées lors de l'orientation temporelle de l'attention comme « biais » de Type 2. Nous reviendrons plus en détail sur cette conclusion plus loin dans la discussion (section 2.1 de la discussion générale et voir ci-dessus). Enfin, dans les chapitres 2 et 4, la méthode utilisée était une combinaison de rapport continu (c'est-à-dire une tâche de reproduction) et de choix forcé de confiance, pour prédire la confiance en fonction de la différence d'erreur essai par essai. Le mélange d'une tâche de reproduction de Type 1 avec méthode du choix forcé de confiance avait un avantage intrinsèque sur les autres méthodes que nous avons utilisées : il permet une étude sans modèle de la relation entre la confiance et les performances, et réduit le biais par rapport à d'autres mesures. Surtout, il ne suppose pas, dans sa forme essentielle, une distribution ou une source déterminée pour les preuves de Type 2. Dans les chapitres 2 et 4, nous avons ensuite pu effectuer une simple analyse de corrélation sur les données de choix forcé de confiance pour évaluer la capacité métacognitive d'un observateur.

Comment ces choix méthodologiques affectent-ils, le cas échéant, nos conclusions sur l'attention et la confiance ? La littérature présente trois relations générales entre l'attention et la confiance (introduction générale, section 3.1) : (1) la confiance ne tient pas compte de l’attention ; (2) la confiance diminue avec l’attention ; (3) ou la confiance augmente avec l'attention. Dans nos études, nous n'avons trouvé aucune preuve que la confiance ne soit pas consciente de l'augmentation de la précision induite par l'attention exogène (1). La nature involontaire de l'orientation attentionnelle (chapitre 1) n'a pas beaucoup changé la corrélation entre la confiance et la précision, contrairement à ce qui a été affirmé précédemment (Kurtz et al., 2017). La confiance a également été fréquemment mise à jour en fonction de la variabilité de l'état d'attention mesurée par la sensibilité métacognitive et le biais (chapitres 1 à 4). De plus, nous n'avons pas trouvé de diminution systématique (2) ni d'augmentation systématique de la confiance (3). En effet, nos résultats mettent en lumière un aspect crucial mais négligé : l'effet de l'attention sur la confiance dépend du temps. Nous avons notamment observé une diminution de la sensibilité métacognitive lorsque l'attention était déplacée ou retardée. Cela semble résulté du fait que la confiance néglige la latence à la fois de l'attention spatiale (chapitre 4) et temporelle (chapitre 3). De plus, la confiance semble également plus faible au niveau du locus attentionnel, lorsque plusieurs stimuli partagent le même épisode attentionnel, un phénomène résultant principalement d'un biais métacognitif (chapitre 3). En revanche, lorsque l'attention n'est pas retardée, la capacité métacognitive est plus importante au sommet de l'épisode attentionnel et diminue progressivement à mesure que l'attention se désengage (chapitre 2).

À première vue, ces résultats peuvent sembler contradictoires, favorisant à la fois l'idée selon laquelle l'attention conduit à une augmentation et à une diminution de la confiance, selon les stimuli, le moment et la métrique utilisée (biais versus sensibilité). Pourtant, ces résultats peuvent être interprété à l’aide d’un concept unique. Dans les paragraphes qui suivent, nous proposons donc un compte rendu intégré de ces effets apparemment disparates, en réexaminant le concept d'épisode attentionnel.

Nous allons d'abord discuter rapidement un résultat imprévu concernant la nature potentielle des indices utilisés lors des jugements de Type 2. Nous avons constaté que dans deux implémentations différentes du paradigme de choix forcé de confiance, et avec deux types de stimulus (par exemple, au Chapitre 2, le stimulus est statique mais pas au Chapitre 4), la différence d'erreur entre deux décisions de Type 1 prédisait fortement les jugements de confiance. De plus, la corrélation entre la différence d'erreur et la confiance a été significativement améliorée lorsque l'on tient compte de l'ampleur globale de l'erreur dans la paire de décisions de Type 1 (fig. 1, A et B de la discussion générale). Cette amélioration a été constatée au niveau des participants individuels au Chapitre 2 et au niveau du groupe pour toutes les tâches du Chapitre 4. Ces résultats et leur cohérence entre les tâches démontrent que l'acuité de la confiance dans la comparaison de deux réponses a été affectée par la précision cumulée de ces réponses, influence qui peut trouver sa source dans la nature des preuves de Type 1 mises à la disposition de la confiance, comme nous le verrons plus loin.

Un principe canonique dans la perception, considéré comme partagé entre les humains et les animaux, est la « loi universelle de généralisation » (Shepard, 1987). Selon Shepard, plus la distance de perception entre deux stimuli est grande, plus la probabilité que les deux stimuli appartiennent à la même catégorie de perception est faible. Une implication fondamentale du principe de généralisation est la décroissance exponentielle, qui est observée avec la distance des stimuli : lorsque deux stimuli sont tous les deux éloignés de la dimension caractéristique centrale, il est possible de confondre les deux et de les catégoriser de manière erronée comme appartenant à la même catégorie perceptive. Ce principe a récemment été utilisé pour expliquer comment les stimuli sont encodés en mémoire de travail, offrant un modèle beaucoup plus simple que les explications existantes dans la littérature (Schurgin, Wixted et Brady, 2019). Les erreurs de large magnitude commises sur une dimension de stimulus (par exemple, l'orientation) sont distribuées de manière plus égale que les petites erreurs car la différence de perception entre elles est *de facto* plus petite. Cette observation découle de la distance dans l'espace psychologique interne : les observateurs représentent les stimuli dans un continuum psychologique qui, contrairement à l'espace des caractéristiques physiques, est intrinsèquement non linéaire. À la lumière de cette représentation non linéaire du monde, il est peu probable que les décisions de Type 2 aient accès à une version linéarisée de l'information utilisée pour les décisions de Type 1. Ainsi, notre constat de non-linéarité empirique observé dans la métacognition pourrait être partiellement dû à la nature même de l'espace psychologique de Type 1.

Nous proposons donc que les indices de Type 1 auxquelles la confiance a accès sont non linéaires. Par exemple, considérons une tâche, telle que celle du Chapitre 4, dans laquelle un observateur doit signaler indiquer la phase d'une horloge (voir le chapitre 4). Les observateurs se voient présenter deux essais consécutifs du même type, puis doivent choisir lequel des deux essais est le meilleur. La décision de Type 1 est un échantillon d'une distribution d'erreurs de forme gaussienne circulaire plus ou moins centrée sur la vraie phase au sein de chaque épisode attentionnel (chapitre 4). Nous supposons, par souci de concision, que le pic de l'épisode attentionnel n'a pas de retard et est centré sur la phase réelle. Le modèle génératif interne de l'observateur obéit donc à une loi normale circulaire, avec une plus grande évidence pour la phase correcte, et une évidence décroissante pour les phases à plus grande distance de celle-ci, d'une façon similaire à la prédiction de la loi de généralisation. La phase à signaler par le participant sera celle avec le signal maximum, probablement à proximité de la phase correcte. Cependant, la fluctuation aléatoire du bruit appliquée à chaque phase candidate le long de l'axe perceptuel pourrait permettre aux phases distantes de gagner, mais moins fréquemment - et donc avec un signal plus faible en moyenne - que les valeurs de phase entourant immédiatement la phase réelle. Cette conception entraîne une conclusion particulière : la différence de signal moyen entre deux erreurs voisines diminue avec leur éloignement progressif du pic de distribution, ce qui rend la discrimination entre elles plus difficile. Le jugement de choix forcé de confiance - dans lequel les participants sélectionnent la réponse de Type 1 avec la confiance la plus élevée - devrait être, idéalement, basée sur la différence d'information (ou « evidence » en anglais) entre ces deux réponses. Une interprétation shepardienne de l'espace perceptif de Type 1 prédit avec précision une sensibilité métacognitive plus faible lorsque les deux erreurs dans la paire sont importantes, ce que nous avons observé dans tous nos données de confiance à choix forcé (fig. 1A contre 1B de la discussion générale). Ce schéma suggère que la confiance suit parfaitement le signal de Type 1, reproduisant également sa non-linéarité. Ce modèle de confiance n'est pas nouveau en soi. Peirce & Jastrow ont postulé que la confiance peut être cartographiée comme reflétant fidèlement le signal de Type 1 (1884; voir la section 1.2.1 de l'introduction générale). Plus récemment, van den Berg & Ma ont proposé une cartographie similaire pour la confiance et la mémoire de travail, où la confiance était liée à la précision du codage de la mémoire de façon non-linéaire (van den Berg et al., 2017). Cependant, cette forme de normalisation à la Weber-Fechner, souvent représenté dans sa forme logarithmique, est différente de notre description de la généralisation : alors que ces auteurs ont proposé que la confiance est une transformation logarithmique du signal interne de Type 1, nous suggérons que la distribution de ce signal de Type 1 est suffisante pour produire des non -linéarités dans certains espaces de décision de Type 2. Pour l'instant, cette hypothèse n'est valable que pour les tâches de confiance à choix forcé présentées dans cette thèse, dans lesquelles deux décisions de Type 1 doivent être comparées. La généralisation de cette approche à d'autres formes de notations de la confiance pourrait s'avérer plus difficile, mais devrait être prise en compte dans de futurs travaux expérimentaux et computationnels.

Intuitivement, une meilleure précision devrait conduire à une plus grande confiance : si l'attention augmente la précision, la confiance devrait également augmenter. Néanmoins, la littérature est encore divisée sur ce point. Comme nous l'avons vu précédemment, une des raisons potentielles de cette division est l'absence d'une mesure quantitative viable du temps dans la manipulation attentionnelle. Nous avons ici défini l'aspect temporel de la relation entre l'attention et la confiance comme une limite qui affecte à la fois le biais et la sensibilité d'un jugement de confiance. Cette limite est basée sur la notion d'épisodes sélectifs empruntée à la littérature de l'attention temporelle (voir la section 2.2.2 de l'introduction générale) : orienter l'attention vers un emplacement donné déclenche un épisode de sélection de type gaussien, étalé dans le temps. L'attention est donc considérée comme ayant une réelle dynamique, avec différents moments ou états : (a) le processus d'orientation, ou engagement, pendant lequel l'attention est allouée; (b) la sélection, avec une meilleure qualité d'encodage au pic; (c) le désengagement, au cours duquel l'attention cesse d'être efficace. Dans les paragraphes suivants, nous détaillerons comment le pic de l'épisode de sélection affecte la confiance d'une manière différence de ses « limites » (c'est-à-dire les processus d'orientation et de désengagement, respectivement). Nous montrerons qu'une simple compréhension de la confiance comme étant limitée par un épisode attentionnel peut expliquer bon nombre des résultats rapportés dans la présente dissertation.

Nous commencerons notre interprétation de la façon dont le temps affecte la relation entre l'attention et la confiance, en considérant la confiance au sommet de l'épisode attentionnel. Lorsque nous considérons le pic de cet épisode, l'attention tant exogène qu'endogène a eu un impact positif sur la confiance. Au chapitre 1, lorsque l'asynchronie indice-cible a été utilisée pour maximiser l'effet de l'attention exogène, la capacité métacognitive n'a pas souffert : l'augmentation précoce de la précision a été suivie d'une augmentation de la confiance et le rapport métacognitif (méta-d'/d') est resté stable dans le temps. Ce résultat confirme que la confiance est capable de s'adapter à l'effet de la capture involontaire de l'attention spatiale et montre que lorsque l'attention est orientée de façon exogène, la capacité métacognitive n'est pas altérée. Il en allait de même pour l'attention spatiale endogène : au chapitre 2, la confiance était plus grande immédiatement après le pic de sélection et diminuait de façon monotone par la suite. Lorsque l'on considère la chronologie d'un épisode attentionnel, le pic de sélection détermine la relation de trois manières : (a) la sous-confiance pour une deuxième cible qui partage le même « pic » qu'une cible précédente ; b) la forte corrélation entre la probabilité de signalement et la confiance moyenne ; (c) et l'effet de la sélection retardée sur la confiance. Le pic est donc positif pour la confiance, et la dissection de la ligne temporelle attentionnelle permet une compréhension approfondie de la façon dont la confiance suit les performances pendant la période la plus forte d'un épisode attentionnel. Premièrement, nous aborderons le biais de sous-confiance observé durant le phénomène dit du « lag-1 sparing ». Malgré un biais très fort favorisant une confiance plus faible lorsque deux cibles partageaient le même épisode attentionnel, nous avons malgré tout systématiquement observé une plus grande confiance pour les réponses correctes. Par conséquent, le partage d'un épisode attentionnel entre deux cibles distinctes pourrait être interprété comme source de confusion principalement au stade de la décision, mais ne semble pas supprimer la capacité métacognitive. La sous-confiance observée souligne ainsi l'importance de l'épisode de sélection dans la construction du biais de confiance.

Dans un deuxième temps, nous pouvons maintenant considérer la forte corrélation entre la probabilité de sélectionner un stimulus (notez que cela est différent de la précision, qui est la probabilité de sélectionner le stimulus cible) et la confiance (fig.2, A et B de la discussion générale). La confiance semblait ici reproduire strictement la probabilité de sélectionner un élément donné au fil du temps, et cette limite a également été observée lorsque le pic de sélection était mal placé (fig.4 du chapitre 3) : même lorsque l'élément au pic de l'épisode attentionnel n'était pas la cible, la confiance était toujours supérieure pour cet élément et diminuée pour les éléments plus éloignés du pic de l'épisode attentionnel (même lorsque le participant a finalement sélectionné l'élément cible correct, voir le chapitre 3). Cela montre que la confiance était moins sensible à la sélection de l'élément correct (ou à la précision) qu'à la probabilité de sélection (ou à l'effet que l'attention avait sur les stimuli). Plutôt que d'utiliser le terme générique de métacognition, nous pourrions plutôt distinguer deux sortes de capacités métacognitives qui permettent une forte corrélation entre la probabilité de sélection et la confiance : la métacognition basée sur l'erreur, qui est la différence de confiance pour les erreurs et les réponses correctes (fig. 2C de la discussion générale), et la métacognition basée sur la position, qui est l'augmentation significative de la confiance pour l'élément le plus fréquemment sélectionné (fig. 2B de la discussion générale).

Cette dichotomie nous amène à notre dernier point : l'effet de la sélection retardée sur la confiance. Même si la métacognition basée sur l'erreur reste la seule mesure véritablement objective de la capacité métacognitive dans le contexte actuel, il convient de noter que la métacognition basée sur la position est néanmoins informative. Cette dernière met en évidence le lien fort que la métacognition cultive avec l'attention temporelle : lorsque l'attention était mal placée, la confiance accordait systématiquement des poids différents à chaque erreur, comme si l'erreur n'était jamais survenue. Cette incapacité de la confiance à prendre en compte le retard de l'attention temporelle a également été observée pour l'attention spatiale (chapitre 4). Pour l'orientation spatiale à la fois exogène et endogène, la confiance a complètement ignoré la latence de l'attention. La sélection attentionnelle est donc un déterminant majeur de la confiance, avant la précision même : lorsque l'attention est efficacement allouée dans l'espace et dans le temps, elle reflète la précision, mais lorsque l'attention est mal placée, cette efficacité cessera d'exister, entraînant une diminution de la sensibilité métacognitive. L'examen de la chronologie du processus nous permet ainsi de disséquer les effets différentiels des signaux utilisés par les jugements de Type 2.

La relation positive que nous avons trouvée entre la confiance et le centre de masse de l'épisode attentionnel prédit que la confiance devrait être plus faible aux « limites » de l'épisode (voir fig. 2B de la discussion générale), mais elle ne nous dit pas grand-chose la nature du schéma de sensibilité métacognitive. Dans le chapitre 2, nous avons utilisé un paradigme à double tâche pour induire un épisode attentionnel endogène lors d'une première tâche, et pour sonder l'effet spatial résiduel de cet épisode sur la seconde tâche. Nous avons appelé ce mécanisme « désengagement » résiduel de facilitation, car le participant devait (au moins partiellement) se désengager de l'emplacement initial pour réussir la deuxième tâche. Dans cette expérience, nous avons observé une diminution de la confiance, reflétant la diminution de l'effet attentionnel. Il est important de noter que la capacité métacognitive a également diminué pour les sondes plus éloignées de la cible initiale, démontrant que l'épisode attentionnel joue un rôle dans la capacité métacognitive globale. Ainsi, il semble que la confiance moyenne et la capacité métacognitive agissent toutes deux de manière similaire aux frontières d'un épisode attentionnel. Dans le chapitre 4, nous avons étudié la capacité métacognitive à l'autre extrémité du spectre, c'est-à-dire lors de l'orientation de l'attention. Malgré des retards de sélection différents selon l'état du processus de l'attention, la précision de la réponse ne semblait pas affectée différemment par l'orientation précoce (condition « pré-cue ») ou tardive (condition « endogène ») de l'attention. Des analyses supplémentaires fondées sur des modèles computationnels ont confirmé la robustesse de ces estimations de précision (voir le matériel supplémentaire présenté au chapitre 4). Enfin, nous avons trouvé que la capacité métacognitive est diminuée spécifiquement *durant* l'orientation de l'attention volontaire. L'orientation initiale de l'attention endogène dans l'espace perturberait ainsi la métacognition. Encore une fois, cette interaction précoce entre l'attention et la confiance fournit la preuve du rôle crucial de l'épisode attentionnel dans le façonnement de la capacité métacognitive.

Dans la section précédente, nous avons vu que les épisodes perceptuels affectaient le processus de décision de Type 2 via des décalages de biais et façonnaient également la métacognition, c'est-à-dire la capacité même de la confiance à refléter les performances de Type 1. En utilisant le cadre théorique d'épisodes perceptifs sélectifs, il semble possible d'expliquer la diversité des résultats empiriques de cette thèse. Cette compréhension « dynamique » de l'attention et de la confiance est née de l'usage de manipulations temporelles systématiques de l'attention à travers nos paradigmes, ce qui faisait jusqu'à présent défaut dans la littérature. Pourtant, en concentrant notre travail sur le concept d'épisodes attentionnels, nous n'avons pas utilisé des conditions dans lesquelles l'attention était vraiment « absente » : l'attention était seulement décalée (comme au chapitre 3), retardée (chapitre 4) ou parfois diminuée via une orientation spatiale non valide (chapitres 1 et 2). Une attention retardée peut être considérée à première vue comme un certain type d'inattention. Mais cette hypothèse ne résiste pas à un examen approfondi. Dans la présente dissertation, nous avons montré que la confiance reste capable de faire la distinction entre différentes erreurs lorsque l'attention est mal allouée, mais cette métacognition basée sur l'attention ne peut pas être mesurée à l'aide de descripteurs standard basés sur la précision (chapitres 3 et 4 et section 2.1 de la discussion générale). Ainsi, la signification précise de ce qu'un expérimentateur sélectionne comme métrique de précision pour le Type 1 est cruciale : observer une plus grande confiance pour les cibles manquées est une chose, mais comprendre pourquoi une telle incohérence de Type 2 se produit est une étape importante vers la compréhension de la confiance. La capacité de la confiance à reproduire la probabilité de report suggère qu'il ne s'agit probablement pas d'un cas d'inattention complète. Cependant, la différence entre une attention retardée et une inattention complète reste à étudier et cette lacune ne permet donc pas de fournir des informations le comportement de la métacognition en l'absence d'attention. Bien que l'aspect de l'inattention n'ait pas fait l'objet de la thèse actuelle, il reste un aspect important à considérer. On pourrait être tenté d'appliquer l'approche épisodique du couple attention-confiance à la compréhension de l'inattention, en considérant que les « limites » de l'épisode équivalent approximativement à l'absence d'attention. Une telle interprétation prédit une forte baisse de la capacité métacognitive durant les épisodes d'inattention. Cependant, cela ne refléterait pas l'inattention en soi, mais plutôt un état d'attention plus faible ou plus « diffuse ». Pour sonder la métacognition durant un véritable état d'inattention, un paradigme doit être soigneusement conçu pour éviter de multiples effets confondants, décrits dans les sections précédentes (voir les sections 2.3.1 et 2.3.2 de l'introduction générale).

Dans ces derniers paragraphes, nous avons proposé un compte rendu épisodique de l'attention et de la confiance, dans lequel la capacité métacognitive est modelée par les épisodes attentionnels. Cette interprétation fait de l'attention un pourvoyeur crucial de signal Type 1 pour les jugements de Type 2 et souligne la nécessité de contrôler systématiquement les effets attentionnels potentiels dans les expériences de confiance.

CONCLUSION

Dans cette thèse, nous avons capitalisé sur la manipulation de la structure temporelle de l'attention sélective pour étudier la relation entre la confiance et la précision dans la prise de décision perceptuelle. Nous avons observé que la confiance est très sensible à la dynamique temporelle de l'attention sélective, au point de se dissocier de la performance dans certains cas. Le travail empirique présenté dans cette thèse souligne ainsi l'importance de l'attention sélective dans la construction de la confiance visuelle et contribue à la compréhension de la nature exacte du signal utilisé lors des jugements métacognitifs.

RÉSUMÉ

La dynamique du monde qui nous entoure nécessite sans cesse d'adapter nos décisions à son incertitude latente. Cette incertitude définit autant notre perception que le fonctionnement même de nos fonctions cognitives. La « métacognition » d'un individu - la manière dont il raisonne sur ses propres perceptions - peut être étudiée en comparant sa confiance à la qualité objective de ses décisions perceptives. Parce que l'attention sélective est une source importante de modulation sensorielle, une bonne métacognition des effets de l'attention sur la perception semble primordiale. La façon dont la confiance émerge du processus d'orientation de l'attention, et se développe ensuite dans l'espace et le temps, fait l'objet de cette thèse. Nous y décrivons notamment la solide dépendance que la confiance cultive à l'égard de l'attention visuelle, une dépendance qui subsiste à chaque étape du processus attentionnel. Les travaux expérimentaux présentés dans cette thèse suggèrent ainsi une dépendance si forte qu'une orientation erronée de l'attention passe souvent inaperçue au niveau métacognitif. Ces résultats témoignent de l'incapacité de la confiance à prendre en compte certaines des limites temporelles de l'attention sélective.

MOTS CLÉS

Métacognition, confiance, attention sélective, décision perceptive

ABSTRACT

Adaptive decision-making requires precise monitoring of decision quality in light of both sensory uncertainty and the variability inherent in cognitive functions. Such monitoring, or metacognitive reasoning, can be assessed by relating subjective confidence in a perceptual decision to objective accuracy. Selective attention is a known modulator of sensory processing, and reliable metacognitive access to attention may be the key to cope with the variability of the environment. The present dissertation investigates the temporal construction of visual confidence during and after the allocation of selective attention either to a point in time (temporal attention) or to a point in space (spatial attention). In both the temporal and spatial domain, we observe that attention constrains metacognitive ability, both during and after allocation. The robust temporal binding observed in the present thesis between attention and metacognition induces dissociations between confidence and accuracy when attention is misallocated. The empirical results presented in this work highlight a systematic inability to integrate the temporal dynamic of selective attention into metacognitive judgments.

KEYWORDS

Metacognition, confidence, selective attention, perceptual decision