Fachbereich Erziehungswissenschaft und Psychologie

der Freien Universität Berlin

DEGREES OF SUBJECTIVE CERTAINTY: PROCESSING AND REPRESENTATION

Dissertation

Zur Erlangung des akademischen Grades

Doctor of Philosophy (Ph.D.)

Vorgelegt von

Mag. rer. nat. Johannes Freiherr Heereman von Zuydtwyck

Berlin, 2016

Erstgutachter: Prof. Dr. Hauke R. Heekeren

Zweitgutachter: Prof. Dr. Dr. Henrik Walter

Tag der Disputation: 19.05.2016

ACKNOWLEDGMENTS

First and foremost, I wish to thank my supervisor Hauke Heekeren for his excellent advice, made-to-measure guidance, and unconditional loyalty.
Thanks Hauke for always having my back. You are an amazing mentor and the best science-sensei I could have had!

I am also full of gratitude to my second supervisor Henrik Walter for his valuable feedback and trans-disciplinary mind-set in particular. The discussions we had mean a lot to me.

Next I want to thank my coauthor Tim Pleskac for all the inspiring meetings we had. I enjoyed each one of them.

Finally, I want to thank the Konrad-Adenauer-Stiftung, the Berlin School of Mind and Brain, and their wonderful personnel.

Without you the last years would not have been what they were.

February 2016

CONTENTS

ZUSAMMENFASSUNG

Bin ich mir wirklich sicher? Diese Frage stellen wir uns wohl täglich. Aber wenn wir uns fragen, wie sicher wir uns sind, machen wir das bevor oder nachdem wir uns entscheiden? Mit ziemlicher Sicherheit hat man schon beides getan und tatsächlich gibt es empirische Evidenz für die Verarbeitung subjektiver Sicherheit bevor und nachdem wir uns entscheiden. Bislang ist jedoch unklar, ob, und wenn ja, wie, die entsprechenden Prozesse sich voneinander unterscheiden. Darüber hinaus ist auch noch unbekannt, wie genau sich die Berechnung subjektiver Sicherheit und der Entscheidungsprozess zueinander verhalten. Ebenfalls unklar ist, in welchem Format subjektive Sicherheit auf neuronaler Ebene repräsentiert ist. In den drei dieser Arbeit zugrundeliegenden Studien untersuche ich diese Fragen zu Repräsentation und Verarbeitung von Graden subjektiver Sicherheit.

In Studie 1 habe ich zwei perzeptuelle Entscheidungsaufgaben mit funktioneller Magnet Resonanz Tomographie (fMRT) kombiniert und gezeigt, dass Grade subjektiver Sicherheit auf neuronaler Ebene in aufgabenunabhängiger Weise repräsentiert sind. Das heißt, ich habe eine neuronale Repräsentation subjektiver Sicherheit gefunden, die über Aufgaben hinweg konstant ist. Dies deutet darauf hin, dass Grade subjektiver Sicherheit in einem abstrakten Format repräsentiert sind, also unabhängig von ihrem Gegenstand.

In Studie 2 habe ich mit einer perzeptuellen Entscheidungsaufgabe gezeigt, dass subjektive Sicherheit in Fehlertrials vor Entscheidungen höher ist als nach Entscheidungen. Gleichzeitig bleibt aber subjektive Sicherheit in korrekten Trials unberührt vom Zeitpunkt des Ratings. Dies deutet darauf hin, dass im Gegensatz zu subjektiver Sicherheit vor Entscheidungen, Sicherheit nach Entscheidungen zusätzlich auf metakognitive Fehlerentdeckungsmechanismen zurückgreift. Somit weist subjektive Sicherheit nach Entscheidungen ein metakognitives Element auf, was für subjektive Sicherheit vor Entscheidungen nicht der Fall ist.

In Studie 3 zeige ich, dass Verarbeitung auf einer zweiten Ebene, also nach Verarbeitung der ersten Antwort (Sicherheitsrating oder Entscheidung), sowohl für subjektive Sicherheit als auch für Entscheidungen stattfindet. Weiterhin zeige ich, dass die Verarbeitung von subjektiver Sicherheit mehr Rechenaufwand erfordert als die von Entscheidungen. Abschließend legen meine Ergebnisse nahe, dass Sicherheits- und Entscheidungsverarbeitung weitgehend entkoppelt sind.

Insgesamt liefert die vorliegende Arbeit neuartige Einsichten in das repräsentationale Format subjektiver Sicherheit, eine konzeptuelle Klärung von subjektiver Sicherheit als Sicherheit in der persönlichen Einschätzung eines Sachverhaltes und eine empirische Abgrenzung zu Nachbarkonzepten wie Metakognition. Dadurch setzt sie Rahmenbedingungen für die weitere Theorienbildung zu Graden subjektiver Sicherheit.

ABSTRACT

Am I really sure? This is a question we ask ourselves practically every day. But when you ask yourself how confident you are, do you ask yourself before or after you make a choice? Most likely you have experienced both and in fact there is good evidence for both pre- and post-decision confidence processing. It remains unclear, however, whether and how these processes differ from each other. More generally, it is an open question how exactly the processing of confidence and choice relate to each other. Finally, it is an open question in which format confidence is represented on the neural level. In the three studies constituting this thesis, I address these issues of representation and processing.

In study one, by combining two perceptual decision tasks with functional Magnetic Resonance Imaging (fMRI), I provide neural evidence for a task-independent representation of degrees of subjective certainty (i.e., a neural representation of subjective certainty prior to choice that remains constant across tasks). The results indicate that confidence is represented in an abstract format, i.e. independent of its referent.

In study two, I use a perceptual decision task to show that confidence in incorrect trials is higher prior to choice than after choice. At the same time, confidence in correct trials remains unaffected by variation of the response order. This indicates that in contrast to pre decision confidence, post decision confidence is informed by metacognitive error detection. While the former is based on incoming evidence and best viewed as degree of epistemic belief, the latter additionally draws on higher order metacognitive monitoring of the decision process. Therefore, confidence understood as degree of epistemic belief prior to choice is not associated with metacognition, while after choice it is.

In study three, I show second stage processing, i.e. continued processing after the first response, of both confidence and choice. In addition, I show that confidence and choice differ in their computational demands, as indicated by longer response times in ratings than

decisions. Finally, my results suggest that confidence and choice processing are largely decoupled.

Taken together, the present thesis provides novel insights into the representational format of confidence, a conceptual clarification of confidence as degree of belief and an empirical dissociation from related concepts, such as metacognition. Thereby, this thesis adds basic conditions to future theorizing about degrees of belief.

LIST OF ORIGINAL RESEARCH ARTICLES

This dissertation is based on the following articles:

Heereman, J., Walter, H., Heekeren, H.R. (2015). A task-independent neural representation of subjective certainty in visual perception. *Frontiers in Human Neuroscience*. 9:551. doi: 10.3389/fnhum.2015.00551

Heereman, J., Heekeren, H.R., Pleskac, T.J. (*under review*). The Difference between Pre- and Post-Decision Confidence. *Psychonomic Bulletin & Review.*
http://link.springer.com/journal/13423
This article may not exactly replicate the final version published. It is not the copy of record.

Heereman, J., Heekeren, H.R., Pleskac, T.J. (*submitted*). Evidence for Decoupled Processing of Confidence and Choice. *Journal of Experimental Psychology: Human Perception and Performance.*
http://www.apa.org/pubs/journals/xhp/index.aspx
Version: 13.02.2015
This article may not exactly replicate the final version published in the APA journal. It is not the copy of record.

1. GENERAL OVERVIEW

What does it depend on whether you believe something or not? Whether you believe it more or believe it less? Once your premises are certain, this question is trivial and logic delivers the answer. For example, if you know that all men are mortal and that Socrates is a man, you are pretty safe to believe that Socrates will sooner or later die. The problem is the premises. Nothing can prove the premises. So how do we arrive at a degree of belief in the beginning? One way to address this question is to study the properties and mechanisms of the process, which relates features of the external world to internal subjective degrees of conscious belief: the study of perception.

Logicians such as Frank Plumpton Ramsey (Ramsey, 1931) and John Maynard Keynes (Keynes, 1952) engaged in an intense dispute about what degrees of belief actually are and whether they are measurable in real numbers. Others have questioned if there are degrees of belief at all (see Kyburg, 2003). However, when logicians speak of degrees of belief, they usually mean *rational* degrees of belief in the context of probability theory. Remarkably, Blaise Pascal (Pascal & Havet, 1852) and Laplace (Laplace & Dale, 1995) have already stressed repeatedly, that empirical degrees of belief deviate from the rationality-criteria of subjective probability. Given that the subject of the present work is empirical belief, the formalization of belief as probability appears inappropriate here.

Still, the theory of probability, in its subjective version, **is** the theory of degrees of belief. While the standard frequentist view holds that probability refers to the relative frequency of events (i.e. objective probability), Bayesians such as Bruno deFinetti have been arguing that degree of belief is the only sensible meaning of probability and that there is no such thing as objective probability (De Finetti, 1970; for an alternative view see Ramsey, 1931). It is this latter subjective notion of probability we are concerned with here. Crucially, this view makes the implicit assumption that degrees of belief can be meaningfully considered independent of their referent. That is, it is assumed that the dimension ‘degree of belief’ is independent of what is believed in. To the best of my knowledge, this important assumption has received no empirical attention so far. Nevertheless, experimental scientists already began to study related questions about belief in the early days of psychophysics. Often, they just called it differently: confidence. Therefore, in the following I use the terms “degree of belief”, “degree of subjective certainty”, and “confidence” largely interchangeably.

a. Origins and current state of the experimental investigation of confidence

To the best of my knowledge, the quantitative experimental study of confidence begins with the work *On small Differences in Sensation* (Peirce & Jastrow, 1884). They adapted a letter scale (= Briefwaage) in a way, which allowed them (via placing different weights on it) to apply different pressures on the subjects’ finger. Subjects had to decide, which of two consecutive pressures was more intense, i.e. which of the two weights placed on the letter scale was heavier. For response collection the experimenters used a playing card and punched holes (0, 1, 2, and 3) in the four corners which served to indicate the degree of confidence (0 = guess, 3 = very confident). To indicate their decision (what they believe) subjects put the card face-up or face-down. To indicate their confidence (how much they believe it), they oriented the respective corner towards the experimenter.

A couple of years later Sumner (1898) set out to study „one's subjective feeling of sureness “. He asked his subjects questions such as „Is the earth round? “, „Is there life on other heavenly bodies? “, or „Is the world becoming better? “. The questions were on separate cards and participants had to arrange the questions in the order of the certainty with which they felt they could answer them. Interestingly, he did not ask subjects for their answer (i.e. choice). He just asked for subjects’ degree of conviction (i.e. confidence) concerning their answers to the questions. Whether rating confidence is associated with implicit choice anyways and whether asking for choice (or not) makes a difference for confidence estimation are open questions that I address in Study two and three. Williamson (1915) presented statements such as "There exists an all-wise Creator of the world", “Only the good die young", or “The whale swallowed Jonah" to his subjects. Similar to Sumner (1898), in a part of Williamson’s experiments subjects had to arrange the statements in an order of belief-strength. In addition, he let subjects compare pairs of such statements and indicate, in which one they believed more. Next, subjects had to rate (on a 4-point-scale) their degree of confidence in their decision. Since then, for nearly a century the decision followed by confidence paradigm has practically had a monopoly on the study of confidence. This methodological domination has even led to the common but wrong (as recently shown by Gherman & Philiastides, 2015; Kiani & Shadlen, 2009) assumption, that confidence is computed only after choice (also see Baranski & Petrusic, 1998).

Also for nearly a century, in striking similarity to Peirce and Jastrow (1884), the standard approach to the study of confidence has been the use of perceptual two-alternative-forced-choice-tasks (2AFC) where task difficulty is manipulated via stimulus discriminability. Standard measures are performance (i.e. choice accuracy), confidence, and response time. These measures are characterized by multiple interdependencies. Confidence decreases as the difficulty level increases, decision response time increases with decreasing confidence and increasing difficulty, accuracy and confidence are positively correlated, and rating response time (when the confidence rating is the second response) decreases with difficulty and increases with confidence (See Figure 1 and Moran, Teodorescu, and Usher (2015) for a complete list of established empirical relationships and discussion).

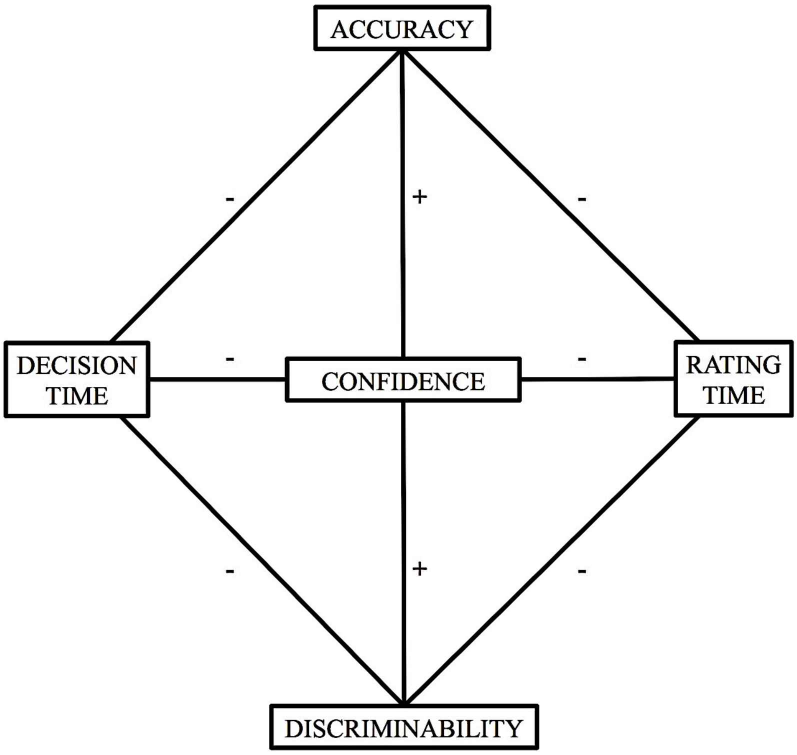


FIGURE 1: Established relationships between confidence, discriminability, accuracy, decision time, and rating time; ± signs represent direction of correlation.

Crucially, it is thus far unclear whether these relationships withstand methodological departure from the classic approach (i.e. choice then confidence). That is, it is unclear in how far these relations generalize to alternative ways of measurement. For example, whether design-changes like a flip of the response order (i.e. to confidence then choice) preserve these relations remains an open question. Also, different types of confidence are often lumped together.

b. Two types of confidence

McDougall (1921) stated that “belief and doubt are upon the plane of intellectual striving while […] confidence and anxiety are upon the plane of practical action.” Similarly, cognitive psychologists have long been dissociating two kinds of confidence: While Type 1 confidence is object related and best expressed as degree of belief in an event or hypothesis (the former pair in McDougall's statement), Type 2 confidence is action related and best expressed as degree of belief in the correctness of a cognitive process, such as decision making (the latter pair in McDougall's statement). To use a statistical analogy: A confidence interval relates to a parameter, not to the correctness of the analysis. A confidence interval has nothing to say about whether you made a mistake in your calculations. Correspondingly, a decision-related confidence rating (Type 2) has nothing to do with a confidence interval, while a stimulus-related confidence rating (Type 1) does indeed. However, when running your analysis, you will probably double-check both, the parameter estimate and whether there is a mistake in your analysis (i.e. error detection). Interestingly, due to the traditional monopoly of the choice followed by confidence paradigm, these processes have only been studied after a decision has been made. Consequently, it is thus far unclear how confidence (both Type 1 and 2) prior to and after choice relate to each other. To address this and other issues, several different ways of confidence measurement have been used.

2. THE MEASUREMENT OF CONFIDENCE

There are both implicit and explicit measures of confidence. In the former, some indirect measure is used as a proxy for confidence, while in the latter, subjects are directly asked.

a. Implicit measures

One popular way of implicit confidence measurement is the use of an opt-out-task (e.g. Gherman & Philiastides, 2015; Kiani & Shadlen, 2009). Subjects perform a two-alternative-forced-choice-task where they receive a reward for correct choices but no reward for incorrect choices. On half of the trials subjects are offered a third alternative, i.e. to opt out of the decision and receive a smaller but certain reward. So if subjects are very confident, the best strategy is to ignore the third option (certain but smaller reward) and try to get the higher reward. If confidence is low the best strategy is to opt out and take the smaller but certain reward. The way subjects use this opt-out option (or not) is taken as a proxy for confidence.

A second recently developed implicit measure of confidence (only used in rodents so far) is subjects' willingness to wait for a reward after choice (e.g. Kepecs, Uchida, Zariwala, & Mainen, 2008; Lak et al., 2014). In this paradigm, subjects also receive a reward for correct choices but none for incorrect ones. In a sub portion of correct trials the reward is omitted. The time animals wait at the reward port before aborting the trial is taken as proxy for confidence. While this approach currently appears to be the best available method of confidence-measurement in non-primates, the contribution of factors such as simple patience to the confidence-data is unclear.

A third implicit measure of confidence is asking subjects to place a bet on the outcome of a decision (e.g. Persaud, McLeod, & Cowey, 2007; Ramsey, 1931). The more confident subjects are the higher should be the bet they place. Here, the amount placed on the bet is taken as proxy for confidence.

One problem these measures have in common is a potential reward confound. Especially when combining behavioral with neural measures the researcher cannot know whether the neural measures are representing confidence, expected reward or a mixture of both. Also, confidence data might be confounded by individual risk-preferences and loss-aversion. These may interact with confidence in unknown ways (for a detailed treatment see Fleming & Dolan, 2010). An additional problem with implicit confidence measures is that they are agnostic about the degree of conscious awareness subjects have with regard to their degree of subjective certainty. For example, in some trials an opt-out or bet might be consciously reflected while in others it might not. Therefore, what exactly is being measured, conscious belief or something else, remains unknown.

b. Explicit measures

The most common way of direct confidence measurement in humans is to ask subjects for post-decision confidence ratings (e.g. Fleming, Huijgen, & Dolan, 2012; Henmon, 1911; Peirce & Jastrow, 1884; Pleskac & Busemeyer, 2010). Alternatively, some have recorded choice and confidence simultaneously, i.e. in one button press (e.g. Heereman & Walla, 2011; Henson, Rugg, Shallice, & Dolan, 2000). The disadvantage of these direct measures (from a behaviorist's point of view) is that they are subjective. The main advantage is that they cannot be confounded by expected reward or risk-preferences. Another advantage is that in contrast to implicit measures, consciousness of confidence is granted. Furthermore, in the present work we introduce variation of the choice-confidence response order as an experimental manipulation and directly compare confidence and choice response times (Study 3). Both approaches require holding the response mode constant. Explicit ratings are the best way to accomplish that. Therefore, in the present work I decided to use explicit ratings.

3. PROCESS MODELS OF CONFIDENCE

Confidence is a direct function of both stimulus discriminability and choice accuracy and varies inversely with response time. Beginning in the late 19th century, researchers started to formalize these relationships between performance and confidence (e.g. Peirce & Jastrow, 1884), decision response time and confidence (e.g. Audley, 1960; Ratcliff, 1978; Volkmann, 1934), as well as discriminability (i.e. evidence) and confidence (Vickers, 1979). Ratcliff (1978) for example proposed a model of decision making in which two conflicting evidence streams drive a random walk towards one of two thresholds (set by the subject). The drift-rate of the walk is determined by the discriminability of the alternatives. Because in this model the only available information about discriminability is response time, Ratcliff suggested that confidence is an inverse function of the subject's response time. Vickers (1979) proposed an accumulator model of decision-making, in which stimulus differences are sampled over time. Positive and negative differences are accumulated in two separate accumulators until one of the two reaches a preset threshold. Confidence is determined by the difference between the two accumulators at the time of choice, i.e. the *balance-of-evidence*.

Most current models of confidence originate from this sequential sampling framework and are based on Wald's (1945) sequential probability ratio test (e.g. Moran et al., 2015; Pleskac & Busemeyer, 2010; Ratcliff & Starns, 2013). Importantly, in contrast to previous models, they manage to capture all known relationships between discriminability, accuracy and response time. They differ along several dimensions (see below) but have some key characteristics in common. They model confidence as a second decision and confidence is primarily related to evidence (i.e. discriminability) rather than RT or accuracy.

Another interesting feature of these models is, that confidence is not tied to the evidence-state at the time of choice. For example, in Pleskac & Busemeyer's (2010) two-stage dynamic signal detection theory (2DSD) evidence-accumulation is allowed to continue after choice, so that choice and confidence are not necessarily based on the same evidence. Another example is the Collapsing Confidence Boundary Model (CCB) (Moran et al., 2015). This model makes the assumption that after choice, confidence boundaries collapse as a function of time. This assumption helps to explain, why the time to determine confidence is inversely related to the degree of confidence. Put simply, if evidence is too weak, decision bounds collapse and come the accumulator's way. Another alternative model has been proposed by Ratcliff and Starns (2013). These authors model confidence ratings as a multiple-choice decision process. One problem with this model is that it includes one accumulator for each confidence level. So if in an experiment there are five levels of confidence, there are five competing accumulators. Crucially, for cases where confidence ratings are continuous (instead of discrete) or implicit (instead of explicit) the model makes no predictions. This is why I exclude it from further treatment.

The main conceptual difference among sequential sampling models is, that they either assume one accumulator for each decision alternative or one accumulator integrating the evidence-difference between alternatives. In order to model confidence, approaches with one accumulator need to assume that the confidence rating is a second decision. In contrast, models with two accumulators don't need that assumption because confidence can be modeled as the difference between the two accumulators (e.g. Vickers, 1979). In the former class of models, confidence and choice are two instantiations of basically the same decision mechanism, whereas in the latter they are not. Interestingly, these assumptions about confidence and choice, as well as the computational properties (such as timing) they imply, have not been directly compared so far. That is, it is thus far unclear whether there is a difference in the processes underlying confidence and choice and how they relate to each other.

4. CONFIDENCE AND THE BRAIN

Only recently have confidence-scientists begun to combine behavioral approaches with neuroscientific methods, such as functional magnetic resonance imaging (fMRI) (e.g. Fleck, Daselaar, Dobbins, & Cabeza, 2006), transcranial magnetic stimulation (TMS) (e.g. Rahnev, Maniscalco, Luber, Lau, & Lisanby, 2012), single-cell-recordings (Kepecs et al., 2008; Kiani & Shadlen, 2009), cortical microstimulation (Fetsch, Kiani, Newsome, & Shadlen, 2014), and local temporal inactivation of brain function using muscimol (Komura, Nikkuni, Hirashima, Uetake, & Miyamoto, 2013; Lak et al., 2014). Below, I summarize research with these neuroscientific approaches on different organisms.

a. Humans

Surprisingly, only a few fMRI studies have investigated neural correlates of decision confidence. The brain area most consistently found to code for decision confidence in humans is the dorsomedial prefrontal cortex (DMPFC) (Fleck et al., 2006; Fleming et al., 2012; Hebart, Schriever, Donner, & Haynes, 2014). Other areas reported to carry a confidence signal in humans include the dorsolateral prefrontal cortex (DLPFC) (Fleck et al., 2006), the ventral striatum and the right anterior insula (Hebart et al., 2014) as well as right posterior parietal cortex (Fleming et al., 2012). How these representations relate to the task at hand and which format they use is an open question. While some authors report evidence supporting the view that confidence is represented in a task-independent format (De Gardelle & Mamassian, 2014), others report evidence suggesting that it is represented in a task-specific format (Kiani & Shadlen, 2009). Crucially, task-independence would imply that confidence is represented in an abstract format. In study one, I address this fundamental question about the neural representation of confidence using fMRI.

Others have used electroencephalography (EEG) to study confidence in the brain and thereby shed more light on the temporal properties of confidence processing. For example, Gherman and Philiastides (2015) showed that a confidence-signal emerges as early as the decision process itself in lateral prefrontal and parietal cortices. Similarly, Zizlsperger, Sauvigny, Händel, and Haarmeier (2014) showed that a cortical perceptual confidence-signal is present as early as 300ms after stimulus onset and evolves in parallel to stimulus representations.

Rounis, Maniscalco, Rothwell, Passingham, and Lau (2010) applied TMS to the DLPFC, which lead to decreased confidence without affecting choice accuracy. In contrast, applying TMS over the visual cortex appears to increase confidence while decreasing accuracy (Rahnev et al., 2012). These results indicate, that the relation between the mechanisms underlying confidence and choice differs across processing stages.

b. Animals

Just as in humans, animal studies have revealed several different brain regions associated with confidence. While Kiani and Shadlen (2009) report that in monkeys, decision confidence is computed along with the decision in sensory-motor neurons in the lateral intraparietal area (LIP), Kepecs et al. (2008) found that in rodents a signal which is correlated with confidence is represented in a population of orbitofrontal neurons. Others identified processes related to confidence in monkeys' pulvinar and lateral geniculate nucleus of the thalamus (Komura et al., 2013).

Similar to the study of confidence in humans, several animal studies have also sought to dissociate the neural processing of choice from the processing of confidence. While some found confidence and choice to be processed via the same neural populations and the same mechanisms (Fetsch et al., 2014; Kiani & Shadlen, 2009), others found areas uniquely contributing to confidence but not choice (Komura et al., 2013; Lak et al., 2014). To investigate causal relations between neural processing and behavior, Fetsch et al. (2014) applied cortical microstimulation to area LIP and reported no differential effect on confidence and choice. In contrast, Lak et al. (2014) showed that temporal inactivation of the orbitofrontal cortex (using g-aminobutyric acid (GABA) receptor agonist muscimol) disrupts confidence reports without affecting decision accuracy. Similarly, inactivation of the pulvinar nucleus decreased monkeys' confidence without affecting categorization performance (Komura et al., 2013).

Altogether, the available evidence strongly suggests that in both humans and animals, some processing components are shared by confidence and choice, while others are specifically related to confidence processing. In both humans and animals, neural correlates of confidence are widely distributed across the brain and it seems, that the neural basis of confidence is not a strictly localized phenomenon.

5. RESEARCH QUESTIONS

The general aim of this thesis is to deepen the understanding of how a degree of belief, i.e. confidence is processed and represented in humans. Particular attention is given to the relation between the processes underlying confidence and choice. In three studies I investigated the following questions:

1. Is there a task-independent neural representation of subjective certainty? That is, is there a representation of degrees of belief independent of its referent as suggested by probability theory? (Study 1)

2. Do pre- and post-decision confidence differ in their underlying processes?
   Specifically, we ask whether pre- and post-decision confidence differ in that the latter includes error detection while the former does not. (Study 2)

3. Does asking for confidence influence decision processing and/or does asking for a decision influence confidence processing? (Study 2)

4. Is there further decision processing after confidence has been estimated, i.e. post confidence decision processing? (Study 3)

5. Does rating confidence require implicit choice? (Study 3)

6. Do confidence and choice processing differ in their computational demands? (Study 3)

6. GENERAL METHODOLOGY

In the following, I will outline the general methodology of the three studies constituting this thesis. Please refer to the methods sections of the three studies for a complete description of methodological details.

a. Tasks and stimuli

Throughout this work, I used perceptual two-alternative-forced-choice tasks combined with confidence ratings. In study one, participants performed a color-motion detection task (modified from Kayser, Erickson, Buchsbaum, & D'Esposito, 2010). The key manipulation was that we combined two tasks and depending on the instruction cue, subjects had to attend motion and ignore color or attend color and ignore motion. After stimulus presentation, subjects had to rate their confidence that motion was to the left or right, or that one of two colors was more present in the stimulus (see Figure 2).

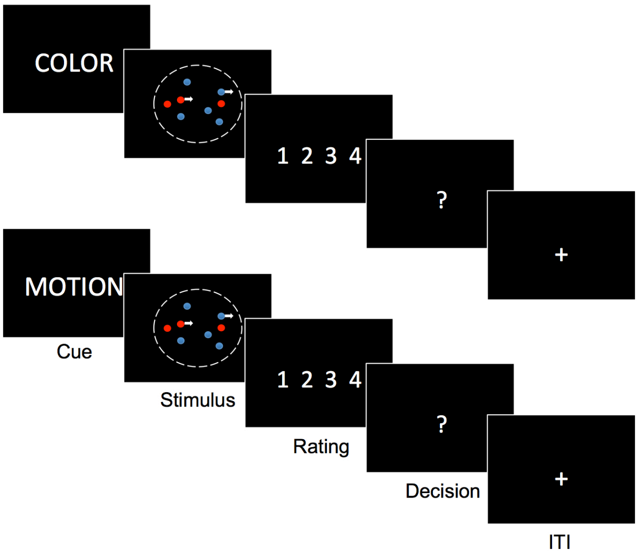


FIGURE 2: Task description: Subjects see on the display a cloud of colored moving dots. In the beginning of each trial they are cued whether to attend motion and ignore color or attend color and ignore motion. Depending on the cue after stimulus presentation (750ms) they have to rate their degree of certainty that motion was to the left or right or that one color was more present than the other (blue or red). Only after the certainty rating, they indicate the respective direction or color.

In study two and three I used a face-car categorization task (used in e.g. Philiastides, Auksztulewicz, Heekeren, & Blankenburg, 2011) where I manipulated difficulty via stimulus visibility (see Figure 3 for sample stimuli). The two key manipulations here were that I 1.) varied whether subjects only had to indicate choice, only confidence, or both and 2.) varied the order of confidence and choice response.

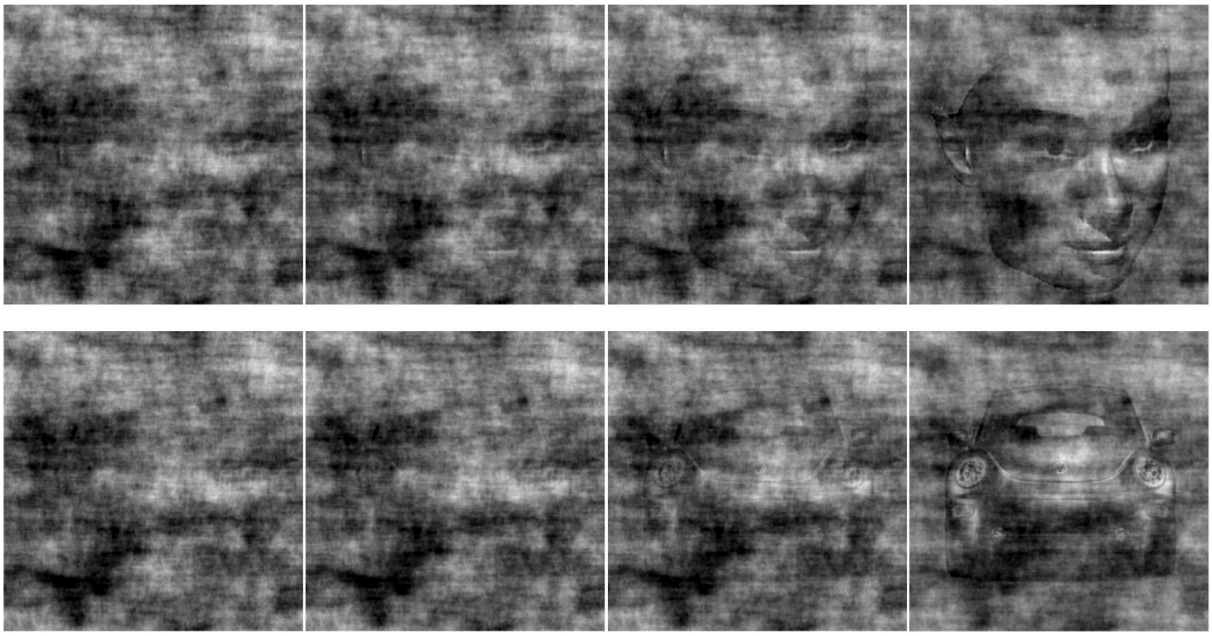


FIGURE 3: Sample stimuli used in Study 2 and 3. Face and car grayscale images equated for spatial frequency, luminance, contrast and magnitude spectrum. Visibility (i.e. phase spectrum) was manipulated using the weighted mean phase (WMP) technique (Dakin, Hess, Ledgeway, & Achtman, 2002).

b. Variables controlled for (quantities held constant)

Depending on the goal of a confidence study, two main strategies have been used: either to keep the stimulus constant or to keep performance constant (e.g., using a staircase procedure). Researchers interested in subjects' introspective ability to evaluate their performance (e.g. Fleming et al., 2012; Lau & Passingham, 2006), usually use staircase-methods to keep performance constant throughout the whole experiment (because they are interested in the relation between confidence and performance). This implies that stimulus values vary. Others, like me, are mainly interested in the relation between stimulus and confidence, so I kept stimulus intensities constant throughout the experiments.

c. Subject-specific stimulus calibration

In order to ensure that mean performance (or confidence) is similar across subjects and to avoid floor and ceiling effects, previous to testing stimulus intensities are usually calibrated for each individual subject. Because most researchers in the field are interested in the relationship between performance and stimulus-difficulty or performance and confidence, the usual approach is to calibrate stimulus intensities to predefined levels of performance. In study two and three I was interested in response times as well as the relation between confidence and performance, so I followed this standard approach.

However, in study one the quantity of interest was confidence, independent of performance and stimulus intensity. Therefore, here I developed a novel calibration procedure where I calibrated stimulus intensities to a predefined level of confidence (and not to a level of performance). To the best of my knowledge in this work I am the first to calibrate to confidence.

d. Measures

1. Behavioral

In study two and three I investigated processes underlying confidence judgments using choice and rating response times, as well as the relation between correctness of choice and confidence. Therefore, throughout all three experiments I measured and analyzed the relative frequency of correct responses (i.e. choice accuracy), the time between stimulus presentation and the first response (i.e. response time 1 (RT1)), the time between the first and second response (i.e. response time 2 (RT2)), and explicit discrete confidence ratings.

In study one, I combined these behavioral measures with fMRI.

2. Functional Magnetic Resonance Imaging

Functional magnetic resonance imaging is an indirect measure of local neural activity. Indirect, because instead of neural activity it relies on task-related changes in local oxygen concentration (Logothetis, 2008; Poldrack & Farah, 2015). The basic approach is to regress behavioral data against fMRI-data, i.e. blood oxygenation level dependent (BOLD) signal, using a general linear model (GLM). However, conventional fMRI analysis methods only offer information about 'where' activity is significantly correlated with the independent variable. In contrast, my approach to analyze neural processes underlying subjective certainty in different perceptual tasks and then perform a conjunction analysis (logical AND) on the respective activation maps allows for inference about properties of the neural representation of confidence in general, i.e. inference beyond localization. This approach appears the most appropriate way to investigate the task-dependency of confidence representation at the neural level.

In addition, the modern study of confidence is highly interdisciplinary, resulting in a variety of qualitatively different results. For example, one will face substantial difficulties when trying to relate results from a purely behavioral study in humans to neural recordings in animals. At the same time, integration of results from the human and animal literature is of utmost importance, because results tend to complement each other. The use of fMRI enables to relate the results to both. For my purposes, fMRI therefore clearly is the method of choice.

7. STUDY SUMMARIES

a. Study 1

Heereman, J., Walter, H., Heekeren, H.R. (2015). A task-independent neural representation of subjective certainty in visual perception. *Frontiers in Human Neuroscience*.

Am I really sure? This is a question not only scientists ask themselves but practically everybody everyday. Previous neuroimaging studies identified neural correlates of decision confidence (e.g. Fleck et al., 2006; Fleming et al., 2012) but the dependence of this neural representation on a particular task remains unclear. A recent study (De Gardelle & Mamassian, 2014) provided behavioral evidence supporting the view that one's subjective confidence in a decision is represented in a task-independent format. Here, combining two perceptual decision tasks with fMRI, I therefore asked whether there is a task-independent neural representation of degrees of subjective certainty (i.e. a neural representation of subjective certainty that remains constant across two tasks). Twenty participants performed a visual color/motion detection task in which they had to rate their perceptual confidence. Importantly, to keep the results independent of task-difficulty and stimulus properties, I used one constant stimulus intensity throughout the whole experiment. I analyzed the two trial types (color and motion trials) separately and used the confidence ratings to parametrically modulate the respective main regressors. On the resulting contrasts I then performed conjunction analyses.

My results show that a substantial part of the neural representation of degrees of subjective perceptual certainty is task-invariant (i.e., a neural representation of subjective certainty that remains constant across two tasks). I found that in two different tasks changes in BOLD signal increased with subjective certainty in the right lingual, calcarine, and left angular gyrus (see Figure 4). BOLD signal decreased with increasing subjective certainty in the left lingual Gyrus, right inferior parietal lobule, bilateral DMPFC/SMA, and left postcentral gyrus (see Figure 5).

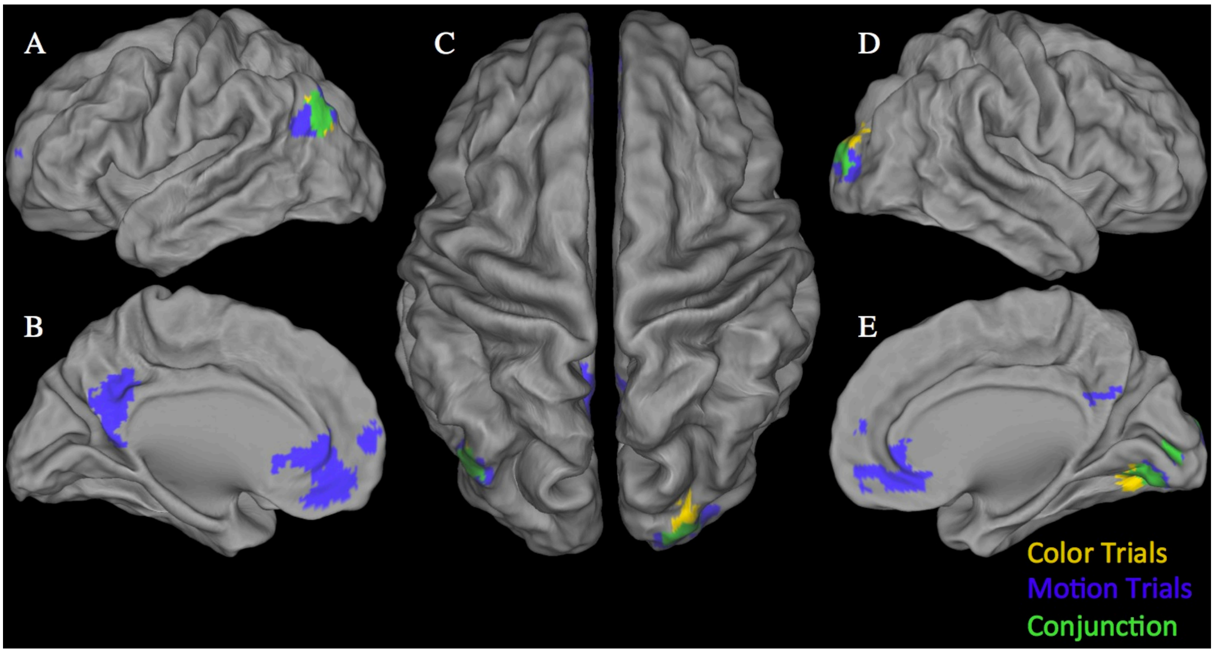


Figure 4: Positive parametric effects in color (yellow) and motion (blue) trials and their conjunction (green). **(A)** Lateral view of left Hemisphere, **(B)** Medial view of left Hemisphere, **(C)** Dorsal View, **(D)** Lateral view of right Hemisphere, and **(E)** Medial view of right Hemisphere. Cluster-defining threshold $p < 0.001$, uncorrected; reported changes in BOLD signal corrected for multiple comparisons at $p < 0.05$, whole brain corrected. Significant conjunctions in: Calcarine, lingual, fusiform, and angular gyrus ($n = 20$, see Study 1 for details).

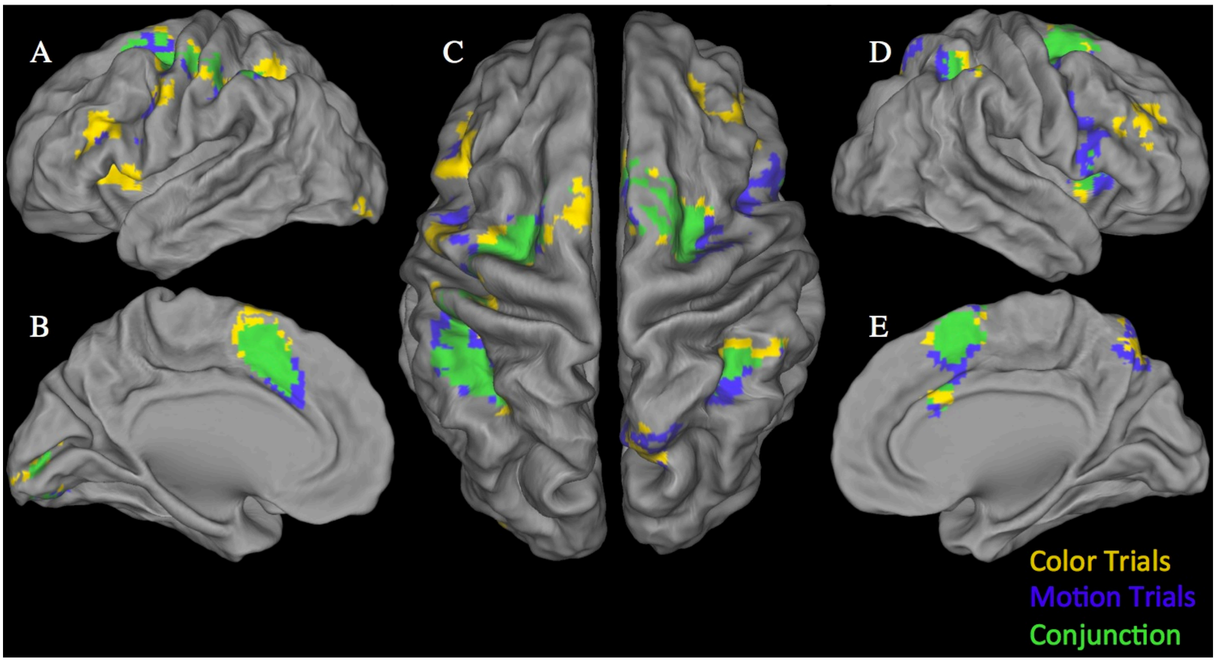


Figure 5: Negative parametric effects in color (yellow) and motion (blue) trials and their conjunction (green). **(A)** Lateral view of left Hemisphere, **(B)** Medial view of left Hemisphere, **(C)** Dorsal View, **(D)** Lateral view of right Hemisphere, and **(E)** Medial view of right Hemisphere. Cluster-defining threshold $p < 0.001$, uncorrected; reported changes in BOLD signal corrected for multiple comparisons at $p < 0.05$, whole brain corrected. Significant conjunctions in: DMPFC, postcentral and lingual gyrus, and insula ($n = 20$, see Study 1 for details).

These changes in BOLD signal were virtually identical in the two tasks. My data therefore support the view, that there is a central module in the brain processing subjective certainty and that, in line with De Gardelle and Mamassian (2014), degrees of subjective certainty are represented in a task-independent format. In conclusion my data provide strong evidence for a task-independent neural representation of subjective certainty.

b. Study 2

Heereman, J., Heekeren, H.R., Pleskac, T.J. (*under review*). The Difference between Pre- and Post-Decision Confidence. *Psychonomic Bulletin & Review.*

When you ask yourself how confident you are, do you ask yourself before or after you commit to choice? Most likely you have experienced both and in fact there is good evidence for both pre-decision (e.g. Gherman & Philiastides, 2015; Zylberberg, Barttfeld, & Sigman, 2012) and post-decision confidence processing (e.g. Baranski & Petrusic, 1998; Pleskac & Busemeyer, 2010). Yet, while pre-decision confidence may have an impact on post-rating decisions (but not vice versa), pre-rating decisions may have an impact on post-decision confidence (but not vice versa). That is, committing an error may reduce post- but not pre-decision confidence. In this study (n=27), using a perceptual face-car-detection task, I show that post-decision confidence ratings outperform pre-decision confidence ratings in predicting choice accuracy (Figure 6). More specifically, pre-decision confidence in incorrect trials is greater than post-decision confidence in incorrect trials. This strongly suggests the involvement of error detection in post-decision confidence processing.

I also show, that not only the requirement of post-decision ratings (Petrusic & Baranski, 2003), but also the requirement of post-rating decisions leads to increased primary response times (RT1) (as compared to choice-only and rating-only, respectively). The former finding has been taken as support for the view that asking for confidence influences the decision process. One interpretation of my finding would be that asking for choice influences confidence processing. However, while it is unclear whether there is a change in the underlying processes, it is clear that there is a change on the motor level. Crucially, Sternberg, Monsell, Knoll, and Wright (1978) and Klapp (1995) showed, that RT1 increases with the number of required consecutive responses. In my view, this suggests that increased RT1 might be due to additional demands on motor-response planning rather than changes in the underlying computational processes.

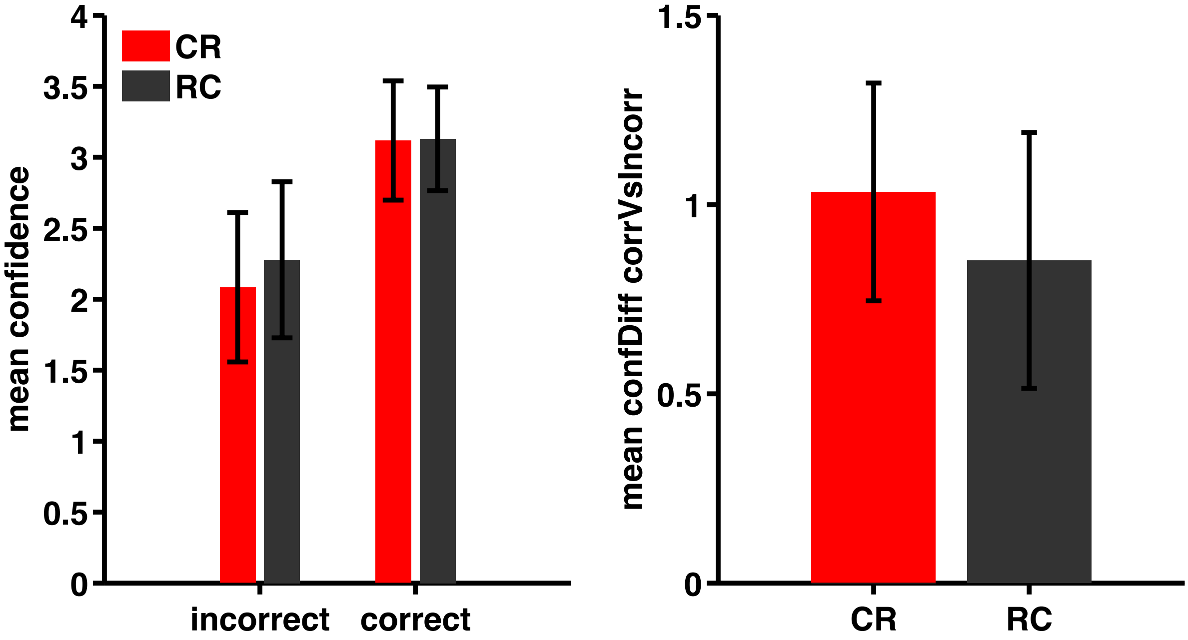


Figure 6: Confidence in correct vs. incorrect trials in choice-then-rating (CR) vs. rating-then-choice (RC) condition. Left: confidence in incorrect trials is higher in the RC than in the CR condition. Right: Confidence-difference between incorrect and correct trials is higher in CR than in RC-condition, indicating the involvement of an error-detection mechanism in post-decision-ratings. (n = 27, error bars represent SD)

In conclusion, my results challenge the view that the requirement of post-decision confidence ratings influences the decision process. Rather, my data support the view that the requirement of a second response (post-choice rating or post-rating choice) increases advance motor-sequence-planning-demands and thereby RT1. Most importantly, here I show that there is in fact a difference between pre- and post-decision confidence ratings: post-decision confidence outperforms pre-decision confidence in dissociating correct from incorrect trials. This is most likely due to post-decision error detection and implies a conceptual difference between pre- and post-decision confidence ratings: while in the case of ratings following choice we would speak of a metacognitive process (or at least a process involving a metacognitive component) in the rating-then-choice case we would not.

c. Study 3

Heereman, J., Heekeren, H.R., Pleskac, T.J. (*submitted*). Evidence for Decoupled Processing of Confidence and Choice. *Journal of Experimental Psychology: Human Perception and Performance.*

After you commit to a decision, do you sometimes reconsider it? There is good evidence for post-decision processing of confidence (Baranski & Petrusic, 1998; Pleskac & Busemeyer, 2010), so your answer is probably yes. It remains unclear, however, whether the other way around, decision processing may continue after rating confidence. Using a perceptual face-car-categorization task, I show that not only post-decision rating response times, but also post-rating decision response times, decrease with difficulty (see Figure 7). This suggests second stage processing of both confidence and choice.

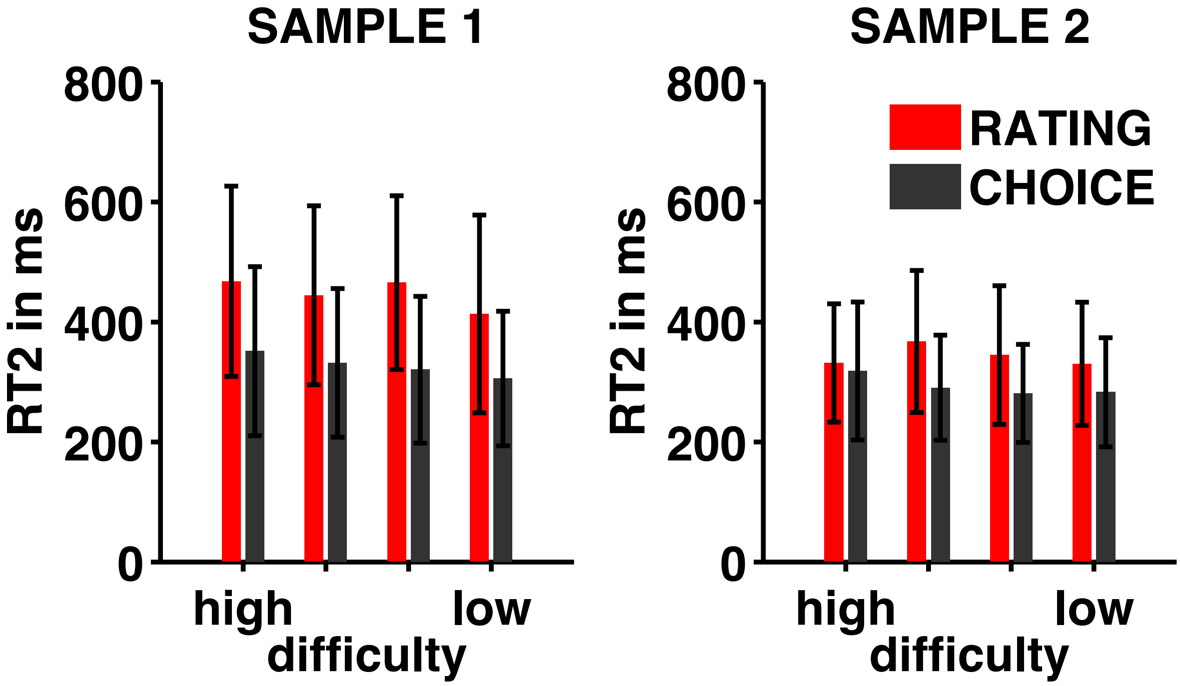


Figure 7: RT2 decreases with difficulty both when the rating and when choice is the second response suggesting not only post decision processing of confidence but also post confidence processing of choice. Also, rating confidence takes longer than choice, suggesting a difference in computational demands. Left: sample one (Study 2, n = 27). Right: sample two (Study 3, n = 27). (error bars represent SD)

In addition, I show that, controlling for the number of response alternatives, subjects are faster when they only rate their confidence alone as compared to when they have to indicate choice and confidence in a combined response. This finding indicates, that estimating confidence doesn't require implicit choice. Also, I find that rating confidence takes longer than making a decision both when the confidence rating is the first and the second response. Finally, I replicate my previous finding that not only the requirement of post-decision ratings, but also the requirement of post-rating decisions is associated with increased primary response times (as compared to choice-only and rating-only, respectively) (Heereman et al., *under review*).

One may argue, that variations of RT2 as a function of evidence are not necessarily an index of post-decision (or post-rating) processing of the second response. Alternatively, my observation may also be explained by assuming that it is not the computation but the response read-out time from a latent variable that varies as a function of evidence strength. However, the fact that RT2 is longer for confidence ratings than for decisions strongly suggests that this is not the case because else RT2 for confidence and decisions should be the same. One may also argue, that longer confidence RT1 indicates that rating confidence implies choice while choice doesn't imply confidence. However, this interpretation is inconsistent with my finding of longer RT1 in a combined choice/rating condition as compared to a rating-only condition.

8. GENERAL DISCUSSION

In this thesis I aimed to deepen the understanding of the cognitive and neural underpinnings of degrees of belief via the study of perceptual confidence. Main topics are the representation of confidence and the relationship between processing of confidence and choice.

One of four general questions of this thesis was, in which format confidence is represented on the neural level. The second was, in how far the temporal order of confidence and choice measurement has an impact on the statistical relation between the two variables. A third question was, how confidence and choice processing relate to each other. A fourth question was, whether and if yes how the processes underlying choice and confidence differ from each other. More specifically, in three studies we showed that:

1. The neural representation of degrees of subjective certainty is largely task-independent.
2. Pre- and post-decision confidence differ in their introspective accuracy (i.e. resolution). This is due to metacognitive error detection in post-decision but not pre-decision confidence ratings.
3. Asking for a second response (confidence or choice) influences motor planning rather than choice (or confidence) processing.
4. Second-stage processing of both confidence and choice.
5. Rating confidence is not associated with implicit choice.
6. Estimating confidence is associated with higher computational demands than making a decision, suggesting a difference in the underlying processes.

In the following I will focus the discussion on representation and processing of confidence. I discuss the respective implications for current models and the contribution of this work to our understanding of confidence. Particular attention will also be given to the demarcation of belief from metacognition and the implication of my results for the theory of subjective probabilities.

a. Representation of confidence

1. Task-independence

In study one, I show that a substantial part of the neural representation of degrees of subjective certainty is task-invariant (see Figure 4 and 5). My data therefore support the view, that there is a central module in the brain processing subjective certainty and that, in line with De Gardelle and Mamassian (2014), degrees of subjective certainty are represented in an abstract and task-independent format. This mode of representation appears important for several reasons. First, the global neuronal workspace model of consciousness (Dehaene & Changeux, 2011) states that consciousness provides a means of global broadcasting of information throughout the brain. Critically, this requires that the respective signal is readable for brain modules on the receiver-side.

Similarly, in the field of value-based decision making it has been argued, that in order to compare the value of different choice-options like "apples vs. pears" for example, the brain needs a common currency (e.g. Montague & King-Casas, 2007; Platt & Plassmann, 2013). In analogy, to broadcast and compare degrees of belief, these must be stored in a common currency, i.e. a format constant across referents and tasks. One may argue, that instead of a common currency, the brain may use a translation mechanism between different formats when necessary. However, such a translation mechanism would be associated with processing time. Crucially, De Gardelle and Mamassian (2014) show, that comparing confidence between two identical and between two different tasks takes the same amount of time. This renders the involvement of an on-the-fly translation mechanism instead of an abstract representation very unlikely. Finally, consider the perceptual face-car task used in study two and three: to decide, whether evidence (which roughly translates to degrees of belief) for face or evidence for car is stronger, subjects need some representation of degrees of belief which is independent of its referent.

2. Redundant representation of confidence

Another key result of study one is, that I find both, areas showing a positive and areas showing a negative parametric effect of subjective certainty (see Figure 3 and 4). Among these I find both activations and deactivations as compared to baseline. So it appears, as if confidence is processed and represented in several different ways in parallel. That is, despite its inherent inefficiency there appears to be a substantial degree of diverse redundancy. While homogenous redundancy refers to systems, where identical components work in parallel, diverse redundancy refers to systems where different components with different functional principles work in parallel. Both forms of redundant systems produce the same results for the same inputs and both forms are associated with the advantage of increased system stability and increased reliability of outputs. But while a mechanism where certainty computation is realized by one principle alone (i.e. homogenous redundancy) is still vulnerable and prone to error, a mechanism where certainty is computed in several different ways in parallel (i.e., diverse redundancy) is very robust against system perturbation.

A similar kind of multi-controller processing has been suggested in dual-systems approaches to decision making (Beierholm, Anen, Quartz, & Bossaerts, 2011; Daw, Niv, & Dayan, 2005). To note however, while these authors assume the respective systems compete, in our view they should rather cooperate.

b. Confidence: epistemic belief or metacognition?

Metacognition refers to “thinking about thinking” (i.e. a metacognitive process has another cognitive process as its subject) (Fleming, Weil, Nagy, Dolan, & Rees, 2010; Nelson & Narens, 1990). In study two and three I found evidence suggesting the involvement of a metacognitive error detection component in post-decision confidence ratings. Specifically, I found post-decision confidence ratings to be more predictive of trial correctness than pre-decision ratings. That is, the resolution of post-decision confidence is superior to resolution of pre-decision confidence (see Figure 5). Post-hoc analyses revealed that this effect is driven by incorrect trials, i.e. confidence in incorrect trials was higher in the rating-then-choice condition than in the choice-then-rating condition while in correct trials I observed no difference in confidence between the two conditions. Relatedly, using EEG, Boldt and Yeung (2015) found neural markers shared by post-decision confidence and error detection (i.e. retrospective detection of incorrect choices) establishing a link between the two. This finding together with my results strongly suggests, that differences between pre- and post-decision confidence are specific for error trials and most likely due to post-decision error detection.

While it has been shown, that confidence is already computed before choice (Gherman & Philiastides, 2015), its modulation by metacognitive error detection appears to happen only after choice. This suggests, that while pre-decision confidence refers to incoming evidence, post-decision confidence refers to both incoming evidence AND the decision process (via error detection). In line with this view, Murphy, Robertson, Harty, and O'Connell (2015) show that a peri-choice signal originating in medial frontal cortex, which has been associated with error detection (Yeung, Botvinick, & Cohen, 2004), provides a source of input to post-decision confidence processing.

In reference to the metacognitive account (“thinking about thinking”) of confidence (e.g. Fleming et al., 2012; Fleming et al., 2010) Kepecs and Mainen (2012) ask: “Why should it be taken for granted that self-reported confidence judgments in humans require an instance of metacognition and uncertainty-monitoring processes?" This question is related to their definition of confidence as "the degree of belief in the truth of a proposition or the reliability of a piece of information". To clarify this issue, confidence as defined here is not necessarily associated with a monitoring-instance. In contrast, post decision confidence as degree of belief in the correctness of a decision seems to be, as indicated by confidence modulation via error detection. In my view, this additional instance is the crucial difference between pre- and post-decision confidence, and between confidence as degree of belief and confidence as metacognitive monitoring. Confidence as degree of belief is epistemologically oriented and a reliability-estimate of external input to the decision process. In contrast, confidence as metacognitive process is introspective and might be best viewed as a reliability estimate of the cognitive process "decision". While the former drives pre-decision confidence, post-decision confidence is driven by both.

In conclusion my main point is, that for researchers primarily interested in degrees of belief (like me), the post decisional metacognitive component is incidental, because it has (conceptually) nothing to do with degrees of belief in the epistemological sense.

c. Asking for confidence does not influence decisions

For the study of decision making, it is an important question whether asking for confidence influences the decision process. The answer to this question decides, in how far results from studies including both choice and confidence measurement are informative about the decision process when confidence is not asked for. Petrusic and Baranski (2003) found, that asking for post decision confidence is associated with increased RT1 (as compared to a choice-only-condition). This finding has been taken as evidence for the view that confidence computation requires additional processing and influences the decision process. However, Lebreton, Abitbol, Daunizeau, and Pessiglione (2015) report neural evidence showing that confidence is processed regardless whether asked for or not. So if it is not additional processing due to confidence computation, what is the reason for increased RT1 when confidence ratings are required after choice? Sternberg et al. (1978) and Klapp (1995) showed that RT1 increases with the number of required consecutive responses, which is due to advance motor-sequence planning. In line with these observations, in two studies I showed that not only the requirement of post-decision confidence ratings is associated with longer decision reaction times (as compared to a choice-only condition) see Figure 8). In consequence, the more parsimonious explanation for increased RT1 conditional on the requirement of a second response appears to be an increase in motor planning rather than altered choice (or confidence) processing.

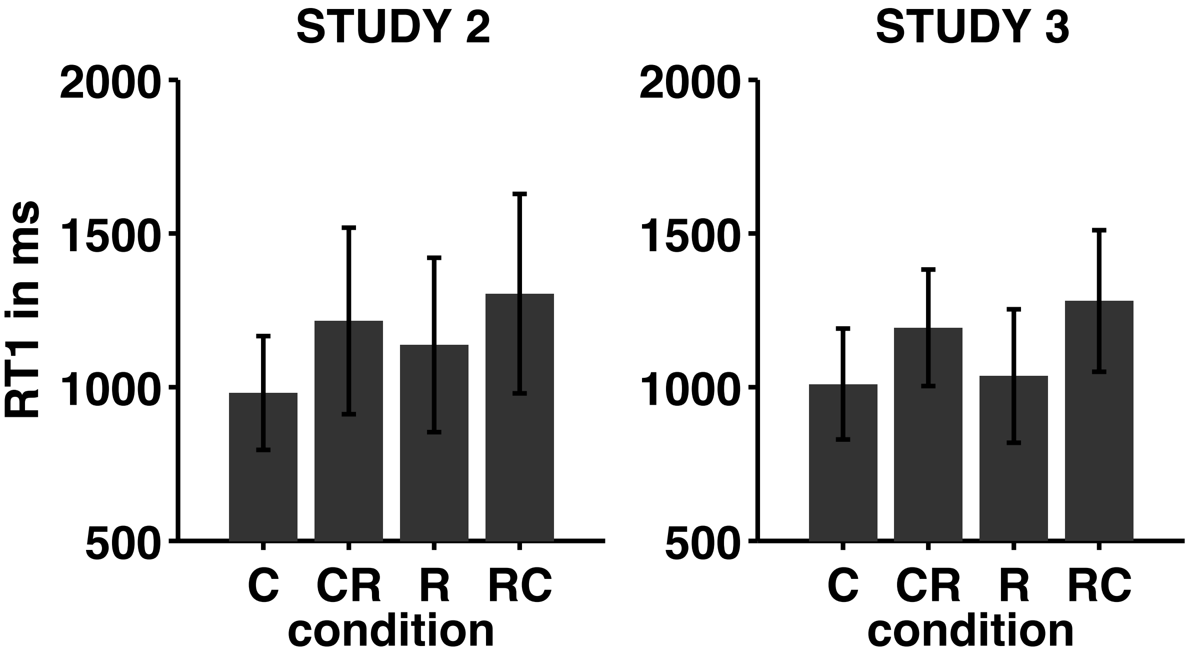


Figure 8: RT1 per condition and effect of asking for 2nd response on RT1 (C = choice vs. CR = choice-rating, R = rating vs. RC = rating-choice); the requirement of a second consecutive response is associated with prolonged RT1 in both two-response conditions. (error bars represent SD)

In conclusion, I don’t see any grounds for the claim that asking for confidence influences decision processing and therefore it is valid to consider insights from decision studies including confidence measurement in the context of pure decision studies.

d. Second-stage processing of both confidence and choice

RT2-changes as a function of stimulus-difficulty in choice-then-rating-designs are a well-replicated finding (e.g. Baranski & Petrusic, 1998; Moran et al., 2015; Pleskac & Busemeyer, 2010). To the best of my knowledge, I am the first to observe the respective effect in a rating-then-choice design. In two studies, I find that secondary response time (RT2) decreases with stimulus-difficulty, regardless whether choice or rating is the second response. That is, both post decision rating-time and post rating decision-time vary as a function of evidence (see Figure 7). This implies that in both response orders the second response (confidence rating or decision) is, at least to some degree, computed after the first response. This extends the original finding of Baranski and Petrusic (1998) concerning post-decision confidence processing to post-rating decision processing.

An alternative explanation of the observed effects would be, that in difficult trials evidence is accumulated over a longer time window than in easy trials, as indicated by RT1 variation as a function of evidence strength. A metacognitive mechanism monitoring the decision process, which most likely contributes to post-decision confidence ratings (see section 'b. Confidence: epistemic belief or metacognition') would have to integrate the monitoring signal in difficult trials over a longer time window than in easy trials. Second stage processing would therefore be associated with higher processing demands, as expressed in longer RT2 in difficult trials. Yet, if the observed RT2 effects were due to integration-time of a metacognitive mechanism, the rating process (when the confidence rating is the first response) should also be subject to metacognitive monitoring. Consequently, one would expect the resulting metacognitive signal to contribute to post-rating choice. That is, post-confidence choice accuracy should be higher than pre-confidence choice accuracy. Crucially, this is not the case.

Therefore, I conclude that not only confidence but also choice can be at least in part computed in a second stage, i.e. after the primary response.

e. Rating confidence does not require implicit choice

In several models, such as the balance-of-evidence hypothesis (Vickers, 1979), judging confidence is assumed to require implicit choice. Others (e.g. Kiani & Shadlen, 2009) have suggested that confidence and choice are computed along with each other in the same process. In study three I find evidence indicating that both views are not entirely correct. Specifically, I find that RT1 is longer when subjects indicate their decision and rating together (i.e. in one button press) as compared to when they only rate their confidence. This suggests that rating confidence is not associated with implicit choice because else RT1 would be constant across the two conditions. Longer RT1 when explicit choice is required indicates, that even when all evidence is accumulated and readily available, actually committing to choice requires an extra computational effort. This indicates that confidence and choice are not computed along with each other in the same process and that judging confidence does not require implicit choice.

f. Are confidence ratings decisions?

Are confidence and choice basically just two instantiations of fundamentally the same decision process? For example, Vickers (1979) models choice as a race-to-threshold process and confidence as the balance of evidence (i.e. difference) between two accumulators at the time of choice. So here, confidence and choice are based on different mechanisms. In contrast, Pleskac and Busemeyer (2010) and Moran et al. (2015) model post-decision confidence as a second decision. In study three I find evidence suggesting that this assumption might not be entirely correct. That is, I observed that the two processes are associated with different computational demands, as indicated by longer both RT1 and RT2 for ratings than for decisions. One may argue, that longer RT1 for ratings is due to ratings including or requiring a prior implicit decision, while decisions do not require prior confidence estimation (as suggested by Vickers (1979)). However, this interpretation is inconsistent with my finding

(Study 3) that combined confidence/choice responses take longer than ratings alone. That is, I find evidence suggesting that ratings do not necessarily require choice. Also, the fact that not only RT1 but also RT2 is longer for ratings makes this interpretation unlikely. Unlikely, because when subjects indicate confidence as the second response, a choice has already been made. So why should confidence take longer than choice, i.e. why should a post decision confidence rating require another decision? That seems very implausible. However, making the assumption, that evidence accumulation continues after choice and that the second response (confidence or choice) is based on an updated evidence state and therefore a new decision, one could accommodate longer RT2 for ratings than decisions. That is, second stage processing might be not just confidence processing, but confidence AND choice processing. Under this assumption, the choice-then-confidence paradigm would include three choices in total: choice and "confidence+choice". That is, post-decision ratings would include a second implicit choice. But again, this is inconsistent with my finding showing that ratings are not associated with implicit choice. Crucially, I also observe that the functions linking evidence to confidence and choice, respectively, differ from each other. But while it is clear that the two computations differ from each other, the specific difference is not. In summary, my data indicate that confidence and choice are either based on the same mechanism but associated with different process parameters (e.g. threshold, accumulation rate) or that the underlying mechanisms differ. But while the underlying mechanisms might be the same, the specific processing would still be associated with different parameters. Processes with different parameters are not the same, so we can conclude that confidence processing differs from decision processing. Therefore, confidence ratings are not decisions.

Future studies should elucidate whether the difference between confidence and choice processing is due to a difference in the underlying mechanism or whether the mechanisms are basically the same and it is only model parameters (such as thresholds/bounds), which vary between the processing of confidence and choice.

g. What is the purpose of computing confidence?

Our degree of confidence prior to choice is an estimate of the reliability of incoming evidence and associated with the actual decision process. It thereby determines whether we are willing to make a decision or rather opt out, i.e. decide not to decide (Gherman & Philiastides, 2015; Kiani & Shadlen, 2009). Degrees of confidence prior to choice may also guide decision making, in that when we feel unconfident, we look for more evidence or additional cues. That is pre-decision confidence influences the decision at hand and guides current behavior.

In contrast, post-decision confidence by definition cannot influence the decision at hand. So what is the purpose of it? In settings, where a decider has the opportunity to change his mind and withdraw his decision or change it, post-decision confidence can drive these changes of mind. Also, when a decider has to make several decisions in succession, post-decision confidence could serve as an estimate of in how far the agent can rely on the correctness of previous decisions, and therefore rely on the respective outcome in subsequent decisions, or not. Another potential role for post-decision confidence has been suggested by Hebart et al. (2014). The authors argue, that post-decision confidence may provide a learning signal in the absence of external feedback. To conclude, while pre-decision confidence informs the decision at hand, post-decision confidence is informed by the decision at hand and may guide how we decide in the future.

h. Implication of task-independent confidence for subjective probability

My finding of a task-independent neural representation of confidence (Study 1) comes along with an interesting implication: The standard notion of belief is a combination of a referent (subject) with a quantifier (i.e. degree of belief). The former refers to what is believed, the second refers to how much it is believed. Subjective probability theory is silent with regard to the referent. It makes the implicit assumption, that the quantifier, formally expressed as a probability, is independent of the referent. To the best of my knowledge, this assumption has not been tested. Although I don't claim to prove the validity of that assumption, my finding of a task-independent neural representation of confidence (i.e. degree of belief) suggests, that this assumption is most likely correct. While the referent (color or motion) varies, the neural representation of the quantifier (degree of belief) remains largely constant. It therefore appears, that not only in normative views but also empirically, referent and quantifier are decoupled.

j. Implications of this dissertation for current process models of confidence and choice

Vickers (1979) models choice as a race-to-threshold process and confidence as the balance-of-evidence (i.e. difference) between two accumulators at the time of choice. With the additional assumption, that the time needed for computing this difference is a function of that difference, the model can nicely capture decreasing RT2 with decreasing difficulty when confidence follows choice. However, the model cannot reproduce this pattern when choice follows confidence because in the model, choice processing must be completed prior to (or at least at the same time as) confidence computation. The collapsing confidence bounds model (CCB) (Moran et al., 2015) faces a similar problem. It makes the assumption that after choice, confidence boundaries collapse as a function of time. This assumption helps to explain, why the time to determine confidence is inversely related to the level of evidence-strength. But while CCB can reproduce the evidence-dependent RT2 pattern when confidence follows choice, it cannot when the response order is flipped. That is, in order to accommodate both RT1 and RT2 variation as a function of evidence when choice follows confidence, in addition to collapsing bounds in second stage processing, it would also have to assume collapsing bounds in first stage processing (processing of first response). RT2-variation as a function of evidence also suggests that choice and confidence are not computed in one and the same process (as proposed by Fetsch et al., 2014; Kiani & Shadlen, 2009). If that was the case, all processing should be completed at the time of the first response (confidence or choice) and RT2 should therefore be constant across levels of evidence strength.

The balance-of-evidence-hypothesis (BOE) (Vickers, 1979) also fails to capture our observation that a combined confidence/choice response takes longer than a simple rating response. This is also due to the fact, that here confidence requires prior choice, which would predict identical RT1 across these two conditions. Interestingly, when confidence follows choice (but not vice versa), BOE can accommodate that confidence-responses take longer than choice responses. This is because in BOE choice and confidence are based on different processes (race-to-threshold vs. difference between accumulators, respectively). In contrast, Pleskac and Busemeyer (2010) and Moran et al. (2015) model post-decision confidence ratings as a second decision. If this assumption was correct, one would expect confidence and choice processing to be associated with similar computational demands, as indicated by similar response times. Crucially, this is not the case. For example, CCB (Moran et al., 2015) would need to be extended, in that it would not only have to assume collapsing bounds in both first and second stage processing, it would also have to assume different bounds for confidence and choice. To account for the observed differences in RT1 and RT2, 2DSD (Pleskac & Busemeyer, 2010) would have to assume either a lower drift rate or a higher threshold for confidence than choice. To conclude, my data render the practice of modeling confidence as a decision at least questionable and I am not aware of a process model capable of accommodating my findings without substantial modifications.

k. Limitations and outlook

For specific limitations please refer to the discussion sections of the three studies. Two general limitations apply to all three studies conducted in this thesis: First, my data are agnostic with regard to the generalizability of my findings beyond the domain of visual perception. Second, in all three studies I follow a correlative approach, which is blind with regard to causal relationships between response modes, resolution and reaction time on the one hand, between neural representations and behavioral confidence on the other.

Future studies should address these issues. Also, in the future, combining neural measures such as EEG and fMRI with my approach of contrasting pre- and post-decision ratings may help disentangling processes shared by confidence and choice computation from processes that uniquely contribute to confidence. Also, in my view the large network of activations and deactivations associated with degrees of confidence (Study 1) and the plethora of brain areas found in other studies to be related to confidence ask for a graph-theoretical treatment (e.g. Bullmore & Sporns, 2009) of the subject. Also, I think that the briefly sketched potential role of diverse redundancy in processing and representation of confidence deserves a more detailed treatment. Finally, during the last two years I have been piloting experiments where evidence and counterevidence are orthogonal. Following this line of research promises a deeper understanding of the signals driving confidence and choice and to shed more light on the role of counterevidence in error detection and confidence processing in particular.

9. GENERAL CONCLUSION

In this thesis, I showed that confidence is largely represented in a task-independent format and therefore at least in part decoupled from its referent. Furthermore, the difference between confidence prior to choice and after choice is that the latter includes error detection, while the former does not. The former is best viewed as degree of epistemic belief while the latter is epistemic belief integrated with higher order metacognitive monitoring of the decision process. Confidence understood as degree of epistemic belief does not require metacognition. Also, I found that confidence and choice processing are at least in part decoupled and do not necessarily require each other. Relatedly, not only confidence but also choice can be processed in a second stage. In addition, I show that confidence is associated with higher computational demands than choice and that the functions linking evidence to these two variables, respectively, differ from each other. These results suggest, that confidence and choice are based on different processes.

Altogether, the present dissertation advances the understanding of the neural underpinnings, the representational format, and the processing of decision confidence. Finally, this thesis provides a much-needed conceptual and empirical dissociation of confidence as epistemic degree of belief from the concept of metacognition. This thesis thereby adds to the basic conditions of future theorizing about degrees of belief.

REFERENCES

Audley, R. (1960). A stochastic model for individual choice behavior. *Psychological review, 67*(1), 1.

Baranski, J. V., & Petrusic, W. M. (1998). Probing the locus of confidence judgments: experiments on the time to determine confidence. *Journal of Experimental Psychology: Human Perception and Performance, 24*(3), 929.

Beierholm, U. R., Anen, C., Quartz, S., & Bossaerts, P. (2011). Separate encoding of model-based and model-free valuations in the human brain. *NeuroImage, 58*(3), 955-962.

Boldt, A., & Yeung, N. (2015). Shared Neural Markers of Decision Confidence and Error Detection. *The Journal of Neuroscience, 35*(8), 3478-3484.

Bullmore, E., & Sporns, O. (2009). Complex brain networks: graph theoretical analysis of structural and functional systems. *Nature Reviews Neuroscience, 10*(3), 186-198.

Dakin, S., Hess, R., Ledgeway, T., & Achtman, R. (2002). What causes non-monotonic tuning of fMRI response to noisy images? *Current Biology, 12*(14), R476-R477.

Daw, N. D., Niv, Y., & Dayan, P. (2005). Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. *Nature neuroscience, 8*(12), 1704-1711.

De Finetti, B. (1970). Logical foundations and measurement of subjective probability. *Acta Psychologica, 34*, 129-145.

De Gardelle, V., & Mamassian, P. (2014). Does confidence use a common currency across two visual tasks? *Psychological science*, 0956797614528956.

Dehaene, S., & Changeux, J.-P. (2011). Experimental and theoretical approaches to conscious processing. *Neuron, 70*(2), 200-227.

Fetsch, C. R., Kiani, R., Newsome, W. T., & Shadlen, M. N. (2014). Effects of cortical microstimulation on confidence in a perceptual decision. *Neuron, 83*(4), 797-804.

Fleck, M. S., Daselaar, S. M., Dobbins, I. G., & Cabeza, R. (2006). Role of prefrontal and anterior cingulate regions in decision-making processes shared by memory and nonmemory tasks. *Cerebral Cortex, 16*(11), 1623-1630.

Fleming, S. M., & Dolan, R. J. (2010). Effects of loss aversion on post-decision wagering: Implications for measures of awareness. *Consciousness and cognition, 19*(1), 352-363.

Fleming, S. M., Huijgen, J., & Dolan, R. J. (2012). Prefrontal contributions to metacognition in perceptual decision making. *The Journal of Neuroscience, 32*(18), 6117-6125.

Fleming, S. M., Weil, R. S., Nagy, Z., Dolan, R. J., & Rees, G. (2010). Relating introspective accuracy to individual differences in brain structure. *science, 329*(5998), 1541-1543.

Gherman, S., & Philiastides, M. G. (2015). Neural representations of confidence emerge from the process of decision formation during perceptual choices. *NeuroImage, 106*, 134-143.

Hebart, M. N., Schriever, Y., Donner, T. H., & Haynes, J.-D. (2014). The relationship between perceptual decision variables and confidence in the human brain. *Cerebral Cortex*, bhu181.

Heereman, J., & Walla, P. (2011). Stress, uncertainty and decision confidence. *Applied psychophysiology and biofeedback, 36*(4), 273-279.

Henmon, V. A. C. (1911). The relation of the time of a judgment to its accuracy. *Psychological review, 18*(3), 186.

Henson, R. N., Rugg, M. D., Shallice, T., & Dolan, R. J. (2000). Confidence in recognition memory for words: dissociating right prefrontal roles in episodic retrieval. *Cognitive Neuroscience, Journal of, 12*(6), 913-923.

Kayser, A. S., Erickson, D. T., Buchsbaum, B. R., & D'Esposito, M. (2010). Neural representations of relevant and irrelevant features in perceptual decision making. *The Journal of Neuroscience, 30*(47), 15778-15789.

Kepecs, A., & Mainen, Z. F. (2012). A computational framework for the study of confidence in humans and animals. *Philosophical Transactions of the Royal Society B: Biological Sciences, 367*(1594), 1322-1337.

Kepecs, A., Uchida, N., Zariwala, H. A., & Mainen, Z. F. (2008). Neural correlates, computation and behavioural impact of decision confidence. *Nature, 455*(7210), 227-231.

Keynes, J. M. (1952). *A Treatise on Probability*. London: Macmillan.

Kiani, R., & Shadlen, M. N. (2009). Representation of confidence associated with a decision by neurons in the parietal cortex. *science, 324*(5928), 759-764.

Klapp, S. T. (1995). Motor Response Programming during Simple and Choice-Reaction Time - the Role of Practice. *Journal of Experimental Psychology-Human Perception and Performance, 21*(5), 1015-1027. doi:Doi 10.1037//0096-1523.21.5.1015

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, 16*(6), 749-755. doi:10.1038/nn.3393

Kyburg, H. E. (2003). Are there degrees of belief? *Journal of Applied Logic, 1*(3), 139-149.

Lak, A., Costa, G. M., Romberg, E., Koulakov, A. A., Mainen, Z. F., & Kepecs, A. (2014). Orbitofrontal cortex is required for optimal waiting based on decision confidence. *Neuron, 84*(1), 190-201.

Laplace, P.-S., & Dale, A. I. (1995). *Philosophical Essay on Probabilities* (Vol. 13): Springer Science & Business Media.

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.

Lebreton, M., Abitbol, R., Daunizeau, J., & Pessiglione, M. (2015). Automatic integration of confidence in the brain valuation signal. *Nat Neurosci, 18*(8), 1159-1167. doi:10.1038/nn.4064

Logothetis, N. K. (2008). What we can do and what we cannot do with fMRI. *Nature, 453*(7197), 869-878.

McDougall, W. (1921). Belief as a Derived Emotion. *Psychological review, 28*(5), 315.

Montague, P. R., & King-Casas, B. (2007). Efficient statistics, common currencies and the problem of reward-harvesting. *Trends in cognitive sciences, 11*(12), 514-519.

Moran, R., Teodorescu, A. R., & Usher, M. (2015). Post choice information integration as a causal determinant of confidence: Novel data and a computational account. *Cognitive psychology, 78*, 99-147. doi:10.1016/j.cogpsych.2015.01.002

Murphy, P. R., Robertson, I. H., Harty, S., & O'Connell, R. G. (2015). Neural evidence accumulation persists after choice to inform metacognitive judgments. *eLife*, e11946.

Nelson, T. O., & Narens, L. (1990). Metamemory: A theoretical framework and new findings. *The psychology of learning and motivation, 26*, 125-141.

Pascal, B., & Havet, E. (1852). *Pensées*: Dezobry et E. Magdeleine.

Peirce, C. S., & Jastrow, J. (1884). On small differences of sensation. *Memoirs of the National Academy of Sciences, 3*, 73-83.

Persaud, N., McLeod, P., & Cowey, A. (2007). Post-decision wagering objectively measures awareness. *Nature neuroscience, 10*(2), 257-261.

Petrusic, W. M., & Baranski, J. V. (2003). Judging confidence influences decision processing in comparative judgments. *Psychonomic bulletin & review, 10*(1), 177-183.

Philiastides, M. G., Auksztulewicz, R., Heekeren, H. R., & Blankenburg, F. (2011). Causal role of dorsolateral prefrontal cortex in human perceptual decision making. *Curr Biol, 21*(11), 980-983. doi:10.1016/j.cub.2011.04.034

Platt, M. L., & Plassmann, H. (2013). *Multistage Valuation Signals and Common Neural Currencies.* Paper presented at the Elsevier Inc.

Pleskac, T. J., & Busemeyer, J. R. (2010). Two-stage dynamic signal detection: a theory of choice, decision time, and confidence. *Psychol Rev, 117*(3), 864-901. doi:10.1037/a0019737

Poldrack, R. A., & Farah, M. J. (2015). Progress and challenges in probing the human brain. *Nature, 526*(7573), 371-379.

Rahnev, D. A., 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, 107*(6), 1556-1563.

Ramsey, F. P. (1931). Truth and probability (1926). *The foundations of mathematics and other logical essays*, 156-198.

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological review, 85*(2), 59.

Ratcliff, R., & Starns, J. J. (2013). Modeling confidence judgments, response times, and multiple choices in decision making: recognition memory and motion discrimination. *Psychological review, 120*(3), 697.

Rounis, E., Maniscalco, B., Rothwell, J. C., Passingham, R. E., & Lau, H. (2010). Theta-burst transcranial magnetic stimulation to the prefrontal cortex impairs metacognitive visual awareness. *Cognitive Neuroscience, 1*(3), 165-175.

Sternberg, S., Monsell, S., Knoll, R. L., & Wright, C. E. (1978). The latency and duration of rapid movement sequences: Comparisons of speech and typewriting. *Information processing in motor control and learning*, 117-152.

Sumner, F. B. (1898). A statistical study of belief. *Psychological review, 5*(6), 616.

Vickers, D. (1979). *Decision processes in visual perception*: Academic Press.

Volkmann, J. (1934). The relation of the time of judgment to the certainty of judgment. *Psychological Bulletin, 31*, 672-673.

Wald, A. (1945). Sequential tests of statistical hypotheses. *The Annals of Mathematical Statistics, 16*(2), 117-186.

Williamson, G. F. (1915). Individual differences in belief, measured and expressed by degrees of confidence. *The Journal of Philosophy, Psychology and Scientific Methods*, 127-137.

Yeung, N., Botvinick, M. M., & Cohen, J. D. (2004). The neural basis of error detection: conflict monitoring and the error-related negativity. *Psychological review, 111*(4), 931.

Zizlsperger, L., Sauvigny, T., Händel, B., & Haarmeier, T. (2014). Cortical representations of confidence in a visual perceptual decision. *Nature communications, 5*, 3940. doi:10.1038/ncomms4940

Zylberberg, A., Barttfeld, P., & Sigman, M. (2012). The construction of confidence in a perceptual decision. *Frontiers in integrative neuroscience, 6*.

EIDESSTATTLICHE ERKLÄRUNG

Hiermit erkläre ich an Eides statt,

- dass ich die vorliegende Arbeit selbständig und ohne unerlaubte Hilfe verfasst habe,

- dass ich mich nicht bereits anderwärts um einen Doktorgrad beworben habe und keinen Doktorgrad in dem Promotionsfach Psychologie besitze und

- dass ich die zugrundeliegende Promotionsordnung vom 02.12.2008 kenne.

Berlin, den 13.02.2015 Johannes Freiherr Heereman

RESEARCH ARTICLES

Study 1

Heereman, J., Walter, H., Heekeren, H.R. (2015). A task-independent neural representation of subjective certainty in visual perception. *Frontiers in Human Neuroscience*. 9:551. doi: 10.3389/fnhum.2015.00551

frontiers
in Human Neuroscience

ORIGINAL RESEARCH
published: 09 October 2015
doi: 10.3389/fnhum.2015.00551

A task-independent neural representation of subjective certainty in visual perception

Johannes Heereman[1,2]*, *Henrik Walter*[2,3] *and Hauke R. Heekeren*[1,2]*

[1] *Biological Psychology and Cognitive Neuroscience, Department of Education and Psychology, Freie Universitaet Berlin, Berlin, Germany,* [2] *Berlin School of Mind and Brain, Humboldt Graduate School, Humboldt-Universitaet zu Berlin, Berlin, Germany,* [3] *Division of Mind and Brain Research, Department of Psychiatry and Psychotherapy, Campus Mitte, Charité Universitaetsmedizin Berlin, Berlin, Germany*

Am I really sure? This is a question not only scientists ask themselves but practically everybody every day. A recent study provides behavioral evidence supporting the view that one's subjective confidence in a decision (i.e., feeling sure that a decision is correct) is represented in a task-independent format. Previous neuroimaging studies identified neural correlates of decision confidence but whether or not these are task-dependent remains unclear. Here, combining two perceptual decision tasks with functional magnetic resonance imaging (fMRI), we provide neural evidence for a task-independent representation of degrees of subjective certainty (i.e., a neural representation of subjective certainty that remains constant across two visual tasks). Importantly, due to the constant stimulus-intensity used this result is independent of task-difficulty and stimulus properties. Our data provide strong evidence for a generic mechanism underlying the computation of subjective perceptual certainty in vision.

Keywords: decision making, confidence, subjective certainty, perception, fMRI

Edited by:
John J. Foxe,
Albert Einstein College of Medicine,
USA

Reviewed by:
Stefan Pollmann,
Otto-von-Guericke University
Magdeburg, Germany
Britt Anderson,
University of Waterloo, Canada

****Correspondence:***
Johannes Heereman and
Hauke R. Heekeren,
Biological Psychology and Cognitive
Neuroscience, Department
of Education and Psychology, Freie
Universitaet Berlin, Habelschwerdter
Allee 45, 14195 Berlin, Germany
johannes.heereman@fu-berlin.de;
hauke.heekeren@fu-berlin.de

Received: *24 May 2015*
Accepted: *21 September 2015*
Published: *09 October 2015*

Citation:
Heereman J, Walter H
and Heekeren HR (2015)
A task-independent neural
representation of subjective certainty
in visual perception.
Front. Hum. Neurosci. 9:551.
doi: 10.3389/fnhum.2015.00551

Introduction

Am I really sure? This is a question not only scientists ask themselves but practically everybody everyday. Shall I get a less interesting but better paying job? Shall I finally end my annoying relationship? Or give it another try? These are decisions we make with more or less subjective certainty. Generally, when we make decisions, we do that with varying degrees of subjective certainty, or confidence. The mechanisms of the emergence of degrees of subjective certainty have been investigated in humans for more than a century (e.g., Peirce and Jastrow, 1884; Vickers, 1979; Fleming et al., 2010, 2012; Pleskac and Busemeyer, 2010; Hebart et al., 2014; Zizlsperger et al., 2014; Gherman and Philiastides, 2015) and more recently in animals (Kepecs et al., 2008; Kiani and Shadlen, 2009; Komura et al., 2013).

So far research has mostly focused on process models of certainty (e.g., Peirce and Jastrow, 1884; Vickers, 1979; Pleskac and Busemeyer, 2010; Hebart et al., 2014; Gherman and Philiastides, 2015) rather than representational models. While process models describe system dynamics and computational mechanisms, representation models describe the processing device itself and the way an entity is represented (such as by a feature list).

An outstanding question on the representational level is how degrees of subjective certainty are represented on the neural level. For example it may be the case that in two different tasks distinct brain areas represent subjective certainty. In line with this view, Kiani and Shadlen (2009) report that choice certainty is computed along with the decision in sensory-motor neurons in Area LIP, i.e., represented in a task-specific fashion. On the other hand, De Gardelle and Mamassian (2014) provide behavioral evidence supporting the view that one's subjective confidence in a decision is represented in a task-independent format. In this study subjects either performed two identical or two different perceptual task-trials in succession. After each pair of trials they had to indicate, in which of the two trials they were more confident in the correctness of their decision. The authors reason, that if confidence was task specific, then comparing confidence across two different tasks should be harder than comparing confidence across two instances of the same task. They found no difference between the two conditions, supporting the view that confidence is accessed as an abstract and task-independent quantity. In consequence, as an alternative hypothesis to the task-specific representation of subjective certainty, subjective certainty in different tasks could share a common neural substrate. Previous neuroimaging studies identified neural correlates of decision confidence, i.e., the degree of belief subjects have in the correctness of their choice (e.g., Fleming et al., 2012; Hebart et al., 2014). The brain area most consistently found to code for post decisional confidence in humans is the dorsomedial prefrontal cortex (DMPFC; Fleck et al., 2006; Fleming et al., 2012; Hebart et al., 2014). Other areas reported to carry a confidence signal in humans include the dorsolateral prefrontal cortex (DLPFC; Fleck et al., 2006), the ventral striatum and the right anterior insula (Hebart et al., 2014) as well as right posterior parietal cortex and bilateral middle frontal gyrus (MFG; Fleming et al., 2012). Crucially, the dependence of these neural representations on a particular task remains unclear. Previous studies couldn't address this question because they either only looked at one task at a time (e.g., Fleming et al., 2012; Hebart et al., 2014), didn't control for difficulty (Fleck et al., 2006), or were pure behavioral studies (e.g., De Gardelle and Mamassian, 2014).

The hypothesis we investigate here is, whether there is a task-independent neural representation of subjective certainty. We reasoned that in a brain region, to be considered the neural substrate of a task-independent certainty-representation, BOLD signal should vary with the degree of subjective certainty, independent of the kind of task. For this study we therefore developed a functional magnetic resonance imaging (fMRI)-design combining two tasks, a color and a motion detection task.

Because most researchers in the field are interested in the relationship between performance and confidence (belief in the correctness of a choice; for critical treatments of this approach see Gigerenzer et al., 1991; Drugowitsch et al., 2014) the usual approach is to calibrate stimulus intensities to predefined levels of performance. In the experimental trials subjects indicate their (binary) decision first and only then rate their confidence (e.g., Fleming et al., 2010; Pleskac and Busemeyer, 2010; Fleming et al., 2012). Crucially, Baranski and Petrusic (2001) report that in this design confidence processing occurs both parallel to choice and after choice. In addition, Pleskac and Busemeyer (2010) and Yu et al. (2015) show that in this design subjects continue to accumulate evidence after choice, so that the confidence ratings are not based on the same evidence as the decision. Finally, Boldt and Yeung (2015) report evidence supporting the view that shared mechanisms underlie post decisional confidence ratings and post decisional error detection. Therefore, because we were interested in the degree of certainty leading to choice (and neither in post decisional confidence nor in error detection) and it's neural correlates we flipped the usual response order and asked for certainty ratings first and let subjects only then indicate their decision. Also, because the quantity of interest here is subjective certainty, in contrast to previous studies, we calibrated stimulus intensity to a predefined average level of subjective certainty instead of performance.

Materials and Methods

Participants

A total of 24 healthy right-handed volunteers with normal or corrected to normal vision participated in the experiment. Eligibility was assessed with a general health questionnaire and an fMRI safety screening form. None had a history of psychiatric or neurological disorder. Two subjects aborted the experiment due to dizziness. Two additional subjects were excluded due to excessive head motion. The latter 2 subjects' data were used for behavioral analyses, resulting in a sample of 22 subjects (mean age = 23.5, min = 20, max = 28, 13 female). (The final fMRI analyses were carried out with data obtained from the remaining 20 subjects (mean age = 23.3, min = 20, max = 28, 11 female). The study was approved by the local ethics committee at the Freie Universitaet Berlin, Germany, and carried out in accordance to the Declaration of Helsinki. All subjects gave informed written consent before the study.

Stimuli

Participants performed a color-motion-detection-task (see **Figure 1**). The dots were blue and red, presented on a black background and were drawn in a circular aperture for the duration of one video frame (60 Hz). The dots were redrawn after ~50 ms at either a random location or a neighboring spatial location to induce apparent motion. The resulting motion effect appeared to move between 3 and 7°/s, and the dots were drawn at a density of 16.7 dots per degree/second. For the color manipulation we used the values from Kayser et al. (2010). The two values were red: RGB = (255 65 2) and blue: RGB = (5 137 255). A subportion of the dots was assigned the target color while the rest of the dots was evenly divided between blue and red. The subset of dots representing the coherent feature (motion and/or color) was changed from frame to frame so that the subset of coherent dots on one frame was not the same as the subset of coherent dots on the previous frame. The task was implemented using Psychtoolbox-3 (Brainard, 1997; Pelli, 1997; Kleiner et al., 2007). In order to avoid floor and ceiling effects, previous to testing stimulus intensities (coherence values) were calibrated to an average level (2.5 on a scale of 1–4) of subjective certainty in a subject-specific manner. The resulting average coherence-levels were 0.178 (min = 0.14, max = 0.217) for motion and 0.197 (min = 0.137, max = 0.269) for color. These subject-specific

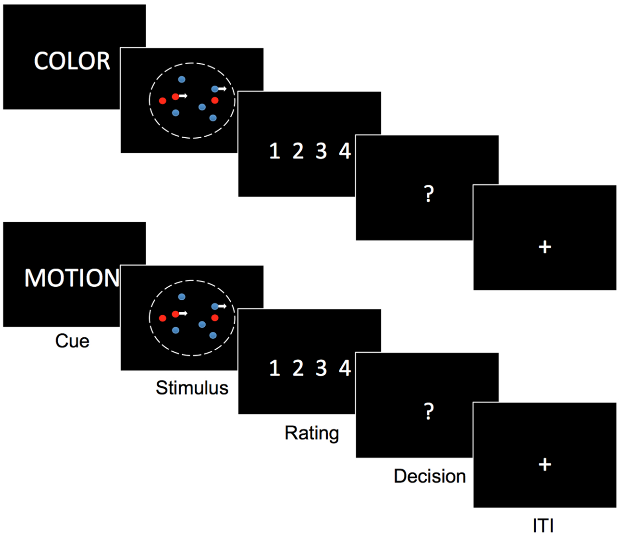


FIGURE 1 | Task description. subjects see on the display a cloud of colored moving dots. In the beginning of each trial they are cued whether to attend motion and ignore color or attend color and ignore motion. Depending on the cue after stimulus presentation (750 ms) they have to rate their degree of certainty that motion was to the left or right or that one color was more present than the other (blue or red). Only after the certainty rating, they indicate the respective direction or color. Task modified from Kayser et al. (2010).

stimulus intensities were then held constant throughout the whole experiment. Subjects indicated their responses using a four-button-response-box in the right hand (for ratings) and a two-button-response-box in the left hand (for direction and color indication). (Current Designs, Philadelphia).

Task

Subjects saw a cloud of colored moving dots. In the beginning of each trial a verbal cue appeared on the screen ('color,' 'motion') instructing subjects to attend motion and ignore color or attend color and ignore motion. Depending on the cue, after stimulus presentation (750 ms) they had to rate their degree of certainty that the net-motion of a dynamic random dot stimulus was to the left or right or that the number of dots in one color was greater than the number of dots in the other color (blue or red). After the rating subjects indicated the respective direction or color. [See **Figure 1**, Task modified from Kayser et al. (2010)]. Note that we flipped the usual order of response prompts, that is we asked for certainty-ratings first and only then let subjects indicate their binary choice. To minimize switch cost trial types were presented in Blocks of 16 trials. Block order was counterbalanced across runs and subjects. All subjects completed 5 runs of 64 trials in a pseudorandomized fashion.

Behavioral Analysis

To analyze the effect of certainty and task on RT we specified a linear mixed model. For the analysis of effects of confidence and task on performance we used a generalized linear mixed model. In both analyses we followed an information theoretic approach via AIC comparison. To arrive at the minimum adequate model we compared a (1) full model including confidence, RT and their interaction term with 3 reduced models: (2) without interaction term, (3) only confidence term, (4) only task term.

fMRI Data-acquisition

Whole-brain functional and anatomical images were acquired using a 3.0 T Magnetom TrioTim MRI scanner (Siemens, Erlangen, Germany) and a 12-channel head coil. A high-resolution 3D T1-weighted dataset was recorded for each participant (176 sagittal sections, 1 mm × 1 mm × 1 mm; 256 matrix × 256 matrix). Functional images were acquired using a T2*- weighted, gradient-echo echo planar imaging (EPI) pulse sequence recording 37 axial slices (no gap) for whole brain coverage at an in-plane resolution of 3 mm × 3 mm × 3 mm (TE = 30 ms; TR = 2 s; FA = 70°; FoV = 192 mm × 192 mm; 64 matrix × 64 matrix). A total of 290 whole-brain volumes were recorded for each of five experimental runs of ~10 min each.

fMRI-preprocessing

Data quality was checked using ArtRepair[1]. Bad slices (scanner-artifacts due to Radiofrequency-coil fluctuations) were detected when the amount of data scattered outside the head (in a slice) is at least T (here default, $T = 5$) above the average amount of data scattered outside the head in the corresponding slices of the best two of the first three volumes.

Bad slices were replaced by a linear interpolation of the corresponding slices in the before and after volume. In addition data were despiked and outlier-volumes replaced by interpolating between the nearest intact volumes. For a discussion of the applied data quality check methods, please refer to Mazaika et al. (2009). We performed all analyses using MATLAB (Mathworks, Natick, MA, USA), SPM8[2], and R[3]. fMRI data were preprocessed using standard procedures in SPM8. EPI images were realigned, coregistered to the respective participant's T1 scan, segmented, normalized to a standard T1 template based on the Montreal Neurological Institute (MNI) reference brain, resampled to 3 mm isotropic voxels, and spatially smoothed with an isotropic 8 mm full width at half maximum (FWHM) Gaussian Kernel.

fMRI-analysis

At the first level, we regressed fMRI time series onto a general linear model (GLM) containing stick functions representing the onset of the stimulus. Separate regressors aligned to stimulus onset modeled color and motion trials, each parametrically (linear) modulated by the reported certainty rating. Regressors were convolved with a double gamma hemodynamic response function (HRF). Motion correction parameters were entered as regressors of no interest and we applied a high-pass filter (128 s cutoff) to exclude low-frequency drifts. First-level contrast images were entered into a second-level ANOVA. The analysis included four first-level contrast images (positive and negative parametric effects of certainty in color and motion trials) from

---

[1] http://cibsr.stanford.edu/tools/human-brain-project/artrepair-software.html
[2] http://www.fil.ion.ucl.ac.uk/spm/
[3] https://www.r-project.org/

each participant. We then performed conjunction analyses (test of conjunction null hypothesis, i.e., logical AND) on the two positive and on the two negative parametric effects. All reported changes in BOLD signal survive $p < 0.05$, Family-wise-error (FWE) -corrected, at the cluster level for multiple comparisons using a cluster-defining threshold of $p < 0.001$, uncorrected.

To further specify the relation between rating-level and changes in BOLD signal, we specified an additional model. Here instead of using the rating level as a parametric modulator of one regressor, we modeled each rating level using a separate regressor. This allowed us to compute percent BOLD signal change for each rating level (e.g., **Figure 5**).

Results

Behavior

Subjects were correct on 80.57% (±14.99) of motion trials and 74.22% (±15.69) of color trials. Mean confidence-ratings were 2.67 (±0.25) in motion trials and 2.65 (±0.38) in color trials. Mean reaction times were 0.84 s (±0.35) in motion trials and 0.88 s (±0.36) in color trials. For the relationship between certainty, task and RT we found that a reduced model with certainty but without task as predictor and without interaction term (AIC = 7515.7) was superior to all other models. (AIC full = 7532.851, AIC task = 8495.642, AIC task + certainty = 7518.444, also see **Table 1**). For the relationship between certainty, task and performance a different reduced model (with certainty and task as predictors but without interaction term; AIC = 6512.2) was superior to all other models (AIC full = 6514.8, AIC certainty = 6551.9, AIC task = 6725.8). (See **Table 1** for details of the winning model and **Figure 1**). Given that we calibrated with respect to confidence and not to performance this latter result is to be expected. We further checked whether relationships between RT, subjective certainty and performance as expected based on the literature were present in our data. As expected, average task performance was higher for high (rating = 4 = 'certain') subjective certainty (color: mean = 0.8518; motion: mean = 0.885) than for low (rating = 1 = 'uncertain') certainty trials [color: mean = 0.5991, paired t-test $t(21) = -5.777$, $p = 0.00000982$; motion: mean = 0.6546, $t(21) = -6.0991$, $p = 0.000004725$] while average reaction times decreased with increasing certainty (color trials: mean $r = -0.3357 \pm 0.234$ SD; motion trials: mean $r = -0.3265 \pm 0.1799$ SD; **Figure 2**). Also as expected, RTs were on average longer in error-trials (color mean = 0.9272, motion mean = 0.9565) than in correct trials [color mean = 0.8723; paired t-test, $t(21) = -2.2647$, $p = 0.03424$, motion mean = 0.8283, paired t-test, $t(21) = -4.0415$, $p = 0.0006$]. Choice-RTs (time between rating and decision) were also longer in error trials (color mean = 0.5446, motion mean = 0.5309) than in correct trials [color mean = 0.4893, paired t-test, $t(21) = -3.8453$, $p = 0.00094$; motion mean = 0.4455, paired t-test, $t(21) = -3.8423$, $p = 0.00095$].

In particular, we wondered whether subject-specific calibrated stimulus intensities were related to mean performance but

TABLE 1 | Minimum adequate model for RT and performance as dependent variable.

	Log RT as dependent variable		Performance as dependent variable	
	Estimate	***SE***	**Estimate**	***SE***
Intercept	−0.12	0.05	0.39	0.2
Certainty 2	0.05	0.02	0.4	0.06
Certainty 3	−0.17	0.02	0.5	0.097
Certainty 4	−0.44	0.02	1.1	0.097
Task 2	–	–	1.47	0.12

The minimum adequate model for RT has only confidence as a predictor, while the model for performance additionally includes task (color or motion) as a predictor. Intercept: mean log RT (respectively, performance) for certainty rating 1 in task 1. Certainty 2–4: increasing certainty ratings.

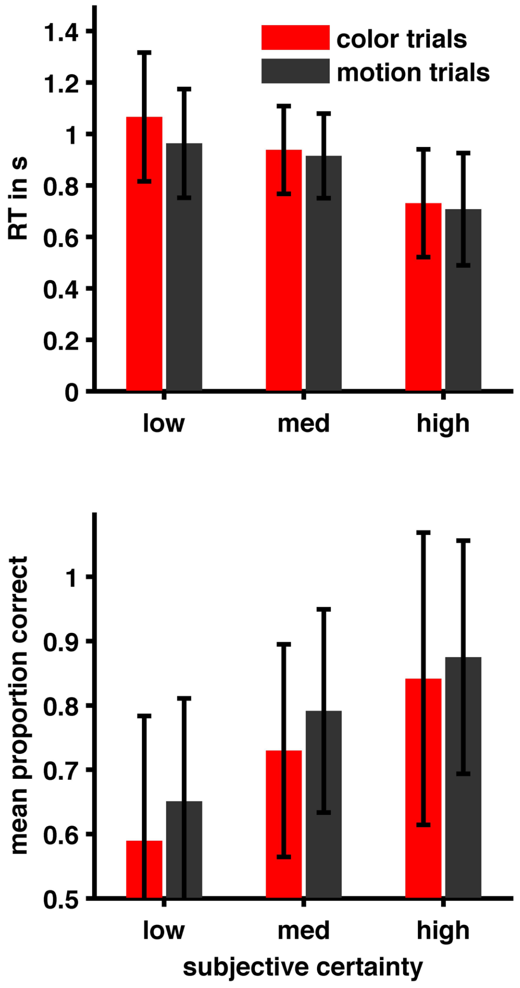


FIGURE 2 | Relation between subjective certainty ratings, performance, and reaction time. Performance increases with subjective certainty while RT decreases (see also Results section and **Table 1**).

found no significant correlation (color trials: $r = 0.2516$, $p = 0.2587$; motion trials: $r = -0.0036$, $p = 0.9857$). The same was the case for the relation between calibrated stimulus intensities and rating-RT (color trials: $r = -0.0616$, $p = 0.7854$; motion trials: $r = 0.2767$, $p = 0.2126$) as well as between intensities and subjective certainty (color trials: $r = -0.4047$, $p = 0.0617$; motion trials: $r = -0.2720$, $p = 0.2207$). We found that calibrated color and motion intensities were correlated across subjects ($r = 0.4771$, $p = 0.0248$), suggesting that we calibrated to a dimension that is shared by the color and the motion task. Also, choice-RT, i.e., the time between choice-screen onset and choice displayed a significant negative correlation with performance in color trials ($r = -0.5053$, $p = 0.0164$) but not in motion trials ($r = -0.2910$, $p = 0.1889$).

Finally, we performed a pairwise comparison of the standard deviations of the certainty-ratings in the two tasks but found no significant difference [$t(21) = 1.885, p = 0.073$].

fMRI-results

Unless indicated otherwise all changes in BOLD signal are reported at a cluster-defining threshold of $p < 0.001$ (uncorrected) and family wise error (FWE) corrected for multiple comparisons at $p < 0.05$.

Color trials: In the color task we found a positive parametric effect of subjective certainty in the right lingual, calcarine, fusiform, and left angular gyrus. We found a negative parametric effect of certainty in the supplementary motor area (SMA) within DMPFC, superior frontal gyrus (SFG), lingual gyrus, inferior frontal gyrus (IFG), Insula, and inferior parietal lobule. (See **Table 2** for a full list of activations).

Motion trials: In the motion task we found a positive parametric effect of subjective certainty in the angular, calcarine, lingual, and fusiform gyrus, middle orbital gyrus within the ventromedial prefrontal cortex (vmPFC), and posterior cingulate cortex. We found a negative parametric effect of certainty in SMA/DMPFC, SFG, lingual gyrus, fusiform gyrus, precuneus, superior parietal lobule, inferior parietal lobule, and IFG (p. triangularis). (See **Table 3**).

Conjunction null: Testing the Conjunction null (logical AND) for the positive parametric effects we found two significant clusters. The first was centered in the right calcarine gyrus extending into the right lingual gyrus and the right fusiform gyrus. The second one was located in the left angular gyrus (see **Table 4**). Testing the Conjunction

TABLE 2 | Areas showing significant correlations with subjective certainty ratings in color trials (cluster-defining threshold $p < 0.001$, uncorrected; reported changes in BOLD signal corrected for multiple comparisons at $p < 0.05$, whole brain corrected).

Region	Voxels at $p < 0.001$	Peak t-score	P (cluster FWE corrected)	Peak voxel MNI coordinates
Positive correlations				
Lingual gyr (r)	339	5.84	<0.001	21, −70, −8
Calcarine gyr (r)		5.19		18, −82, 4
Fusiform gyr (r)		4.36		30, −67, −5
Angular gyr (l)	83	4.77	<0.019	−45, −76, 28
Negative correlations				
DMPFC/SMA (l/r)	1604	7.49	<0.001	−9, 11, 49
Sup front gyr (l)		6.4		−24, −4, 52
Postcentral gyr (l)		6.29		−42, −19, 52
Lingual gyr (l)	157	6.57	<0.001	−12, −88, −5
Inf occipital gyr (l)		3.9		−33, −88, −11
		3.6		−24, −91, −14
Inf front gyr (l)	260	5.43	<0.001	−45, 14, 1
		4.27		−51, 26, 28
Insula (l)		4.08		−30, 29, 4
Inf pariet lobule (r)	165	4.71	<0.001	36, −43, 52
Supramarginal gyrus (r)		3.73		33, −40, 43
Sup pariet lobule (r)		3.65		42, −31, 31
Insula (r)	82	4.43	<0.003	45, 14, −2
		4.37		33, 17, 4
Inf front gyr (r)		3.66		54, 17, 10
Precuneus (r)	69	4.24	<0.005	12, −64, 52
Middle front gyr (r)	95	3.99	<0.001	33, 38, 28
Middle front gyr (r)		3.75		27, 47, 28
Inf front gyr (r)		3.67		51, 29, 28

FWE, family wise error; l, left; r, right; MNI, Montreal Neurological Institute; SMA, supplementary motor area; DMPFC, dorsomedial prefrontal cortex, inf, inferior; sup, superior; pariet, parietal; front, frontal; gyr, gyrus, lob, lobule.

TABLE 3 | Areas showing significant correlations with subjective certainty ratings in motion trials (cluster-defining threshold $p < 0.001$, uncorrected; reported changes in BOLD signal corrected for multiple comparisons at $p < 0.05$, whole brain corrected).

Region	Voxels at $p < 0.001$	Peak t-score	P (cluster FWE corrected)	Peak voxel MNI coordinates
Positive correlations				
Angular gyr (l)	173	6.06	<0.001	−45, −70, 25
Inf pariet gyr (l)		5.3		−48, −73, 37
Calcarine gyr (r)	370	5.79	<0.001	15, −85, 1
Lingual gyr (r)		5.00		12, −76, −8
Fusiform gyr (r)		4.98		24, −73, −11
vmPFC (l/r)	562	5.74	<0.001	−3, 44, −11
		5.00		3, 35, −11
Rectal gyr (l/r)		4.57		−3, 44, −20
Posterior cingulate (l)	267	5.5	<0.001	−9, −52, 31
		4.55		−15, −52, 13
Calcarine gyr (l)		3.83		−6, −52, 7
Negative correlations				
SMA/DMPFC (l/r)	1970	7.09	<0.0001	−6, 8, 49
		6.04		21, 5, 67
Sup Front gyr (r)		6.01		24, 2, 58
Lingual gyr (l)	116	5.32	<0.004	−12, −88, −5
		4.69		−15, −85, −14
Fusiform gyr (l)		4.33		−24, −76, −8
Precuneus (r)	167	5.26	<0.001	9, −64, 49
		4.81		15, −76, 55
Sup pariet lob		3.73		18, −64, 61
Inf pariet lob (r)	212	4.7	<0.0001	39, −43, 46
		5.56		36, −49, 52
		4.14		33, −43, 37
Inf front gyr (l)	83	4.28	<0.019	−51, 29, 25

vmPFC, ventromedial prefrontal cortex.

TABLE 4 | Conjunction: areas showing task-independent significant correlations with subjective certainty ratings in both color and motion trials.

Region	Voxels at $p < 0.001$	Peak t-score	P (cluster FWE corrected)	Peak voxel MNI coordinates
Positive correlations				
Calcarine gyr (r)	232	5.18	<0.001	18, −82, 4
Lingual gyr (r)		4.98		15, −76, −8
Fusiform gyr (r)		4.98		24, −73, −11
Angular gyr (l)	72	4.77	<0.032	−45, −76, 28
Negative correlations				
DMPFC (l/r)	1133	7.09	<0.001	−6, 8, 49
Postcentral gyr (l)		5.67		−48, −19, 55
SMA (r)		5.6		15, 11, 67
Lingual gyr (l)	78	5.32	<0.024	−12, −88, −5
Insula (r)	68	4.37	<0.04	33, 17, 4
		4.37		45, 14, −2
Inf front gyr (r)		3.66		54, 17, 10
Inf pariet lob (r)	84	4.24	<0.018	39, −43, 52
Inf pariet cortex (r)		3.48		39, −31, 37
Supramarg gyr (r)		3.41		54, −34, 46

Cluster-defining threshold $p < 0.001$, uncorrected; reported changes in BOLD signal corrected for multiple comparisons at $p < 0.05$, whole brain corrected.

null of the certainty activation maps in both trial-types for the negative parametric effects we found four significant clusters.

The first and biggest was centered in DMPFC extending from left SMA into the left postcentral gyrus and right SMA. The others were located at left lingual gyrus, right insula, extending into right

IFG (p. opercularis) and right inferior parietal lobule extending into the right supramarginal gyrus (see **Table 4**). For an overview of positive and negative parametric effects see **Figures 3** and **4**, respectively.

Control analysis: To exclude the possibility that our main conjunction-results were biased by the rating-distribution we performed a confirmatory analysis with homogenized rating-frequencies. We randomly selected trials so that for each rating-level we used the same amount of trials. We performed this analysis within a mask generated from the contrasts from the previous parametric whole brain analysis (initial threshold $p < 0.001$, with heterogeneous rating-distributions). As not all subjects used all rating levels in all runs, for this model we did the following: If a subject in at least one run didn't use rating level 1 (uncertain) or 2 (rather uncertain) we collapsed these two rating levels. If a subject in at least one run didn't use rating level 3 (rather certain) or 4 (certain) we collapsed these two rating levels. We confirmed all our main results (although the exact peak-coordinates differ slightly, compare **Tables 4** and **5**) for the conjunction of the positive parametric effects (see **Table 5**).

For the conjunction of negative parametric effects we confirmed our frontal and occipital activations, but couldn't confirm the right insular and right parietal cluster (see **Table 5**). We wondered whether this was due to the different rating distribution or due to the reduced trial-number [and therefore a signal-to-noise-ratio (SNR) issue]. We reasoned, that if changes in BOLD signal, although non-significant, followed the same

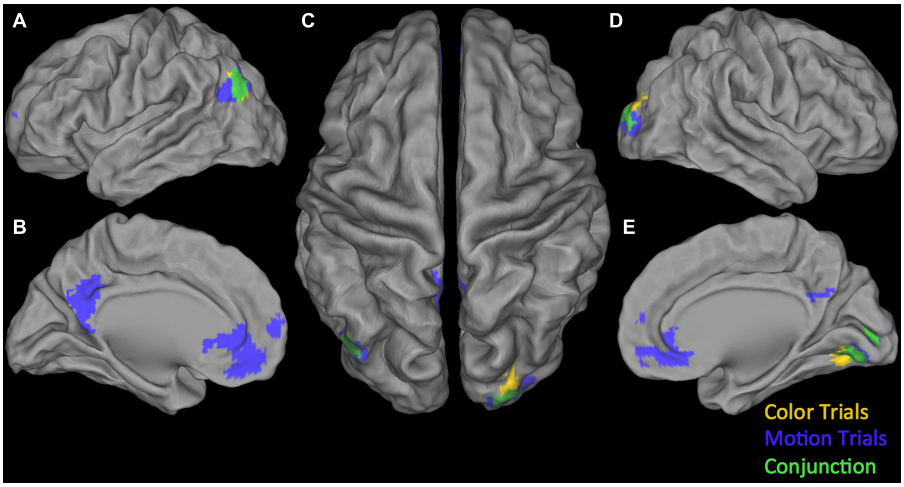


FIGURE 3 | Overview of positive parametric effects in color (yellow) and motion (blue) trials and their conjunction (green). (A) Lateral view of left Hemisphere, **(B)** Medial view of left Hemisphere, **(C)** Dorsal view, **(D)** Lateral view of right Hemisphere, and **(E)** Medial view of right Hemisphere. See **Tables 2–4** for details.

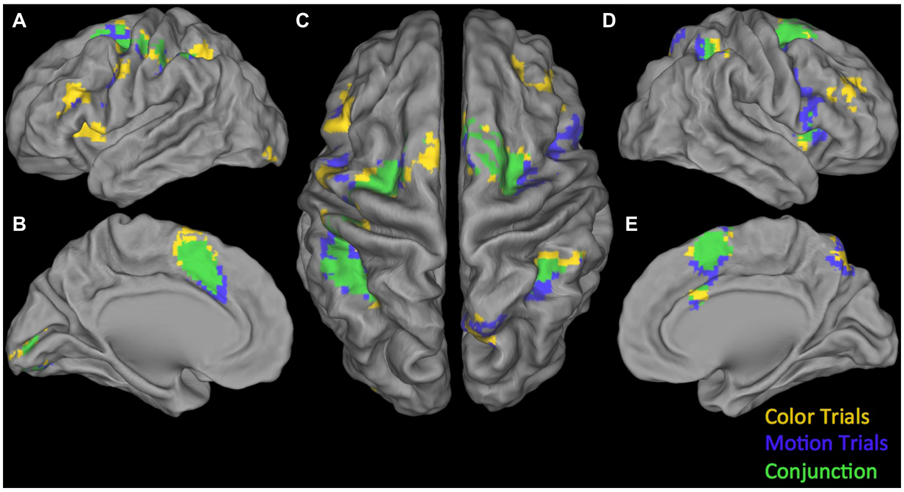


FIGURE 4 | Overview of negative parametric effects in color (yellow) and motion (blue) trials and their conjunction (green). (A) Lateral view of left Hemisphere, **(B)** Medial view of left Hemisphere, **(C)** Dorsal view, **(D)** Lateral view of right Hemisphere, and **(E)** Medial view of right Hemisphere. See **Tables 2–4** for details.

TABLE 5 | Conjunction: areas showing task-independent significant correlations with subjective certainty ratings in both color and motion trials.

Region	Voxels at $p < 0.001$	Peak t-score	P (cluster FWE corrected)	Peak voxel MNI coordinates
Positive correlations				
Calcarine gyr (r)	143	4.43	<0.001	15, −85, 1
Fusiform gyr (r)		4.39		24, −70, −8
Lingual gyr (r)		4.39		15, −79, −8
Angular gyr (l)	33	4.26	<0.003	−45, −76, 34
Negative correlations				
DMPFC/ SMA	86	5.42	<0.011	−6, 14, 49
		3.38		6, 20, 61
Calcarine gyr (l)	28	5.14	<0.017	−9, −88, −2
Lingual gyr (l)		3.73		−12, −82, −14
Postcentral gyr (l)	41	4.6	<0.007	−45, −19, 52
Sup front gyr (r)	74	4.52	<0.001	21, 5, 64
		4.51		24, 5, 52
SMA (r)		3.93		15, 11, 67
Sup front gyr (l)	26	4.52	<0.02	−21, −4, 52

*Rating frequencies are **homogenized** and the contrast is masked by the significant clusters obtained from the unhomogenized conjunction (see **Table 3**; cluster-defining threshold $p < 0.001$, uncorrected; reported changes in BOLD signal corrected for multiple comparisons at $p < 0.05$, small volume corrected).*

trend (i.e., higher betas for lower confidence) we would consider this an SNR-issue. For this purpose we specified an additional model, where we modeled trials associated with a specific rating level using separate regressors and extracted betas at the previously found peak coordinates in the insular cluster (xyz = 33, 17, 4) and the inferior parietal cluster (xyz = 39, −43, 52; see **Figure 5**). Both areas showed a negative parametric trend of certainty so we consider these differences in the results of the two models (heterogeneous vs. homogenized rating frequencies) to be due to SNR rather than the altered rating distribution.

In an additional control analysis we limited our analysis to correct trials only. Again, we confirmed all our main results (see **Table 6**). For further discussion we limit ourselves to changes in BOLD signal that were consistent across tasks and models.

Discussion

It is an outstanding question to what extent the neural computation of degrees of subjective certainty is task-specific. Here, we show that a substantial part of the neural representation of degrees of subjective perceptual certainty is task-invariant. We found that in two different tasks changes in BOLD signal increased with subjective certainty in the right lingual, calcarine, and left angular gyrus; BOLD signal decreased with increasing subjective certainty in the left lingual Gyrus, right inferior parietal lobule, bilateral DMPFC/SMA, and left postcentral gyrus. These changes in BOLD signal were virtually identical in the two tasks. Our data therefore support the view, that there is a central module in the brain processing subjective certainty and that, consistent with De Gardelle and Mamassian (2014), degrees of subjective certainty are represented in a task-independent format. This supports the notion of a generic neural mechanism underlying the computation of certainty.

Similarity to an deviations from other studies' results: The observed conjunction effects in DMPFC (−6, 8, 49, MNI coordinates of peak voxel) were located closely to confidence-related changes in BOLD signal reported by Fleck et al. (2006; −11, 15, 49), Hebart et al. (2014; −9, 15, 54), and (Fleming et al., 2012; −3, 14, 46), stressing the robustness of this finding. In our motion condition we largely replicate the confidence-related findings by Hebart et al. (2014) who also used a direction-of-motion discrimination task. Importantly, however, we could not replicate these results in the color task, although we kept everything but the instruction cue constant across the two tasks. One possible explanation for this observation is that variability in certainty-ratings may be lower in the color task than in the motion task. We performed a pairwise comparison of the standard deviations of the certainty-ratings in the two tasks but found no significant difference. This suggests that differences in the activation pattern between the two tasks are genuine. Also, we did not find the positive parametric effect of certainty in the striatum reported by Schwarze et al. (2013) or Hebart et al. (2014). Regarding the role of the ventral striatum in confidence processing there are two noteworthy recent reports: Daniel and Pollmann (2012) found a positive parametric relationship between prediction error on confidence and striatal activation. Whenever confidence was higher than could be expected from previous trials, striatal activation was also higher, indicating a role of the striatum in coding changes in confidence (rather than coding confidence itself). Schwarze et al. (2013) reported that the striatum contributes to confidence processing when the task at hand is very difficult. Using an *'unusually difficult'* task characterized by low levels of both confidence and accuracy Schwarze et al. (2013) observed a positive correlation between confidence and changes in activity in the ventral striatum. The authors explain their finding in terms of reward: Humans are typically uncertainty averse (Camerer and Weber, 1992; Hirsh et al., 2012). When difficulty is high and in consequence high-confidence trials rare, the subjective value of high confident decisions is expected to be higher than usual, i.e., subjects might experience the infrequent high confidence trials as rewarding.

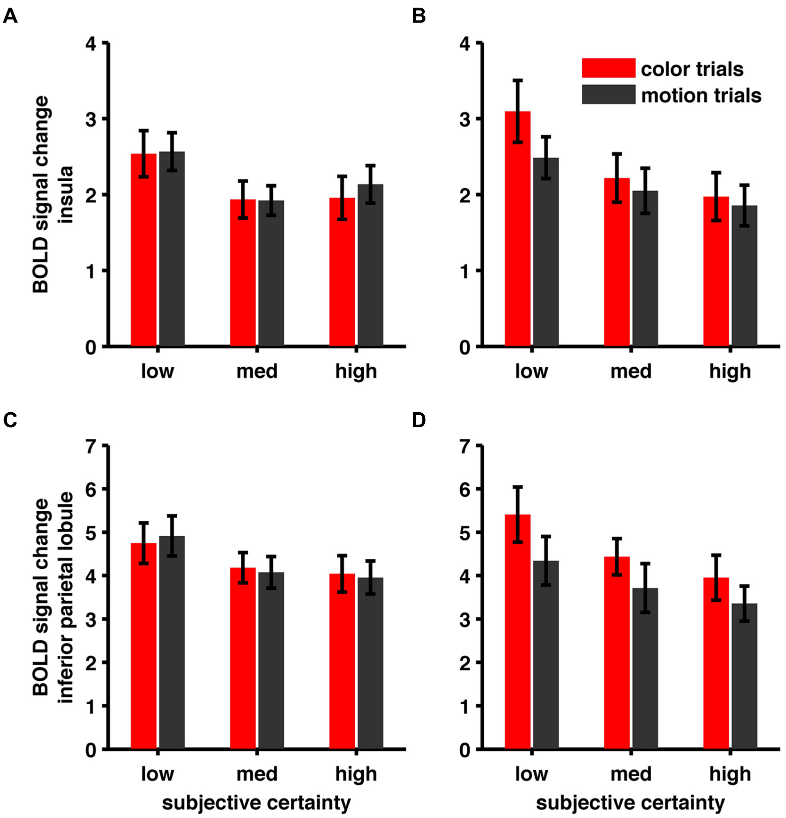


FIGURE 5 | (A) BOLD signal change in Insula (xyz = 33, 17, 4), heterogeneous rating frequencies. **(B)** BOLD signal change in Insula, homogenized rating frequencies **(C)** BOLD signal change in inferior parietal lobule (xyz = 39, −43, 52), heterogeneous rating frequencies. **(D)** BOLD signal change in inferior parietal lobule, homogenized rating frequencies.

TABLE 6 | Conjunction: Areas showing task-independent significant correlations with subjective certainty ratings in both color and motion trials.

Region	Voxels at $p < 0.001$	Peak t-score	P (cluster FWE corrected)	Peak voxel MNI coordinates
Positive correlations				
Calcarine gyr (r)	147	5.23	<0.001	21, −82, 4
Lingual gyr (r)		4.84		15, −73, −8
Angular gyr (l)	14	3.67	<0.014	−42, −70, 37
Negative correlations				
DMPFC/ SMA (l)	365	5.35	<0.001	−6, 11, 49
SMA (r)		4.89		15, 11, 67
Postcentr gyr (l)	36	4.82	<0.012	−45, −19, 52
Inf pariet lob (r)	20	4.29	<0.036	36, −46, 52
lingual gyr (l)	17	4.08	<0.046	−12, −88, −5
Inf pariet lob (l)	40	3.68	<0.01	−36, −43, 46

*Here **only correct trials** are modeled and the contrast is masked by the significant clusters obtained from the 'main'- conjunction (see **Table 3**; cluster-defining threshold $p < 0.001$, uncorrected; reported changes in BOLD signal corrected for multiple comparisons at $p < 0.05$, small volume corrected).*

Combining these two findings it appears that the striatum signals when confidence is higher than expected and that this prediction error is experienced as rewarding.

As detailed in the next section, possible explanations for differences between the results of our study and those of earlier studies may be found in the details of the respective experimental designs: Relevant factors may be (a) the quantity kept constant, (b) the precise question (performance or stimulus-related) subjects are asked, (c) the time confidence is rated, and (d) whether feedback or reward were provided.

Depending on the goal of a confidence/certainty – study mainly two strategies have been used: either keep the stimulus constant or keep performance constant (e.g., using a staircase procedure). Researchers interested in subjects' introspective

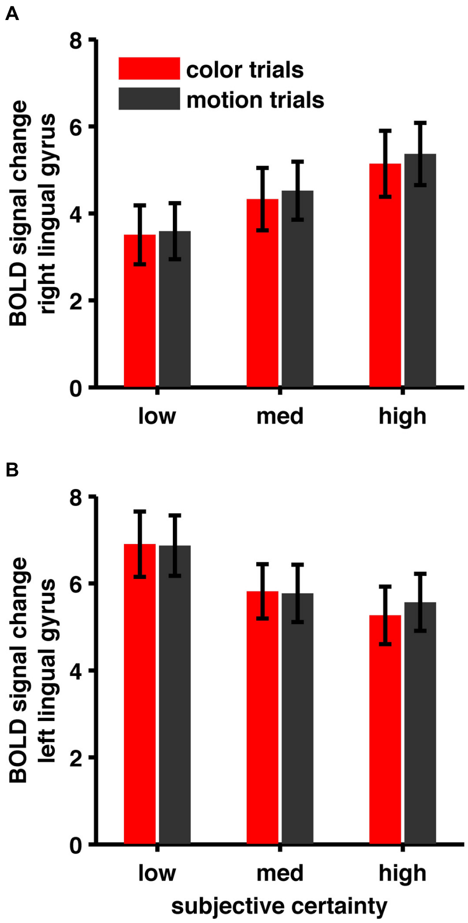


FIGURE 6 | (A) BOLD signal change in right lingual gyrus, **(B)** BOLD signal change in left lingual gyrus. Areas show opposite patterns. While BOLD signal in right lingual gyrus shows a positive correlation, BOLD signal in left lingual gyrus shows a negative correlation with subjective certainty ratings.

ability to evaluate their performance (e.g., Lau and Passingham, 2006; Fleming et al., 2012), usually use staircase-methods to keep performance constant throughout the whole experiment (because they are interested in the relation between confidence and performance). This implies that stimulus values vary. Others, like us in the present study, are interested in the relation between stimulus and confidence, so we calibrated stimulus intensities to a predefined level of confidence (and not to a level of performance), and keep this stimulus level constant throughout the experiment. This implies that performance may vary.

Researchers interested in subjects' introspective ability to evaluate their performance (e.g., Lau and Passingham, 2006; Fleming et al., 2012), usually use performance-related ratings (because they are interested in the relation between confidence and performance). That is: subjects rate how confident they are that their decision was correct. We asked subjects for their subjective certainty with regard to the stimulus. So we asked a different question. While the usual performance-related question ("How confident are you that your decision was correct?") is primarily concerned with introspection and metacognition, the stimulus-related question ("How certain are you with regard to the stimulus identity?") is of a more epistemic nature. It is concerned with the subject's estimation of the 'here and now' (Komura et al., 2013) and importantly in their belief in their percept of the world.

In most recent studies subjects are asked to indicate their decision and after a forced delay indicate their degree of certainty that their decision was correct. If the primary target of investigation is choice certainty, the degree of subjective certainty associated with (i.e., directly preceding) the actual choice, this is problematic. Pleskac and Busemeyer (2010) show, that in a perceptual task evidence accumulation continues after choice, so that post-decisional confidence ratings are not based on the same evidence underlying choice. Therefore we flipped the response-order of decision and rating and asked subjects for the rating first, that is before they indicated their decision. We reasoned that already during the decision process there should be a graded certainty signal (as recently shown by Gherman and Philiastides (2015).

In contrast to animal studies most studies of confidence in humans do not give feedback or offer reward. However, some give feedback during training (e.g., Hebart et al., 2014) which may still affect neural processes in the main experiment. Varying these four factors (quantity kept constant, performance- vs. stimulus-related question, time of rating, and feedback/reward) in a systematic and independent fashion could shed light on the origins of the observed differences in experimental results.

One possible shortcoming of this study is that variations in certainty might partially be due to fluctuations in attention. However, it has been shown (Macdonald et al., 2011) that attention and confidence, although they are both related to performance, are not necessarily correlated. In addition, although we held stimulus information constant, there are fluctuations in the momentary evidence, which may partially explain the observed variation in subjective certainty at an otherwise fixed stimulus-level [For effects of the temporal distribution of evidence on confidence see Zylberberg et al. (2012)]. However, this doesn't affect that mean-coherence, and therefore average information available to the subject, was constant.

In the present study we asked subjects for their perceptual certainty before choice and we observed a parametric modulation of BOLD signal by certainty-ratings already during stimulus presentation. Using EEG, Gherman and Philiastides (2015) showed that a confidence-signal, which could not be explained by stimulus difficulty or performance, emerges as early as the decision process itself. Similar results were obtained by Zizlsperger et al. (2014) showing that a perceptual confidence-signal, which is dissociable from representations of sensory evidence and performance, is present as early as 300 ms after stimulus onset. The temporal aspect of these results clearly challenges the generality of the metacognitive account of confidence/certainty. According to this account (e.g., Nelson, 1990; Fleming et al., 2010), confidence is modeled as the result of a noisy read-out of a decision variable. This model therefore excludes the existence of 'pre-decision-confidence' as observed by us and Gherman and Philiastides (2015). The existing literature makes it unlikely that confidence is a purely post-decisional process. Its processing might either start pre-decisionally or alternatively, pre- and post- decisional certainty processing may serve different purposes and may be based on partially different mechanisms. While pre-decision certainty is associated with incoming evidence and the actual decision process, post-decision confidence may be best viewed as an error-detection-signal (Boldt and Yeung, 2015) and, as suggested by Hebart et al. (2014), provide a learning signal in the absence of feedback. Whether this signal is metacognitive in nature and the result of a noisy read-out of the decision variable at the time of choice is unclear. Yu et al. (2015) report compelling evidence for post-decisional processing of confidence. Specifically, they show that with longer forced delays between decision and rating the resolution of the confidence-accuracy relationship increases. While confidence in correct responses stays relatively stable over different forced-delay durations, confidence in incorrect responses decreases with increasing forced delay durations. This is in line with research on the link between decision confidence and error detection (Macdonald et al., 2011; Boldt and Yeung, 2015) and at the same time challenges the metacognitive account of confidence. Further research and a formal comparison of process models are needed to shed light on this issue.

In the present study even when in addition to the stimulus, performance was held constant, we still observe parametric effects of certainty (see **Table 6**). This indicates that the processes we observe here are not directly performance-related. The observation of stimulus- and performance-independent certainty signals in our study and reported above (Zizlsperger et al., 2014; Gherman and Philiastides, 2015) is further supported by McSorley et al. (2014). They report eye-tracking-evidence for the claim that in perceptual decision making, decision and confidence are based on different information sources or processing mechanisms. In line with this dissociability, Vlassova et al. (2014) show that unconscious information changes decision accuracy but not confidence. The dissociability of performance- and confidence-related processes gets further support from the animal literature: Komura et al. (2013) report that silencing of the pulvinar nucleus decreased monkeys' confidence without affecting performance. Similarly, Lak et al. (2014) found that orbitofrontal cortex inactivation disrupts confidence processing without affecting decision accuracy. However, Fetsch et al. (2014) report opposing results suggesting that the same neural signals support choice, reaction time, and confidence in a decision. Taken together the aforementioned findings indicate that behavioral confidence and performance do not necessarily go hand in hand. On the neural level the picture is more complex: Komura et al. (2013), Fetsch et al. (2014), and Lak et al. (2014) recorded or manipulated neural activity at different sites (LIP, pulvinar thalamic nucleus, and OFC). Taken together, their results suggest that within the brain network processing confidence and performance, some nodes may represent confidence and performance jointly (e.g., LIP), while other nodes represent these quantities separately (e.g., pulvinar and OFC). This view accommodates the observations, that on the one hand confidence and performance are usually correlated, and that on the other hand there are instances in which they are not.

One particularly interesting observation is that we find both, areas showing a positive and areas showing a negative task-independent parametric effect of subjective certainty (see **Figure 6** for an example). While some authors only report negative parametric effects (e.g., Fleming et al., 2012), others report both (Hebart et al., 2014) but focus on regions displaying a positive parametric effect of confidence. The usual claim is, that either areas showing a positive effect or areas showing a negative effect represent confidence proper.

In our view the co-observation of these effects raises an intriguing alternative explanation: Imagine the brain is in a very high (or very low) activity state. A simple one-directional mechanism where certainty is computed in a way that increasing certainty is associated with increasing neural activity, would loose its calibration or better, its capability to adequately code certainty. If alternatively, certainty (as observed here) is computed in an interplay of increases and decreases in neural activity, the baseline state of brain activity is largely canceled out and the mechanism preserves its calibration. Also, a mechanism where certainty computation is realized by one principle alone is vulnerable and therefore prone to error. In contrast, a mechanism where certainty is computed in parallel (i.e., redundantly) and in several different ways would be very robust against system perturbation.

Finally, our results not only shed light on the neural representation of subjective certainty but by showing evidence for task-independence of certainty-processing also legitimize a broader and more general interpretation of previous studies of certainty. However, further research is needed to clarify the generalizability of our findings to auditory and somatosensory settings, as well as to non-perceptual tasks.

Acknowledgments

JH was supported by a PhD-scholarship from the Konrad-Adenauer-Stiftung. We want to thank the personnel of the Berlin School of Mind and Brain for support. We are indebted to Ulf Toelch, Yulia Oganian, Martin Hebart and Dar Meshi for helpful discussions and advice.

References

Baranski, J. V., and Petrusic, W. M. (2001). Testing architectures of the decision–confidence relation. *Can. J. Exp. Psychol.* 55, 195–206. doi: 10.1037/h0087366

Boldt, A., and Yeung, N. (2015). Shared neural markers of decision confidence and error detection. *J. Neurosci.* 35, 3478–3484. doi: 10.1523/JNEUROSCI.0797-14.2015

Brainard, D. H. (1997). The psychophysics toolbox. *Spat. Vis.* 10, 433–436. doi: 10.1163/156856897X00357

Camerer, C., and Weber, M. (1992). Recent developments in modeling preferences: uncertainty and ambiguity. *J. Risk Uncertain.* 5, 325–370. doi: 10.1007/BF00122575

Daniel, R., and Pollmann, S. (2012). Striatal activations signal prediction errors on confidence in the absence of external feedback. *Neuroimage* 59, 3457–3467. doi: 10.1016/j.neuroimage.2011.11.058

De Gardelle, V., and Mamassian, P. (2014). Does confidence use a common currency across two visual tasks? *Psychol. Sci.* 25, 286–288. doi: 10.1177/0956797614528956

Drugowitsch, J., Moreno-Bote, R., and Pouget, A. (2014). Relation between belief and performance in perceptual decision making. *PLoS ONE* 9:e96511. doi: 10.1371/journal.pone.0096511

Fetsch, C. R., Kiani, R., Newsome, W. T., and Shadlen, M. N. (2014). Effects of cortical microstimulation on confidence in a perceptual decision. *Neuron* 83, 797–804. doi: 10.1016/j.neuron.2014.07.011

Fleck, M. S., Daselaar, S. M., Dobbins, I. G., and Cabeza, R. (2006). Role of prefrontal and anterior cingulate regions in decision-making processes shared by memory and nonmemory tasks. *Cereb. Cortex* 16, 1623–1630. doi: 10.1093/cercor/bhj097

Fleming, S. M., Huijgen, J., and Dolan, R. J. (2012). Prefrontal contributions to metacognition in perceptual decision making. *J. Neurosci.* 32, 6117–6125. doi: 10.1523/JNEUROSCI.6489-11.2012

Fleming, S. M., Weil, R. S., Nagy, Z., Dolan, R. J., and Rees, G. (2010). Relating introspective accuracy to individual differences in brain structure. *Science* 329, 1541–1543. doi: 10.1126/science.1191883

Gherman, S., and Philiastides, M. G. (2015). Neural representations of confidence emerge from the process of decision formation during perceptual choices. *Neuroimage* 106, 134–143. doi: 10.1016/j.neuroimage.2014.11.036

Gigerenzer, G., Hoffrage, U., and Kleinbölting, H. (1991). Probabilistic mental models: a Brunswikian theory of confidence. *Psychol. Rev.* 98, 506–528. doi: 10.1037/0033-295X.98.4.506

Hebart, M. N., Schriever, Y., Donner, T. H., and Haynes, J. D. (2014). The relationship between perceptual decision variables and confidence in the human brain. *Cereb. Cortex* doi: 10.1093/cercor/bhu181 [Epub ahead of print].

Hirsh, J. B., Mar, R. A., and Peterson, J. B. (2012). Psychological entropy: a framework for understanding uncertainty-related anxiety. *Psychol. Rev.* 119, 304–320. doi: 10.1037/a0026767

Kayser, A. S., Erickson, D. T., Buchsbaum, B. R., and D'esposito, M. (2010). Neural representations of relevant and irrelevant features in perceptual decision making. *J. Neurosci.* 30, 15778–15789. doi: 10.1523/JNEUROSCI.3163-10.2010

Kepecs, A., Uchida, N., Zariwala, H. A., and Mainen, Z. F. (2008). Neural correlates, computation and behavioural impact of decision confidence. *Nature* 455, 227–231. doi: 10.1038/nature07200

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

Kleiner, M., Brainard, D., Pelli, D., Ingling, A., Murray, R., and Broussard, C. (2007). What's new in psychtoolbox-3. *Perception* 36, 1.

Komura, Y., Nikkuni, A., Hirashima, N., Uetake, T., and Miyamoto, A. (2013). Responses of pulvinar neurons reflect a subject's confidence in visual categorization. *Nat. Neurosci.* 16, 749–755. doi: 10.1038/nn.3393

Lak, A., Costa, G. M., Romberg, E., Koulakov, A. A., Mainen, Z. F., and Kepecs, A. (2014). Orbitofrontal cortex is required for optimal waiting based on decision confidence. *Neuron* 84, 190–201. doi: 10.1016/j.neuron.2014.08.039

Lau, H. C., and Passingham, R. E. (2006). Relative blindsight in normal observers and the neural correlate of visual consciousness. *Proc. Natl. Acad. Sci. U.S.A.* 103, 18763–18768. doi: 10.1073/pnas.0607716103

Macdonald, J. S. P., Mathan, S., and Yeung, N. (2011). Trial-by-trial variations in subjective attentional state are reflected in ongoing prestimulus EEG alpha oscillations. *Front. Percept. Sci.* 2:82. doi: 10.3389/fpsyg.2011.00082

Mazaika, P. K., Hoeft, F., Glover, G. H., and Reiss, A. L. (2009). Methods and software for fMRI analysis of clinical subjects. *Neuroimage* 47, S58. doi: 10.1016/S1053-8119(09)70238-1

McSorley, E., Lyne, C., and Mccloy, R. (2014). Dissociation between the impact of evidence on eye movement target choice and confidence judgements. *Exp. Brain Res.* 232, 1927–1940. doi: 10.1007/s00221-014-3884-2

Nelson, T. O. (1990). Metamemory: a theoretical framework and new findings. *Psychol. Learn. Motiv.* 26, 125–173. doi: 10.1016/S0079-7421(08)60053-5

Peirce, C. S., and Jastrow, J. (1884). On small differences of sensation. *Mem. Natl. Acad. Sci.* 3, 73–83.

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: transforming numbers into movies. *Spat. Vis.* 10, 437–442. doi: 10.1163/156856897X00366

Pleskac, T. J., and Busemeyer, J. R. (2010). Two-stage dynamic signal detection: a theory of choice, decision time, and confidence. *Psychol. Rev.* 117, 864–901. doi: 10.1037/a0019737

Schwarze, U., Bingel, U., Badre, D., and Sommer, T. (2013). Ventral striatal activity correlates with memory confidence for old-and new-responses in a difficult recognition test. *PLoS ONE* 8:e54324. doi: 10.1371/journal.pone.0054324

Vickers, D. (1979). *Decision Processes in Visual Perception*. New York, NY: Academic Press.

Vlassova, A., Donkin, C., and Pearson, J. (2014). Unconscious information changes decision accuracy but not confidence. *Proc. Natl. Acad. Sci. U.S.A.* 111, 16214–16218. doi: 10.1073/pnas.1403619111

Yu, S., Pleskac, T. J., and Zeigenfuse, M. D. (2015). Dynamics of postdecisional processing of confidence. *J. Exp. Psychol. Gen.* 144, 489. doi: 10.1037/xge0000062

Zizlsperger, L., Sauvigny, T., Händel, B., and Haarmeier, T. (2014). Cortical representations of confidence in a visual perceptual decision. *Nat. Commun.* 5, 3940. doi: 10.1038/ncomms4940

Zylberberg, A., Barttfeld, P., and Sigman, M. (2012). The construction of confidence in a perceptual decision. *Front. Integr. Neurosci.* 6:79. doi: 10.3389/fnint.2012.00079

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Study 2

Heereman, J., Heekeren, H.R., Pleskac, T.J. (*under review*). The Difference between Pre- and Post-Decision Confidence. *Psychonomic Bulletin & Review.*

http://link.springer.com/journal/13423

The Difference between Pre- and Post-Decision Confidence

Johannes Heereman and Hauke R. Heekeren

Center for Cognitive Neuroscience Berlin

Timothy J. Pleskac

Max Planck Institute for Human Development

Author Note:

Johannes Heereman, Center for Cognitive Neuroscience Berlin and Biological Psychology and Cognitive Neuroscience, Department of Education and Psychology, Freie Universitaet Berlin; Hauke R. Heekeren, Center for Cognitive Neuroscience Berlin and Biological Psychology and Cognitive Neuroscience, Department of Education and Psychology, Freie Universitaet Berlin; Timothy J. Pleskac, Max Planck Institute for Human Development, Berlin.

JH was supported by a PhD-scholarships from the Konrad-Adenauer-Stiftung and the Dahlem Research School. We want to thank Marios Philiastides for support with stimulus production and Ulf Toelch for helpful discussions.

Correspondence should be addressed to: Johannes Heereman, Habelschwerdter Allee 45, 14195 Berlin, Phone: +49/17661913775, E-mail: johannes.heereman@fu-berlin.de

Abstract:

When you ask yourself how confident you are, do you ask yourself before or after you commit to choice? Most likely you have experienced both and in fact there is good evidence for both pre- and post-decision confidence processing. It remains unclear, however, whether and how these processes differ from each other. Using a perceptual decision task, in the present study we show that post decision confidence ratings outperform pre-decision confidence ratings in predicting choice accuracy. More specifically, pre-decision confidence in incorrect trials is greater than post decision confidence in incorrect trials. At the same time confidence in correct trials remains unaffected by variation of the response order. These results are consistent with recent reports suggesting a link between post decision confidence and error-detection. In addition, we find that not only the requirement of post decision ratings, but also the requirement of post rating decisions leads to increased primary response times (as compared to choice-only and rating-only, respectively). This suggests, that other than previously thought, the increased response time is due to additional demands on motor-response planning rather than changes in the underlying computational processes.

Keywords: decision making, confidence, subjective certainty, error detection, response time

INTRODUCTION

When you make a decision, do you ask yourself how confident you are? Do you ask yourself before or after you commit to choice? The principles of confidence computation and their relation to decision formation have been subject to intense investigation and discussion within experimental psychology (Baranski & Petrusic, 1998, 2001; De Gardelle & Mamassian, 2014; Henmon, 1911; Petrusic & Baranski, 2003; Pleskac & Busemeyer, 2010; Vickers, 1979; Yu, Pleskac, & Zeigenfuse, 2015; Zylberberg, Barttfeld, & Sigman, 2012) and, more recently, cognitive neuroscience (Fetsch, Kiani, Newsome, & Shadlen, 2014;

Fleming, Weil, Nagy, Dolan, & Rees, 2010; Hebart, Schriever, Donner, & Haynes, 2014; Heereman, Walter, & Heekeren, 2015; Kepecs, Uchida, Zariwala, & Mainen, 2008; Kiani & Shadlen, 2009; Komura, Nikkuni, Hirashima, Uetake, & Miyamoto, 2013; Lau & Passingham, 2006; Meyniel, Sigman, & Mainen, 2015).

There has been substantial interest in the timing of confidence computations relative to the decision-process (Baranski & Petrusic, 1998; Moran, Teodorescu, & Usher, 2015; Petrusic & Baranski, 2003; Pleskac & Busemeyer, 2010), i.e., whether confidence computation is best viewed as a pre-decisional or post-decisional process. Notably, there is good evidence for both pre decision (Gherman & Philiastides, 2015; Kubanek, Hill, Snyder, & Schalk, 2015; Zizlsperger, Sauvigny, Händel, & Haarmeier, 2014; Zylberberg et al., 2012) and post decision confidence processing (Baranski & Petrusic, 1998; Moran et al., 2015; Pleskac & Busemeyer, 2010; Yu et al., 2015) suggesting confidence is being processed regardless if a choice is reported or not. But while pre-decision confidence may have an impact on post-rating decisions (but not vice versa), pre-rating decisions may have an impact on post-decision confidence (but not vice versa). For example, committing an error may have an impact on post- but not on pre-decision confidence. Interestingly, Yeung and Summerfield (2012) as well as Boldt and Yeung (2015) report evidence suggesting a link between post-decision confidence and error detection. More specifically, using EEG they find neural markers shared by post decision confidence and error detection (Boldt & Yeung, 2015). One hypothesis that follows from these results is that error-detection processes may contribute to post decision confidence ratings. Crucially, if error-monitoring processes are only activated when subjects commit to choice (i.e. after an observed choice) then comparing pre- and post-decision confidence ratings should separate out the contribution of error processing to confidence ratings. So here we ask, whether pre- and post-decision confidence differ in that the latter includes error detection while the former does not.

Also, Baranski and Petrusic (2001) and Petrusic and Baranski (2003) find that asking for post-decision confidence ratings is associated with increased primary response times (RT1) (as compared to choice without subsequent rating). This finding has been taken as support for the view, that asking for confidence influences the decision process. Alternatively, prolonged RT1 when a second consecutive response is required might also be explained by an increased demand for motor-response programming. It has been shown that the requirement of two or more consecutive responses (as compared to only one response) and particularly the associated button presses results in more complex advance motor-sequence planning and with that increases in primary RTs (Klapp, 1995; Sternberg, Monsell, Knoll, & Wright, 1978). Here, we therefore ask whether increased RT1 is due to altered cognitive processes or rather due to advance motor-planning of the second response.

To address these questions we adapted a behavioral face-car-categorization-task (Philiastides, Auksztulewicz, Heekeren, & Blankenburg, 2011) with four blocked conditions: subjects either 1) only indicate choice (C), 2) indicate choice then a confidence rating (CR), 3) only rate their confidence (R), and 4) rate their confidence and indicate choice afterwards (RC). Because of our use of pre-decision ratings in the present study we asked subjects: “How confident are you in your percept?” and not: “How confident are you that your decision is correct?”

Under the hypothesis that error detection processes (if they are separate from confidence processing) are engaged when people commit to choice (i.e. make an overt response) we reasoned, that if post-decision confidence includes error-detection, post-decision confidence ratings should outperform pre-decision ratings in predicting trial correctness. That is, resolution (resolution refers to the ability to distinguish correct from incorrect responses) of post-decision confidence should be superior to resolution of pre-decision confidence.

Furthermore, if in a choice-then-rating-design increased RT1 (as compared to a choice without subsequent rating) is due to advance motor-sequence planning, we should also observe increased RT1 in a rating-then-choice-design (as compared to a rating without subsequent choice).

METHOD

Participants

Because there were no prior investigations of our main hypothesis we performed a pilot study (N=7). We used G*Power (Faul, Erdfelder, Lang, & Buchner, 2007) to determine the required sample size based on the effect size in the pilot study ($\eta_p^2 = .24$, critical F = 7.64). This resulted in an actual power of .994 at a total sample size of N=30 (set $\alpha = .01$). 30 subjects (17 female) aged between 19-38 years (mean = 25.03 years) participated in the experiment. We excluded 3 participants for whom their confidence ratings suggested they did not perform the task: there was no correlation between RT1 and confidence and no relationship between confidence and their choice accuracy. This resulted in a sample of 27 subjects (16 Female, mean-age= 25.4, min=19, max=38).

All subjects performed 960 experimental trials (240 trials x 4 conditions). Trials were excluded from further analysis if subjects failed to respond or responded in the wrong order (1.875 % of the data, 18 trials per subject on average). We further excluded trials where RT1 was shorter than 200ms and/or differed more than three standard-deviations from the mean (0.792 % of data, 7.6 trials per subject on average). In total we excluded 2.67 % of data which resulted in ~ 934 remaining trials per subject. All had normal or corrected-to-normal vision and reported no history of neurological problems. Written informed consent was obtained according to procedures approved by the local ethics committee of the Freie Universitaet Berlin.

Stimuli

We used a set of 10 face (face database, Max Planck Institute for Biological Cybernetics, Tuebingen, Germany, http://faces.kyb.tuebingen.mpg.de/) (Troje & Bülthoff, 1996) and 10 car grayscale images (size 500x500 pixels, 8-bits/pixel) (used in Philiastides et al., 2011). Spatial frequency, luminance and contrast were equalized across all images. The magnitude spectrum of each image was adjusted to the average magnitude spectrum of all images used. The phase spectrum of the images was manipulated to obtain noisy stimuli of varying levels of sensory evidence (that is: we manipulated the percentage phase coherence of the images) (Dakin, Hess, Ledgeway, & Achtman, 2002).

Stimuli were presented centrally on a plain grey background on a computer screen using Psychtoolbox (Brainard, 1997; Kleiner et al., 2007; Pelli, 1997). The display was situated 0.4 m away from the subject. We used four different phase coherence values per stimulus type (face or car). At each level of stimulus sensory evidence we generated multiple frames for each image. Within each level of sensory evidence the overall amount of noise remained unchanged, while the spatial distribution of the noise varied across individual frames such that when presented (rapid serial visual presentation (RSVP), 30 frames/second, i.e. 33.33 ms per frame without gaps) different parts of the underlying image were revealed sequentially. Using QUEST (Watson & Pelli, 1983) we determined the subject- and stimulus-specific phase coherences corresponding to a performance of 60%, 70%, 80%, and 90% correct. Mean (±SD) values were 0.134 (±0.027), 0.174 (±0.032), 0.208 (±0.028), and 0.267 (±0.035) for faces. Mean values for car stimuli were 0.155 (±0.03), 0.179 (±0.028), 0.203 (±0.029), and 0.248 (±0.035).

Behavioural Paradigm

Participants performed a visual face vs. car categorization task by discriminating dynamically updating sequences of face and car images (*Figure 1, task description*). Prior to calibration (400 trials) and the first experimental session, subjects were familiarized with the stimuli and the task (160 practice trials). Altogether, all subjects completed 560 trials before testing.

All subjects performed the task in four experimental blocked conditions: subjects had to indicate 1.) their binary categorization response (face or car) (C), 2.) a binary response followed by a confidence rating in their percept (CR), 3.) a perceptual confidence rating followed by the binary response (RC), or 4.) a perceptual confidence rating only (R). Condition order was counterbalanced across subjects. Image sequences were presented in an RSVP sequence. Each trial began with a single sequence with a series of images from one of the two stimulus classes (i.e. either a face or a car) at one of the 4 possible phase coherences. Subjects indicated their choice by pressing one of two buttons on a QWERTY keyboard for the categorization response (if required) and one of four buttons for the confidence-rating (if required). For the categorization response subjects pressed the left and right arrows with their right hand. For the rating subjects pressed one of the buttons 1 (uncertain) – 4 (certain) with their left hand. We instructed participants to respond as accurately and quickly as possible. As soon as a response was made the RSVP sequence was interrupted. The RSVP sequence was allowed to remain on the screen for a maximum of 1000ms. Then subjects had maximally two more seconds to respond. In the two conditions where a second response was required (choice-then-rating and rating-then-choice) subjects were also required to respond within 2 seconds. If subjects failed to respond within this period the trial was marked as a no-choice trial and was excluded from further analysis. For each of these 4 conditions we presented 240 trials in four blocks of 60 trials each to allow subjects to rest briefly between blocks. 50% were face-trials and 50% were car-trials. All Participants performed 960 experimental trials.

Analyses

We performed all analyses using MATLAB (Mathworks, Natick, MA, USA), and R (https://www.r-project.org/).

RESULTS

In the following we use response time 1 (RT1) for the time from stimulus onset to the first button press (regardless whether this is a choice- or rating-response) (*see Figure 1*). While for statistics we log-transformed response times, for illustrative purposes (e.g. means and figures) we did not. Significance levels for analyses of variance are based on Greenhouse-Geisser adjusted degrees of freedom although the degrees of freedom reported are based on the design. We found no interactions of the factor stimulus (face or car trial) with any of our independent variables so we collapsed face and car trials.

Before we turned to the test of our hypotheses we confirmed that established relationships between stimulus-difficulty (here 1- phase-coherence), RT1, choice accuracy, and confidence were present in our data (*see Supplement*).

A repeated measures ANOVA with accuracy and condition as within-subject factors showed the usual effect of accuracy, $F(1, 26) = 324.4$, $p < .001$, $\eta_p^2 = .93$, but not condition, $F(1, 26) = 3.6$, $p = .07$, $\eta_p^2 = .12$, on confidence. Crucially, the interaction, $F(1, 26) = 9.02$, $p < .01$, $\eta_p^2 = .26$, was significant (*see Figure 2*). In both conditions confidence was higher for correct than for incorrect trials (***choice-then-rating:*** $M_{corr} = 3.12$, $SD_{corr} = .42$, $M_{incorr} = 2.09$, $SD_{incorr} = .53$, $t(26) = 18.66$, $p < .001$, $d = 7.32$, $r = .96$; *rating-then-choice:* $M_{corr} = 3.13$, $SD_{corr} = .36$, $M_{incorr} = 2.28$, $SD_{incorr} = .55$, $t(26) = 13.1$, $p < .001$, $d = 5.14$, $r = .93$). Importantly, the mean confidence-difference between correct and incorrect trials in the choice-then-rating-condition (M_{diff} = 1.03, SD = .29) was significantly higher than in the rating-then-choice-condition (M_{diff} = .85, SD = .34, t(26) = 3.0, p < .01, d = 1.178, r = .51). Further post-hoc analyses (*Bonferroni-corrected alpha = .05/3 = .0167*) show that this effect is driven by error trials: Confidence in incorrect trials was higher in the rating-then-choice- (M = 2.28, SD = .55) than in the choice-then-rating-condition (M = 2.08, SD = .52, t(26) = 2.65, p< .017, d = 1.04, r =.46). Crucially, for correct trials we found no confidence-difference between the rating-then-choice (M = 3.13, SD = .36) and the choice-then-rating condition (M = 3.12, SD = .42, t(26) = 0.25, p = .8, d = .097, r = .049).

A repeated measures ANOVA with difficulty and condition as within-subject factors showed the usual effect of difficulty, F(3, 78) = 101.9, p < .001, η_p^2 = .91, on choice accuracy (i.e. proportion correct). Neither condition, F(1, 26) = 0.2, p = .67, η_p^2 = .007, nor the interaction, F(3, 78) = .4, p = .7, η_p^2 = .05, showed a significant effect on choice accuracy (*see Figure 3*).

As expected, mean RT1 was longer in the choice-then-rating-condition (M = 1216ms, SD = 303) than in the choice-only-condition (M = 981ms, SD = 185, t(26) = 6.33, p < .001, d = 2.48, r = .78). Crucially, as hypothesized, mean RT1 in the rating-then-choice condition (M = 1304ms, SD = 324) was longer than in the rating-only-condition (M = 1137ms, SD = 283, t(26) = 4.38, p < .001, d = 1.72, r = .65) (*see Figure 4*).

DISCUSSION

In this experiment, we observe that post-decision-confidence-ratings outperform pre-decision-confidence-ratings in dissociating correct from incorrect trials (i.e. the former have superior resolution) (*see Figure 2*). This suggests the involvement of a metacognitive error-detection component in post-decision confidence-ratings (but see discussion below). Also, we replicate the finding of longer decision-reaction-times if a subsequent confidence rating is required (as compared to a choice-only-condition) (Petrusic & Baranski, 2003). Their explanation for this result was that participants were engaging in confidence processing during the choice. However, we also show that asking for a choice response after a confidence rating also leads to longer rating-reaction-times (RT1) (as compared to a rating-only-condition) (*see Figure 4*). In our view our novel finding asks for a reinterpretation of the original observation by Petrusic and Baranski (2003) in terms of motor-response planning (see discussion below).

Resolution of pre- vs. post-decision confidence

As predicted, we find post-decision-confidence-ratings to be more predictive of trial-correctness than pre-decision-ratings. That is, the resolution of post-decision confidence is superior to resolution of pre-decision confidence. Post-hoc analyses reveal that this effect is driven by incorrect trials, i.e. confidence in incorrect trials was higher in the rating-then-choice-condition than in the choice-then-rating-condition while in correct trials we observed no difference in confidence between the two conditions.

One interpretation of this pattern is, that differences between pre- and post-decision confidence are specific for error trials and due to post decisional error detection. Alternatively, lower confidence in errors in post- as compared to pre-decision confidence ratings could be due to continued evidence accumulation after choice combined with state-dependent decay. In line with the former view, using EEG Boldt and Yeung (2015) report neural markers shared by post decision confidence and error detection (i.e. retrospective detection of incorrect choices) establishing a link between the two. However, our finding is very similar to Yu et al. (2015) who found that increasing the time between decision and post decision confidence rating leads to decreased confidence in incorrect trials but does not affect confidence in correct trials. That is, both an increased delay between decision and confidence and varying the choice-rating order (this study) show a similar attenuation of decision confidence in erroneous trials. Yu et al. (2015) explain their finding with continued evidence accumulation after choice combined with evidence decay or leakage. Put simply, while in correct trials the decay of evidence supporting the correct decision is compensated by new incoming evidence also supporting the correct decision, in incorrect trials the evidence supporting the incorrect decision also decays but is replaced with evidence supporting the correct decision. Thereby confidence in correct trials remains unchanged while confidence in incorrect trials decreases with time, recreating our results.

To better dissociate the two alternative explanations we reasoned, that if our observation was due to the mechanism suggested by Yu et al. (2015), decision accuracy should be higher in the post-rating choices than in the pre-rating choices. That is, if evidence decays and is replaced by new incoming evidence supporting the correct decision, performance should on average increase with time. We found no difference in accuracy between the two conditions (choice-then-rating vs. rating-then-choice) (see Figure 3). This lack of an effect on accuracy suggests, that the difference between pre- and post- choice confidence is due to a contribution of error-monitoring to confidence ratings when they follow a choice. In our view, this implies a conceptual difference between pre- and post-decision confidence ratings: while in the case of ratings following choice we would speak of a metacognitive process (or at least a process involving a metacognitive component) in the rating-then-choice case we would not.

Response time as a function of response sequence length

Increased choice-RT conditional on the requirement of a consecutive rating has been taken as evidence for the view that confidence computation requires additional processing. Crucially, Lebreton, Abitbol, Daunizeau, and Pessiglione (2015) report neural evidence showing that confidence is processed regardless whether asked for or not. In our view this renders the interpretation of prolonged choice-RT1 due to confidence processing rather unlikely. In light of our novel finding (longer rating-RTs if choice is required afterwards) we rather suggest that prolonged RT1 is due to the requirement of a second motor-response rather than confidence (or choice) processing per se. This is in line with Sternberg et al. (1978) and Klapp (1995) who showed, that RT1 increases with the number of required consecutive responses. Our interpretation makes a clear prediction: The requirement of any second button press should lead to a similar effect of prolonged RT1.

Conclusion

Our results challenge the view that the requirement of post-decision confidence ratings influences the decision process. Rather our data support the view that the requirement of a second response (post-choice rating or post-rating choice) increases motor-planning-demands and thereby RT1.

Most importantly here we show that there is in fact a difference between pre- and post-decision confidence ratings: post-decision confidence outperforms pre-decision confidence in dissociating correct from incorrect trials which is most likely due to post-decision error-detection. In the future, combining neural measures such as EEG and fMRI with our approach of contrasting pre- and post-decision ratings may help disentangling processes shared by confidence and choice computation from processes that uniquely contribute to confidence.

REFERENCES:

Baranski, J. V., & Petrusic, W. M. (1998). Probing the locus of confidence judgments: experiments on the time to determine confidence. *Journal of Experimental Psychology: Human Perception and Performance, 24*(3), 929.

Baranski, J. V., & Petrusic, W. M. (2001). Testing architectures of the decision–confidence relation. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale, 55*(3), 195.

Boldt, A., & Yeung, N. (2015). Shared Neural Markers of Decision Confidence and Error Detection. *The Journal of Neuroscience, 35*(8), 3478-3484. doi:10.1523/JNEUROSCI.0797-14.2015

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial vision, 10*, 433-436.

Dakin, S., Hess, R., Ledgeway, T., & Achtman, R. (2002). What causes non-monotonic tuning of fMRI response to noisy images? *Current Biology, 12*(14), R476-R477.

De Gardelle, V., & Mamassian, P. (2014). Does confidence use a common currency across two visual tasks? *Psychological science, 25*(6), 1286-1288. doi:10.1177/0956797614528956

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior research methods, 39*(2), 175-191.

Fetsch, C. R., Kiani, R., Newsome, W. T., & Shadlen, M. N. (2014). Effects of cortical microstimulation on confidence in a perceptual decision. *Neuron, 83*(4), 797-804.

Fleming, S. M., Weil, R. S., Nagy, Z., Dolan, R. J., & Rees, G. (2010). Relating introspective accuracy to individual differences in brain structure. *science, 329*(5998), 1541-1543.

Gherman, S., & Philiastides, M. G. (2015). Neural representations of confidence emerge from the process of decision formation during perceptual choices. *NeuroImage, 106*, 134-143.

Hebart, M. N., Schriever, Y., Donner, T. H., & Haynes, J.-D. (2014). The relationship between perceptual decision variables and confidence in the human brain. *Cerebral Cortex*. doi:10.1093/cercor/bhu181

Heereman, J., Walter, H., & Heekeren, H. R. (2015). A task-independent neural representation of subjective certainty in visual perception. *Frontiers in human neuroscience, 9*(551). doi:10.3389/fnhum.2015.00551

Henmon, V. A. C. (1911). The relation of the time of a judgment to its accuracy. *Psychological review, 18*(3), 186.

Kepecs, A., Uchida, N., Zariwala, H. A., & Mainen, Z. F. (2008). Neural correlates, computation and behavioural impact of decision confidence. *Nature, 455*(7210), 227-231.

Kiani, R., & Shadlen, M. N. (2009). Representation of confidence associated with a decision by neurons in the parietal cortex. *science, 324*(5928), 759-764.

Klapp, S. T. (1995). Motor Response Programming during Simple and Choice-Reaction Time - the Role of Practice. *Journal of Experimental Psychology-Human Perception and Performance, 21*(5), 1015-1027. doi:Doi 10.1037//0096-1523.21.5.1015

Kleiner, M., Brainard, D., Pelli, D., Ingling, A., Murray, R., & Broussard, C. (2007). What's new in Psychtoolbox-3. *Perception, 36*(14), 14-14. Retrieved from <Go to ISI>://WOS:000250594600049

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, 16*(6), 749-755. doi:10.1038/nn.3393

Kubanek, J., Hill, N. J., Snyder, L. H., & Schalk, G. (2015). Cortical alpha activity predicts the confidence in an impending action. *Frontiers in neuroscience, 9*, 243. doi:10.3389/fnins.2015.00243

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.

Lebreton, M., Abitbol, R., Daunizeau, J., & Pessiglione, M. (2015). Automatic integration of confidence in the brain valuation signal. *Nat Neurosci, 18*(8), 1159-1167. doi:10.1038/nn.4064

Meyniel, F., Sigman, M., & Mainen, Z. F. (2015). Confidence as Bayesian Probability: From Neural Origins to Behavior. *Neuron, 88*(1), 78-92. doi:10.1016/j.neuron.2015.09.039

Moran, R., Teodorescu, A. R., & Usher, M. (2015). Post choice information integration as a causal determinant of confidence: Novel data and a computational account. *Cognitive psychology, 78*, 99-147. doi:10.1016/j.cogpsych.2015.01.002

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial vision, 10*(4), 437-442. doi:Doi 10.1163/156856897x00366

Petrusic, W. M., & Baranski, J. V. (2003). Judging confidence influences decision processing in comparative judgments. *Psychonomic bulletin & review, 10*(1), 177-183.

Philiastides, M. G., Auksztulewicz, R., Heekeren, H. R., & Blankenburg, F. (2011). Causal role of dorsolateral prefrontal cortex in human perceptual decision making. *Curr Biol, 21*(11), 980-983. doi:10.1016/j.cub.2011.04.034

Pleskac, T. J., & Busemeyer, J. R. (2010). Two-stage dynamic signal detection: a theory of choice, decision time, and confidence. *Psychol Rev, 117*(3), 864-901. doi:10.1037/a0019737

Sternberg, S., Monsell, S., Knoll, R. L., & Wright, C. E. (1978). The latency and duration of rapid movement sequences: Comparisons of speech and typewriting. *Information processing in motor control and learning*, 117-152.

Troje, N. F., & Bülthoff, H. H. (1996). Face recognition under varying poses: The role of texture and shape. *Vision research, 36*(12), 1761-1771.

Vickers, D. (1979). *Decision processes in visual perception*: Academic Press.

Watson, A. B., & Pelli, D. G. (1983). QUEST: A Bayesian adaptive psychometric method. *Perception & psychophysics, 33*(2), 113-120.

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

Yu, S., Pleskac, T. J., & Zeigenfuse, M. D. (2015). Dynamics of postdecisional processing of confidence. *J Exp Psychol Gen, 144*(2), 489-510. doi:10.1037/xge0000062

Zizlsperger, L., Sauvigny, T., Händel, B., & Haarmeier, T. (2014). Cortical representations of confidence in a visual perceptual decision. *Nature communications, 5*, 3940. doi:10.1038/ncomms4940

Zylberberg, A., Barttfeld, P., & Sigman, M. (2012). The construction of confidence in a perceptual decision. *Frontiers in integrative neuroscience, 6*(79), 79. doi:10.3389/fnint.2012.00079

FIGURES

Figure 1: Task description: Subjects saw a sequence of dynamically updating stimuli (30 frames/second, 1 second max.) on the display. In the first condition (choice-then-rating) they indicated whether they saw a face or a car and then rated on a scale from one (uncertain) to four (certain) their confidence in their percept. In the second condition subjects only indicated their choice. In the third condition (rating-then-choice) subjects rated their confidence first and only then indicated their choice. In the fourth condition subjects only rated their confidence in their percept without indicating their choice. Subjects could give their first response already during stimulus presentation (rating or choice depending on condition). In the two two-response-conditions this was followed by a prompt for the second response (choice or rating depending on condition). RT1 = time from stimulus-onset to first response.

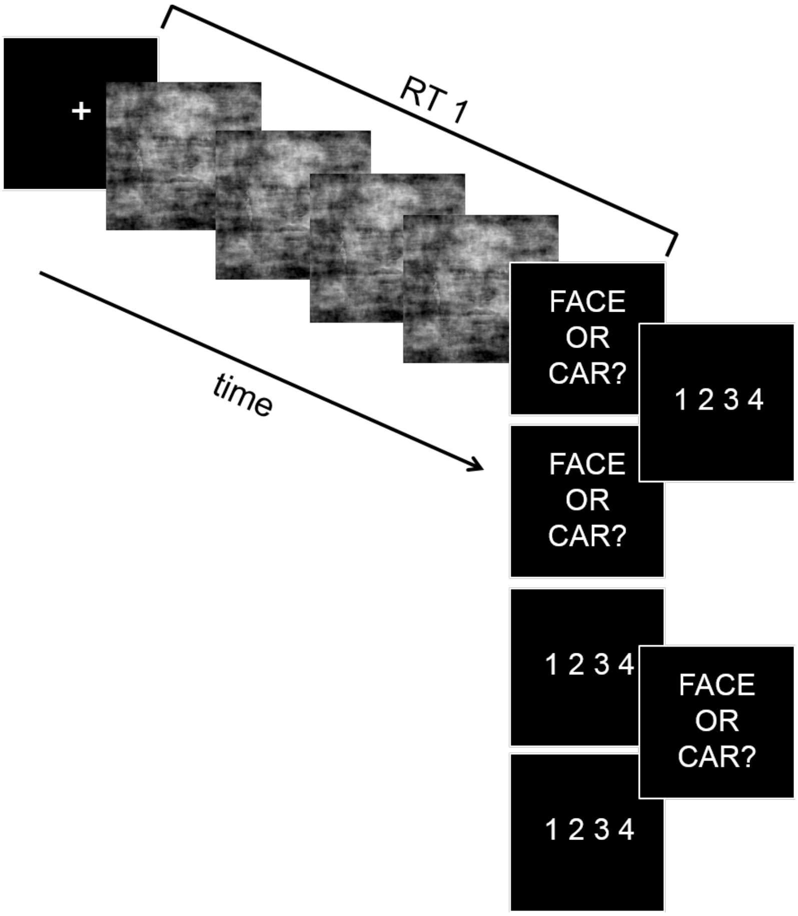


Figure 2: Confidence in correct vs. incorrect trials in choice-then-rating (CR) vs. rating-then-choice (RC) condition. *Left:* confidence in incorrect trials is higher in the RC than in the CR condition. *Right:* Confidence-difference between incorrect and correct trials is higher in CR than in RC-condition indicating the involvement of an error-detection mechanism in post-decision-ratings. (error bars represent SDs)

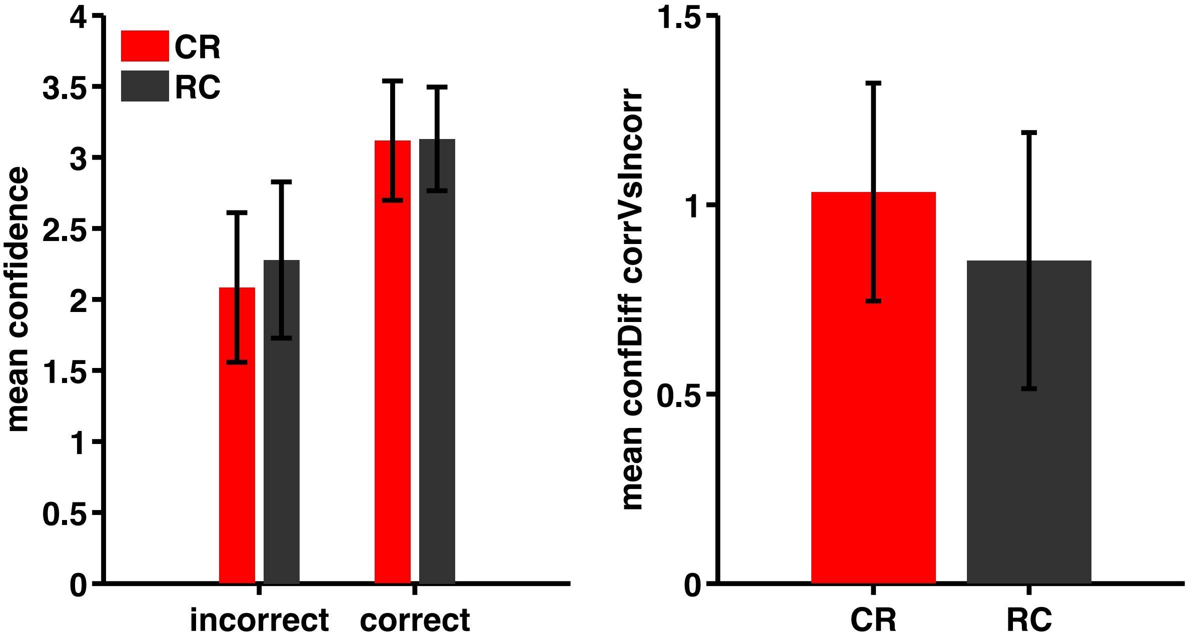


Figure 3: Proportion correct in the choice-then-rating and the rating-then-choice condition. The nearly identical proportion of correct trials supports the view that higher resolution of post-decision ratings is due to error-detection rather than continued evidence accumulation. (error bars represent SDs)

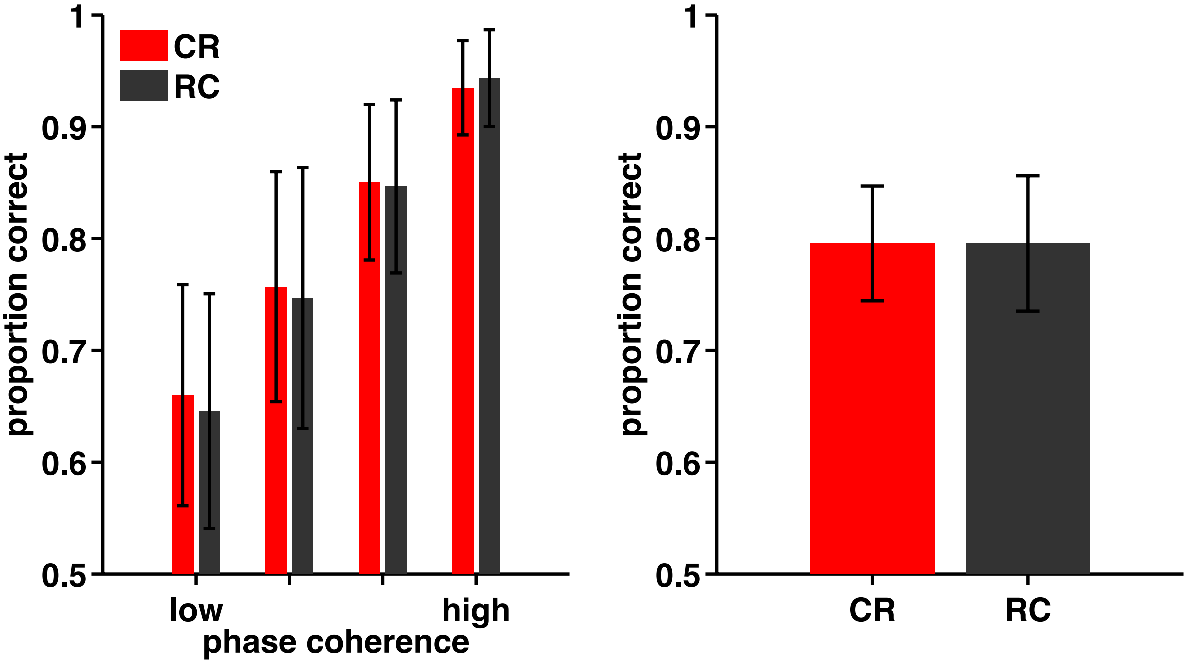
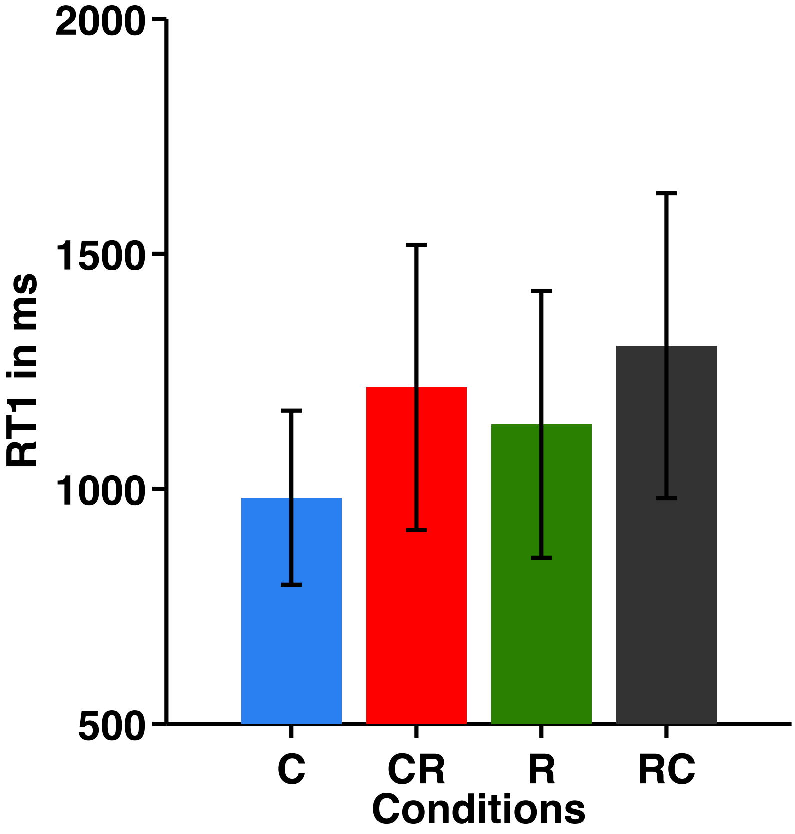


Figure 4: RT1 per condition and effect of asking for 2nd response on RT1 (C vs. CR, R vs. RC); The requirement of a second consecutive response leads to prolonged RT1 in both two-response conditions. (error bars represent SDs)



Supplemental Information: The Difference between pre- and post-decision Confidence.

Johannes Heereman, Hauke R. Heekeren, and Timothy J. Pleskac

Before we turned to the test of our hypotheses, we checked whether established relationships between stimulus-difficulty (here 1- phase-coherence), RT1 (duration between stimulus onset and first response), choice accuracy, and confidence were present in our data.

RT1 decreases with difficulty: A repeated measures ANOVA with difficulty and condition as within-subject factor revealed an effect of both difficulty, F(3,78) = 73.48, $p < .001$, $\eta_p^2 = .82$, and condition, F(3,78) = 31.2, $p < .001$, $\eta_p^2 = .74$, on RT1. The interaction, F(9, 234) = 2.77, $p < .05$, $\eta_p^2 = .43$, was also significant.

Choice accuracy increases with decreasing difficulty: A repeated measures ANOVA with difficulty and condition as within-subject factor revealed an effect of difficulty, F(3, 78) = 133.2, $p < .001$, $\eta_p^2 = .94$, but not condition, F(2, 52) = 1.8, p = .18, $\eta_p^2 = .17$) on performance. The interaction, F(6, 156) = .9, p = .47, $\eta_p^2 = .17$, was not significant.

Confidence increases with decreasing difficulty: A repeated measures ANOVA with difficulty and condition as within-subject factor revealed an effect of difficulty, F(3, 78) = 146.7, $p < .001$, $\eta_p^2 = .92$, but not condition, F(2, 52) = 1.1, p = .3, $\eta_p^2 = .16$, on confidence. The interaction, F(6, 156) = 2.7, $p < .05$, $\eta_p^2 = .54$, was also significant.

RT1 decreases with increasing confidence: A repeated measures ANOVA with confidence and condition as within-subject factor revealed an effect of confidence, F(3, 72) = 72.2, $p < .001$, $\eta_p^2 = .85$, condition, F(2, 48) = 12.2, $p < .001$, $\eta_p^2 = .6$, and their interaction, F(6, 144) = 7.1, $p < .001$, $\eta_p^2 = .7$, on RT1. As expected, we observe the typical negative correlation between decision-confidence and RT1 (here r = -0.38) (e.g. Henmon 1911; Baranski & Petrusic 1998; Pleskac & Busemeyer 2010).

Choice accuracy increases with confidence: A repeated measures ANOVA with confidence and condition as within-subject factor revealed an effect of confidence, F(3, 78) = 59.8, $p < .001$, $\eta_p^2 = .9$, but not condition, F(1, 26) = 0.001, p = .99, $\eta_p^2 = 0.00$, or their interaction, F(3, 78) = .67, p = .54, $\eta_p^2 = .1$, on performance.

Study 3

Heereman, J., Heekeren, H.R., Pleskac, T.J. (*submitted*). Evidence for Decoupled Processing of Confidence and Choice. *Journal of Experimental Psychology: Human Perception and Performance.*

http://www.apa.org/pubs/journals/xhp/index.aspx
Version: 13.02.2015

Evidence for Decoupled Processing of Confidence and Choice

Johannes Heereman and Hauke R. Heekeren

Center for Cognitive Neuroscience Berlin

Timothy J. Pleskac

Max Planck Institute for Human Development

Author Note:

Johannes Heereman, Center for Cognitive Neuroscience Berlin and Biological Psychology and Cognitive Neuroscience, Department of Education and Psychology, Freie Universitaet Berlin; Hauke R. Heekeren, Center for Cognitive Neuroscience Berlin and Biological Psychology and Cognitive Neuroscience, Department of Education and Psychology, Freie Universitaet Berlin; Timothy J. Pleskac, Max Planck Institute for Human Development, Berlin.

JH was supported by PhD-scholarships from the Konrad-Adenauer-Stiftung and the Dahlem Research School.

Correspondence should be addressed to: Johannes Heereman, Habelschwerdter Allee 45, 14195 Berlin, Phone: +49/17661913775, E-mail: johannes.heereman@fu-berlin.de

Abstract:

After you commit to a choice, do you sometimes question it? There is good evidence for post decision processing of confidence, so your answer is probably yes. It remains unclear, however, whether the other way around, decision processing continues after rating confidence. Using a perceptual decision task, we show that not only post decision rating response times, but also post rating decision response times decrease with difficulty, indicating second-stage processing of both. In addition, our results indicate that rating confidence is not associated with implicit choice. Furthermore, we observe that rating confidence and making a decision differ in their computational demands, as indicated by longer response times in ratings. Finally, we find that evidence has a differential effect on confidence and choice processing.

Keywords: decision making, confidence, subjective certainty, response time

INTRODUCTION

When you make a decision, do you estimate your confidence first and decide only then? Or the other way around? And is estimating confidence associated with an implicit decision? The principles of confidence computation and their relation to decision formation have been of longstanding interest in experimental psychology (Baranski & Petrusic, 1998, 2001; De Gardelle & Mamassian, 2014; Henmon, 1911; Petrusic & Baranski, 2003; Pleskac & Busemeyer, 2010; Vickers, 1979; Yu, Pleskac, & Zeigenfuse, 2015; Zylberberg, Barttfeld, & Sigman, 2012) and, more recently, cognitive neuroscience (Fetsch, Kiani, Newsome, & Shadlen, 2014; Fleming, Weil, Nagy, Dolan, & Rees, 2010; Hebart, Schriever, Donner, & Haynes, 2014; Heereman, Walter, & Heekeren, 2015; Kepecs, Uchida, Zariwala, & Mainen, 2008; Kiani & Shadlen, 2009; Komura, Nikkuni, Hirashima, Uetake, & Miyamoto, 2013; Lau & Passingham, 2006; Meyniel, Sigman, & Mainen, 2015).

Baranski and Petrusic (1998) observed, that when confidence ratings follow choice, the time between the first and second response (RT2) varies as a function of difficulty. This finding has been interpreted as post decision processing (i.e. second-stage processing) of confidence and in fact several recent studies have provided evidence supporting this view (Baranski & Petrusic, 1998; Moran, Teodorescu, & Usher, 2015; Murphy, Robertson, Harty, & O'Connell, 2015; Pleskac & Busemeyer, 2010; Yu et al., 2015). These results show, that confidence and choice are not necessarily computed along with each other. However, this decoupling of choice and confidence processing has only been shown for situations where confidence follows choice. Choice processing when choice follows a confidence rating (i.e. second stage processing of choice) has not been investigated so far. For example, instead of processing being completed at the time of rating, it may also continue. That is, final commitment to choice may be withheld until after confidence-estimation. Such a finding together with previous reports about post-decision confidence processing would imply, that choice and confidence processing do not necessarily depend on or require each other. So here we ask, whether decisions can be processed ‘post confidence.’

In a similar vein we wondered, whether confidence is associated with implicit choice. Intuitively, one would think, that rating confidence in a percept (e.g. “how confident are you that this is a dog or a cat?”) requires implicit choice. That is, you are confident that ‘this is a dog’ or you are confident that ‘this is a cat’. You are not just confident. In other words: in this kind of setting confidence is expected to be directed. Here, we test that intuition and ask, whether rating confidence is associated with implicit choice.

A third question is, whether confidence and choice are just two instantiations of basically the same decision mechanism, or not. For example, Vickers (1979) models choice as a race-to-threshold process and confidence as the balance-of-evidence (i.e. difference) between two accumulators at the time of choice. In contrast, Pleskac and Busemeyer (2010) and Moran et al. (2015) model confidence as a second decision. Interestingly, although there have been many studies investigating the relation of confidence and choice (e.g. Fleming, Huijgen, & Dolan, 2012; Fleming, Weil, Nagy, Dolan, & Rees, 2010; Gigerenzer, Hoffrage, & Kleinbölting, 1991; Henmon, 1911; Peirce & Jastrow, 1884) the respective underlying computational processes and their timing have not been directly compared so far. That is, it is an open question whether there is a difference in the processes underlying confidence and choice. Specifically, here we ask whether confidence and choice processing differ in their computational demands.

To address these questions we adapted a behavioral face-car-categorization-task (Philiastides, Auksztulewicz, Heekeren, & Blankenburg, 2011) with five blocked conditions: subjects either 1) only indicate choice, 2) indicate choice then a confidence rating, 3) only rate their confidence, 4) rate their confidence and indicate choice afterwards, and 5) indicate choice and confidence in one button press. Because of our use of pre-decision ratings in the present study we asked subjects: “How confident are you in your percept?” and not: “How confident are you that your decision is correct?”

We reasoned, that if choice can be processed after confidence (i.e. in a second stage), RT2 should vary as a function of difficulty not only in a choice-then-rating but also in a rating-then-choice design.

Furthermore, if rating confidence is associated with implicit choice, RT1 (the duration between stimulus onset and the first response) should be constant across conditions where (controlling for the number of response alternatives) subjects indicate choice and confidence simultaneously (i.e. combined in one button press) and conditions where they only rate their confidence. Here we therefore test, whether RT1 is longer in a combined choice/rating condition as compared to a rating-only condition.

Finally, if there is a difference in computational demands between confidence and choice processing, mean confidence response time and mean choice response time should differ.

METHOD

Participants

In the context of another study (Heereman, Heekeren, and Pleskac, *under review*) we obtained data from a sample of 27 subjects. In the present work we perform analyses including those data and matched the sample size to optimize comparability. So here, a new sample of 27 subjects (14 female) aged between 19-30 years (mean = 22.47 years) participated in the experiment. All had normal or corrected-to-normal vision and reported no history of neurological problems. Written informed consent was obtained according to procedures approved by the local ethics committee of Freie Universitaet Berlin.

Stimuli

We used a set of 10 face (face database, Max Planck Institute for Biological Cybernetics, Tuebingen, Germany, http://faces.kyb.tuebingen.mpg.de/) (Troje & Bülthoff, 1996) and 10 car grayscale images (size 500x500 pixels, 8-bits/pixel) (used in Philiastides et al., 2011). Using QUEST (Watson & Pelli, 1983) we calibrated stimuli in a subject-specific fashion to discriminability-levels corresponding to a performance of 60%, 70%, 80%, and 90% correct. Stimuli were presented centrally on a plain grey background on a computer screen using Psychtoolbox (Brainard, 1997; Kleiner et al., 2007; Pelli, 1997) (*also see supplemental methods).*

Behavioural Paradigm

Participants performed a visual face vs. car categorization task by discriminating dynamically updating sequences of face and car images (*Figure 1, task description*). All subjects performed the task in five experimental blocked conditions: subjects had to indicate 1.) their binary categorization response (face or car) (C), 2.) a binary response followed by a binary confidence rating in their percept (CR), 3.) a binary perceptual confidence rating followed by the binary response (RC), 4.) a binary perceptual confidence rating only (R), or 5) a binary categorization response and a binary perceptual confidence rating in one button press (i.e. face-high-confidence, face-low-confidence, car-low-confidence, car-high-confidence) (C2R2). Note that to compare response times between ratings and choices we controlled for the number of response alternatives, i.e. both choice and rating were binary. Condition order was counterbalanced across subjects. The use of left vs. right hand for confidence and choice responses was also counterbalanced across subjects. Image sequences were presented in an RSVP sequence. Each trial began with a single sequence with a series of images from one of the two stimulus classes (i.e. either a face or a car) at one of the 4 possible phase coherences. Subjects indicated their choice by pressing one of two buttons on a QWERTY keyboard for the categorization response (if required) and one of two buttons for the confidence rating (if required).

We instructed participants to respond as accurately and quickly as possible. As soon as a response was made the RSVP sequence was interrupted. The RSVP sequence was allowed to remain on the screen for a maximum of 1000ms. Then subjects had maximally two more seconds to respond. In the two conditions where a second response was required (choice-then-rating and rating-then-choice) subjects were also required to respond within 2 seconds. If subjects failed to respond within this period, the trial was marked as a no-choice trial and was excluded from further analysis. For each of the 5 conditions we presented 160 trials in four blocks of 40 trials each to allow subjects to rest briefly between blocks. 50% were face-trials and 50% were car-trials. All subjects performed 800 experimental trials. (*Also see supplemental methods*)

RESULTS

In the following we use response time 1 (RT1) for the time from stimulus onset to the first button press and response time 2 (RT2) for the time between the first and second button press (regardless whether this was a choice- or rating-response) (*see Figure 1*).

To identify whether participants engaged in second-stage processing of choice we examined the degree to which post rating decision-times vary as a function of evidence. Our analyses show, that there is an effect of difficulty on RT2 not only for confidence ratings, $F(3,78) = 4.31$, $p < .016$, $\eta_p^2 = .45$, but also for choices, $F(3,78) = 4.18$, $p < .018$, $\eta_p^2 = .32$. (*see Figure 2, right*). So subjects engaged in second-stage processing of both confidence and choice. We obtained consistent results using an independent dataset (*see supplemental results, Figure S1, Table S1+S2*).

To test whether rating confidence is associated with implicit choice, we compared RT1 in the combined choice/rating condition (*see Figure 1, Condition 5*) with RT1 in a condition where participants only rated their confidence (without explicit choice). Because in the present study confidence ratings were binary, in order to control for the number of response alternatives (i.e. 4 alternatives in both conditions), for the rating-only condition we used data from a similar study, where subjects rated confidence on a 4-point scale (Heereman, Heekeren, & Pleskac, *under review*). RT1 in the rating-only-condition, M = 1137ms, SD = 284, was significantly shorter than in the choice-rating-combined condition, M = 1279ms, SD = 243, $t(49.5) = 2.2$, $p < .037$, d = .63, r = .3 (*see Figure 3 and Table S3*).

To test whether confidence and choice processing differ in terms of their computational demands, we compared the respective response times (both RT1 and RT2), controlling for the number of response alternatives (both binary). We found that confidence and choice differ in both RT1, $F(1, 26) = 19.5$, $p < .001$, $\eta_p^2 = .42$, and RT2, $F(1, 26) = 9.78$, $p < .005$, $\eta_p^2 = .27$. RT1 in the rating-then-choice condition, M = 1281, SD = 231, was significantly longer than in the choice-then-rating condition, M = 1193ms, SD = 190, $t(26) = 4.2$, $p < .001$, d = .42, and RT2 was longer in the choice-then-rating-condition, M = 344ms, SD = 102, than in the rating-then-choice-condition, M = 293, SD = 86, $t(26) = 3.1$, $p < .005$, d = .54. So both RT1 and RT2 were longer for ratings than choices (*see Figure 2*). Interestingly, we observed an interaction of response-type and difficulty on both RT1, $F(3, 78) = 9.04$, $p < .001$, $\eta_p^2 = .44$, (*see Table S4*) and RT2, $F(3, 78) = 6.97$, $p < .001$, $\eta_p^2 = .45$. (*see Table S1*) The same interaction on RT1 was also present when we compared choice-only and rating-only trials, $F(3, 78) = 8.5$, $p < .001$, $\eta_p^2 = .42$. So in both RT1 analyses and in the RT2 analysis, we found a differential effect of difficulty on choice and confidence processing. However, we found no difference in RT1 between the choice-only, M = 1010ms, SD = 180, and rating-only condition, M = 1036ms, SD = 217, $F(1, 26) = .5$, $p = .5$, $\eta_p^2 = .02$ (*see Figure 4 and Table S5*).

DISCUSSION

In this experiment, we find evidence suggesting that not only confidence but also choice can be computed in a second stage (i.e. after the first response). Furthermore, our results indicate that judging confidence does not require implicit choice. Also, we show that confidence processing is associated with higher computational demands than choice.

RT2-changes as a function of stimulus-difficulty in choice-then-rating-designs are a well-replicated finding (Baranski & Petrusic, 1998; Moran et al., 2015; Pleskac &

Busemeyer, 2010). To our knowledge, we are the first, however, to observe the respective effect in a rating-then-choice design. That is, RT2 changes as a function of stimulus-difficulty in both conditions (*See Figure 2*). This implies that in both response orders the second response (confidence rating or decision) is, at least to some degree, computed after the first response. This extends the original finding of Baranski and Petrusic (1998) of post decision confidence processing to post rating decision processing. In addition, our finding of second stage processing of not only confidence but also choice shows that the two processes do not depend on or require each other. That is, if there was a processing hierarchy among confidence and choice, at least in one of the two response orders all processing should be completed at the time of the first response and thereby RT2 should not vary as a function of evidence. RT2-variation in both response orders therefore suggests, that confidence and choice processing are (at least in part) decoupled.

A second insight from the present work is, that RT1 is longer (controlling for the number of response alternatives) when subjects indicate both decision and rating together (i.e. in one button press) as compared to when they only rate their confidence. This indicates that rating confidence does not require implicit choice because else RT1 should be constant across the two conditions. Longer RT1 when choice is required also suggests, that even when all evidence is accumulated and readily available, actually committing to choice requires an extra computational effort.

Finally, we find confidence processing to be associated with higher computational demands than choice. That is, when comparing choice-then-rating with rating-then-choice trials we find that both RT1 and RT2 are longer for ratings than for decisions. This effect appears to be more pronounced in easy than in difficult trials, as indicated by the respective interactions of response-type and difficulty (*see Figure 5*). This interaction is even more apparent in choice-only vs. rating-only trials where we only see a trend towards longer rating-

RT1 in easy trials (*see Figure 4*). We want to point out that we had no hypothesis regarding interactions. Still, the fact that across three analyses we consistently find a differential effect of evidence (difficulty) on confidence and choice response times suggests, that confidence and choice differ not only in their computational demands, but also in their processing. If they were based on the same processes, the effect of difficulty on choice and confidence response times should be the same. In our view, the observation that this is not the case indicates, that in contrast to a modeling assumption common in the evidence accumulation framework (e.g. Moran et al., 2015; Pleskac & Busemeyer, 2010), the functions linking evidence with confidence and choice, respectively, might differ.

In summary, our results advance the understanding of confidence and choice processing in several ways. First, we find evidence indicating second-stage processing of both confidence and choice. Our results also suggest that confidence and choice processing do not require each other and are at least in part decoupled. Next, we show that even when all evidence is accumulated and readily available, actually committing to choice requires an extra computational effort. Also, we find that confidence and choice processing are associated with different computational demands. Finally, we observe that evidence has a differential effect on confidence and choice processing. Therefore, future studies should shed more light on the exact difference between the processes underlying confidence and choice.

REFERENCES

Baranski, J. V., & Petrusic, W. M. (1998). Probing the locus of confidence judgments: experiments on the time to determine confidence. *Journal of Experimental Psychology: Human Perception and Performance, 24*(3), 929.

Baranski, J. V., & Petrusic, W. M. (2001). Testing architectures of the decision–confidence relation. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale, 55*(3), 195.

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial vision, 10*, 433-436.

De Gardelle, V., & Mamassian, P. (2014). Does confidence use a common currency across two visual tasks? *Psychological science*, 0956797614528956.

Fleming, S. M., Huijgen, J., & Dolan, R. J. (2012). Prefrontal contributions to metacognition in perceptual decision making. *The Journal of Neuroscience, 32*(18), 6117-6125.

Fleming, S. M., Weil, R. S., Nagy, Z., Dolan, R. J., & Rees, G. (2010). Relating introspective accuracy to individual differences in brain structure. *science, 329*(5998), 1541-1543.

Gigerenzer, G., Hoffrage, U., & Kleinbölting, H. (1991). Probabilistic mental models: a Brunswikian theory of confidence. *Psychological review, 98*(4), 506.

Henmon, V. A. C. (1911). The relation of the time of a judgment to its accuracy. *Psychological review, 18*(3), 186.

Kleiner, M., Brainard, D., Pelli, D., Ingling, A., Murray, R., & Broussard, C. (2007). What's new in Psychtoolbox-3. *Perception, 36*(14), 14-14.

Moran, R., Teodorescu, A. R., & Usher, M. (2015). Post choice information integration as a causal determinant of confidence: Novel data and a computational account. *Cognitive psychology, 78*, 99-147. doi:10.1016/j.cogpsych.2015.01.002

Murphy, P. R., Robertson, I. H., Harty, S., & O'Connell, R. G. (2015). Neural evidence accumulation persists after choice to inform metacognitive judgments. *eLife*, e11946.

Peirce, C. S., & Jastrow, J. (1884). On small differences of sensation. *Memoirs of the National Academy of Sciences, 3*, 73-83.

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial vision, 10*(4), 437-442.

Petrusic, W. M., & Baranski, J. V. (2003). Judging confidence influences decision processing in comparative judgments. *Psychonomic bulletin & review, 10*(1), 177-183.

Philiastides, M. G., Auksztulewicz, R., Heekeren, H. R., & Blankenburg, F. (2011). Causal role of dorsolateral prefrontal cortex in human perceptual decision making. *Curr Biol, 21*(11), 980-983. doi:10.1016/j.cub.2011.04.034

Pleskac, T. J., & Busemeyer, J. R. (2010). Two-stage dynamic signal detection: a theory of choice, decision time, and confidence. *Psychol Rev, 117*(3), 864-901. doi:10.1037/a0019737

Troje, N. F., & Bülthoff, H. H. (1996). Face recognition under varying poses: The role of texture and shape. *Vision research, 36*(12), 1761-1771.

Vickers, D. (1979). *Decision processes in visual perception*: Academic Press.

Yu, S., Pleskac, T. J., & Zeigenfuse, M. D. (2015). Dynamics of postdecisional processing of confidence. *Journal of Experimental Psychology: General, 144*(2), 489.

Zylberberg, A., Barttfeld, P., & Sigman, M. (2012). The construction of confidence in a perceptual decision. *Frontiers in integrative neuroscience, 6*.

FIGURES

Figure 1, Task description: Subjects saw a sequence of dynamically updating stimuli (30 frames/second, 1 second max.) on the display. In the first condition (choice-then-rating) they indicated whether they saw a face or a car and then rated on a binary scale (low confidence vs. high confidence) their confidence in their percept. In the second condition subjects only indicated their choice. In the third condition (rating-then-choice) subjects rated their confidence first and only then indicated their choice. In the fourth condition subjects only rated their confidence in their percept without indicating their choice. In a fifth condition subjects indicated confidence and choice in one response (face high, face low, car low, car high). Subjects could give their first response already during stimulus presentation (rating or choice depending on condition). In the two two-response-conditions this was followed by a prompt for the second response (choice or rating depending on condition). RT1 = time from stimulus-onset to first response; RT2 = time from first to second response.

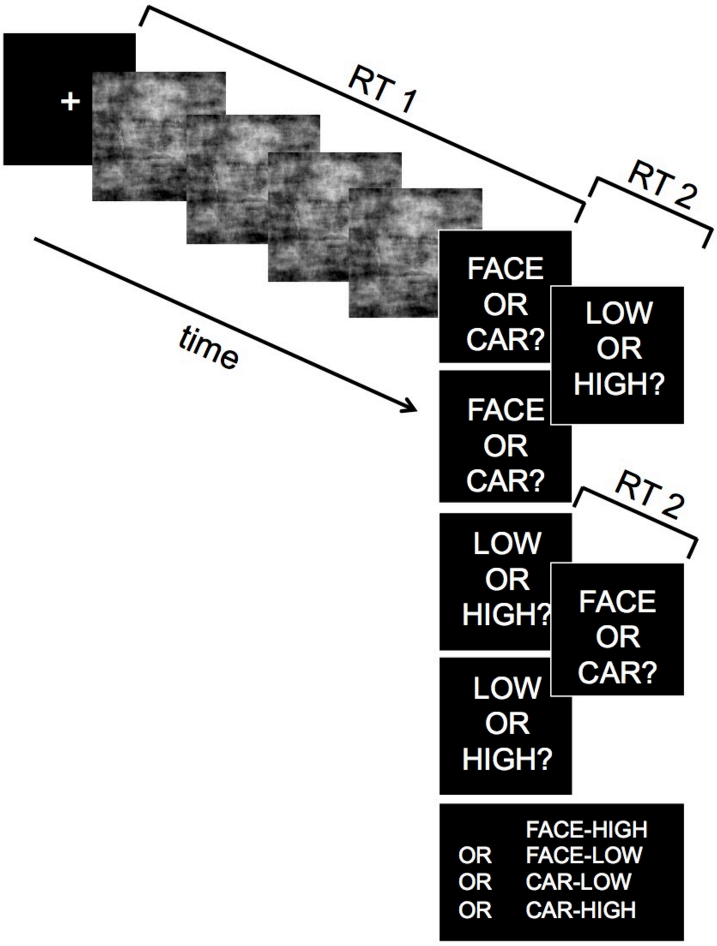


Figure 2: Both RT1 and RT2 are longer for ratings than for decisions indicating a difference in computational demands between the two processes. *Left:* RT1 for binary choice (followed by rating) and binary rating (followed by choice). *Right:* RT2 for binary rating (preceded by choice) and binary choice (preceded by rating). RT2 decreases with difficulty in binary choices and binary ratings suggesting second-stage processing of both. (error bars represent SDs)

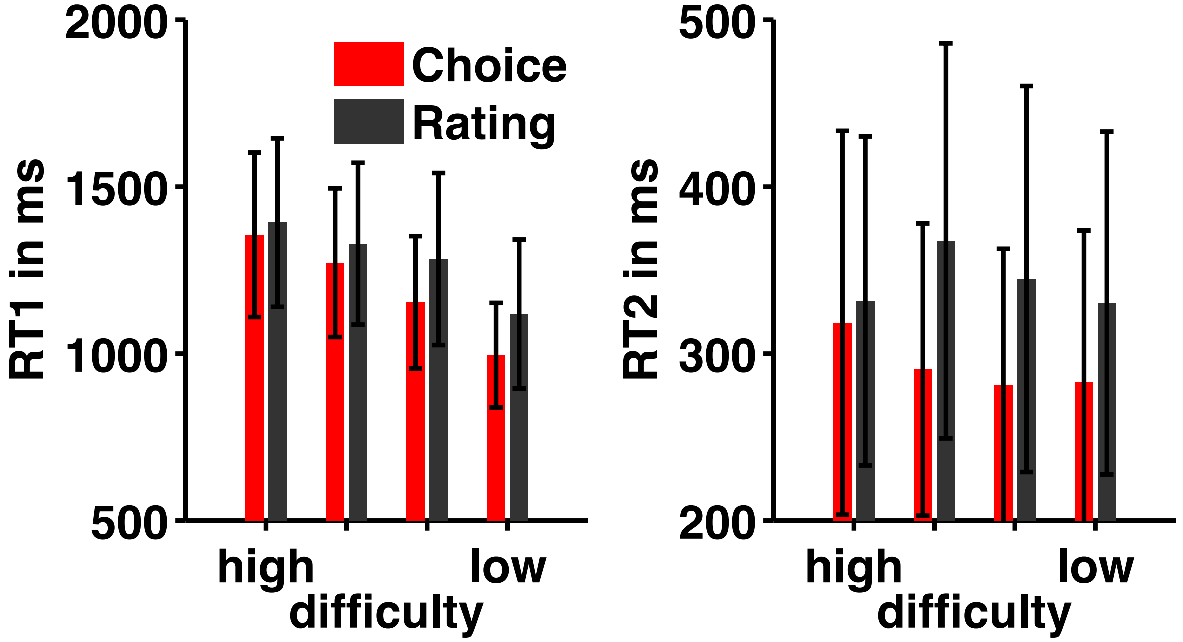


Figure 3: 4-point rating (R4) vs. 4-point combined choice (binary)/rating (binary) (C2R2). The shorter RT1 in R4 suggests, that estimating confidence does not necessarily require prior implicit choice. (error bars represent SDs)

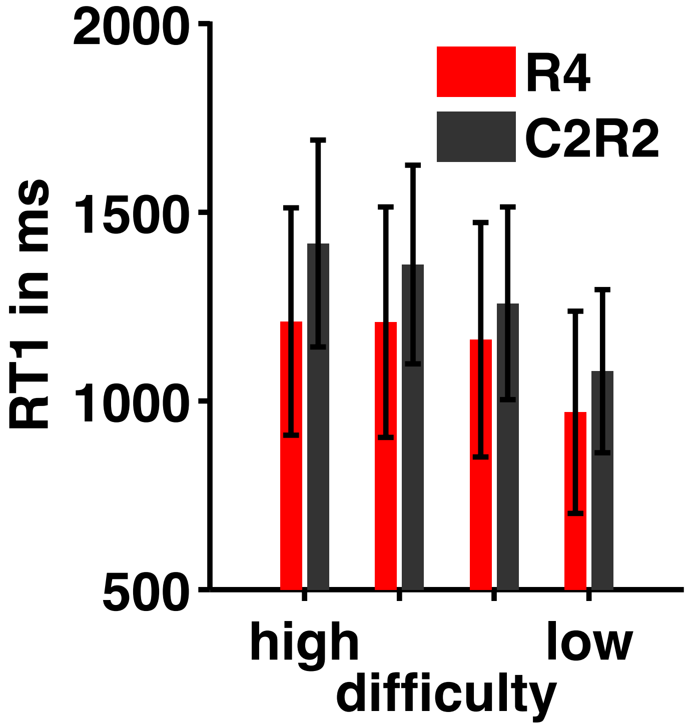
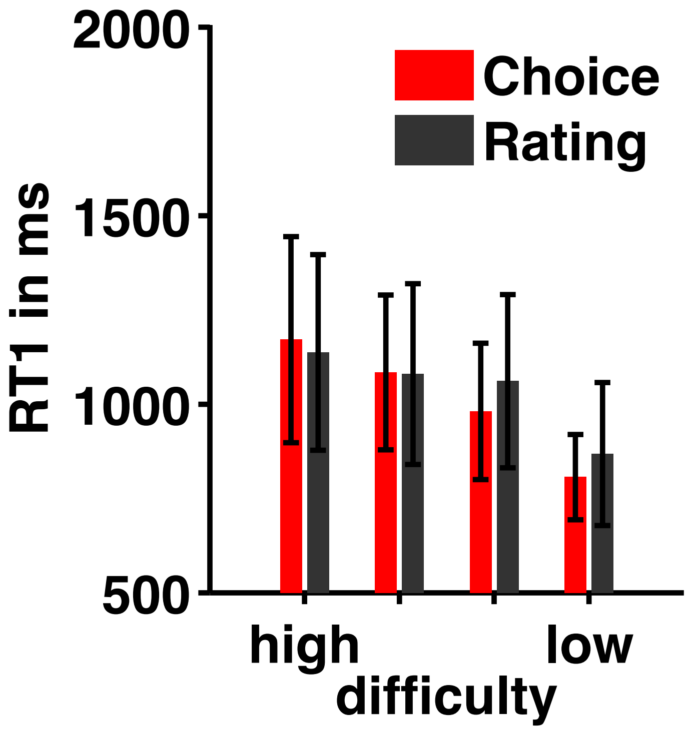


Figure 4: RT1 for binary choice-only (C) and binary rating-only (R). Although in easy trials the data show a trend towards longer RT1 in rating-only trials, the difference in RT1 does not reach significance. (error bars represent SDs)



Supplemental Information: Evidence for Decoupled Processing of Confidence and Choice

Johannes Heereman, Hauke R. Heekeren, and Timothy J. Pleskac

SUPPLEMENTAL METHODS

Stimuli

We used a set of 10 face (face database, Max Planck Institute for Biological Cybernetics, Tuebingen, Germany, http://faces.kyb.tuebingen.mpg.de/) (Troje & Bülthoff, 1996) and 10 car grayscale images (size 500x500 pixels, 8-bits/pixel) (used in Philiastides, Auksztulewicz, Heekeren, & Blankenburg, 2011). Spatial frequency, luminance and contrast were equalized across all images. The magnitude spectrum of each image was adjusted to the average magnitude spectrum of all images used. The phase spectrum of the images was manipulated to obtain noisy stimuli of varying levels of sensory evidence (i.e., we manipulated the percentage phase coherence of the images) (Dakin, Hess, Ledgeway, & Achtman, 2002).

Stimuli were presented centrally on a plain grey background on a computer screen using Psychtoolbox (Brainard, 1997; Kleiner et al., 2007; Pelli, 1997). The display was situated 0.4 m away from the subject. We used four different phase coherence values per stimulus type (face or car). At each level of stimulus sensory evidence, we generated multiple frames for each image. Within each level of sensory evidence, the overall amount of noise remained unchanged, while the spatial distribution of the noise varied across individual frames such that when presented (rapid serial visual presentation (RSVP), 30 frames/second, i.e. 33.33 ms per frame without gaps) different parts of the underlying image were revealed sequentially. Stimulus-difficulty refers to 1- phase-coherence. Using QUEST (Watson & Pelli, 1983) we determined the subject- and stimulus-specific phase coherences corresponding to a performance of 60%, 70%, 80%, and 90% correct. Mean (±SD) values were 0.126 (±0.027), 0.179 (±0.033), 0.207 (±0.032), and 0.259 (±0.032) for faces. Mean values for car stimuli were 0.146 (±0.025), 0.165 (±0.021), 0.194 (±0.015), and 0.235 (±0.012).

Preparation

Prior to calibration (400 trials) and the first experimental session, subjects were familiarized with the stimuli and the task (200 practice trials). Altogether, all subjects completed 600 trials before testing.

Response collection

Subjects indicated their choice by pressing one of two buttons on a QWERTY keyboard for the categorization response (if required) and one of two buttons for the confidence rating (if required). In condition 1-4 for the categorization response half of the subjects pressed the left and right arrows with their right hand. For the rating they pressed one of the buttons ‘a’ (uncertain) – ‘d’ (certain) with their left hand. In the fifth condition where confidence and categorization had to be indicated in one button press, they pressed one of four buttons (‘h’, ’j’, ’k’, or ’l’) with their left hand. For the other 50% of participants the response-to-hand-mapping was the opposite.

Data cleaning

Trials were excluded from further analysis if subjects failed to respond or responded in the wrong order (0.84 % of the data, 7 trials per subject on average). We further excluded trials where RT1 was shorter than 200ms and/or differed more than three standard-deviations from the mean (0.79 % of data, 6 trials per subject on average). In total we excluded 1.63 % of data which resulted in ~ 787 remaining trials per subject.

Analyses

We performed all analyses using MATLAB (Mathworks, Natick, MA, USA), and R (https://www.r-project.org/).

Presentation of Results

While for statistics we log-transformed response times, for illustrative purposes (e.g. means and figures) we did not. Significance levels for analyses of variance are based on Greenhouse-Geisser adjusted degrees of freedom although the degrees of freedom reported are based on the design. While for within-subject ANOVAS we report standard partial eta squared (η_p^2) for mixed ANOVAS we report generalized partial eta squared (η_G^2) (Bakeman, 2005). We found no interactions of the factor stimulus (face or car trial) with any of our independent variables so we collapsed face and car trials in all subsequent analyses.

SUPPLEMENTAL RESULTS

Second stage processing of confidence and choice

We obtained consistent results using an independent dataset from Heereman et al. (*under review*): Two repeated measures ANOVAs with difficulty as within-subject factor revealed an effect of difficulty on RT2 in both the choice-then-rating-condition, *F*(3,78) = 8.5, p < .001, $\eta_p^2 = .36$, and the rating-then-choice-condition, *F*(3,78) = 10.9, p < .001, $\eta_p^2 = .46$. (see Table 2 and Figure S1, left).

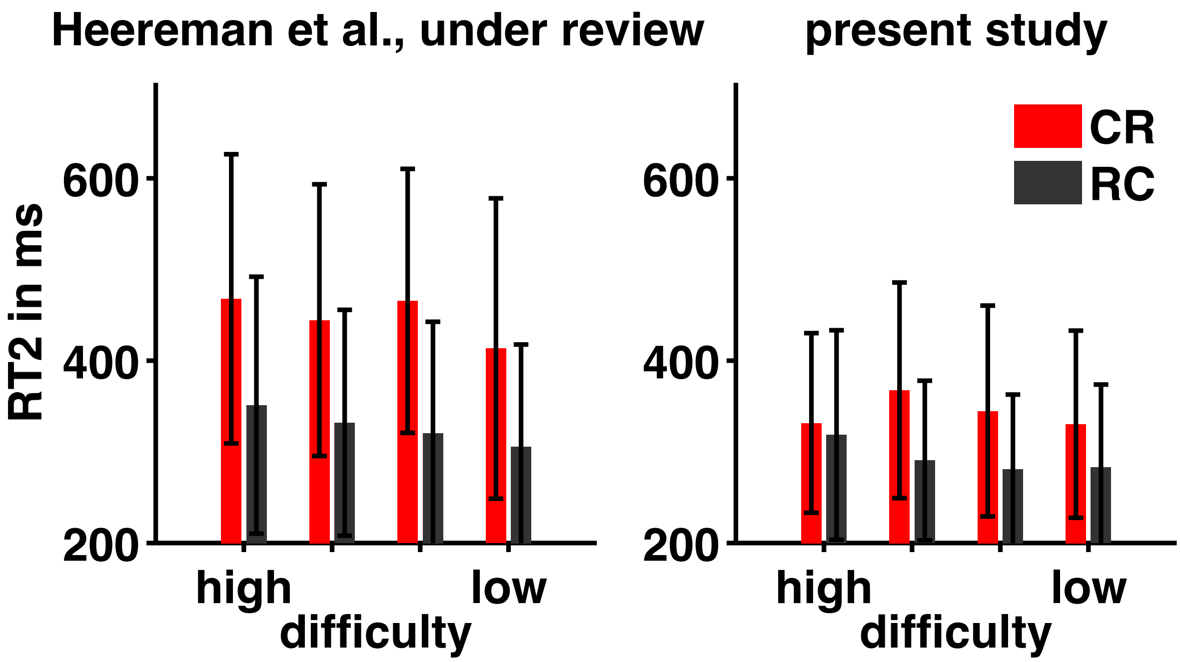


Figure S1: RT2 decreases with difficulty both when the rating (CR) and when choice (RC) is the second response suggesting not only post decision processing of confidence but also post confidence processing of choice (error bars represent SDs)

Replication of results from Heereman et al., (*under review*)

RT1 as function of response sequence length

Choice vs. choice-then-rating

We replicate the finding of longer decision-reaction-times if a subsequent confidence rating is required (as compared to a choice-only-condition) (Petrusic & Baranski, 2003; Heereman, Heekeren, and Pleskac, *under review*). As expected, a repeated measures ANOVA with difficulty and condition as within-subject factor showed the usual effect of difficulty, *F*(3,78) = 113.7, p < .001, $\eta_p^2 = .89$, and one of condition, *F*(1, 26) = 83.8, p < .001, $\eta_p^2 = .76$, on RT1. The interaction, *F*(3, 78) = 2.78, p = .063, $\eta_p^2 = .25$, was not significant. Mean RT1 was longer in the choice-then-rating-condition (M = 1216ms, SD = 303) than in the choice-only-condition (M = 981ms, SD = 185, t(26) = 6.33, p < .001, d = 2.48, r = .78).

Rating vs. rating-then-choice

We also replicate that asking for a choice response after a confidence rating leads to longer rating-reaction-times (RT1) (as compared to a rating-only-condition) (Heereman, Heekeren, and Pleskac, *under review*). As expected, a repeated measures ANOVA with difficulty and condition as within-subject factor showed the usual effect of difficulty, $F(3,78) = 78.1$, $p < .001$, $\eta_p^2 = .91$, and one of condition, $F(1, 26) = 42.63$, $p < .001$, $\eta_p^2 = .62$, on RT1. The interaction, $F(3, 78) = 2.9$, $p = .064$, $\eta_p^2 = .26$, was not significant. Mean RT1 in the rating-then-choice condition (M = 1304ms, SD = 324) was longer than in the rating-only-condition (M = 1137ms, SD = 283, $t(26) = 4.38$, $p < .001$, $d = 1.72$, $r = .65$).

Resolution of pre- vs. post-decision confidence

We fail to replicate a difference in resolution (i.e. the ability of confidence ratings to dissociate correct from incorrect trials) between pre and post decision confidence ratings (Heereman, Heekeren, and Pleskac, *under review*). A repeated measures ANOVA with choice accuracy and condition as within-subject factor showed the usual effect of accuracy, $F(1, 26) = 306.8$, $p < .001$, $\eta_p^2 = .92$, but none of condition, $F(1, 26) = .05$, $p = .8$, $\eta_p^2 = .002$, on confidence. The interaction, $F(1, 26) = 1.9$, $p = .18$, $\eta_p^2 = .07$, was also not significant. Given that we observe the same trend (*see Figure S1*) as in our previous study, in our view the lack of a significant interaction (i.e. a difference in resolution between pre- and post decision confidence-ratings) is most likely due to lower sensitivity of the binary rating used here as compared to the 4-point rating-scale used in our previous study.

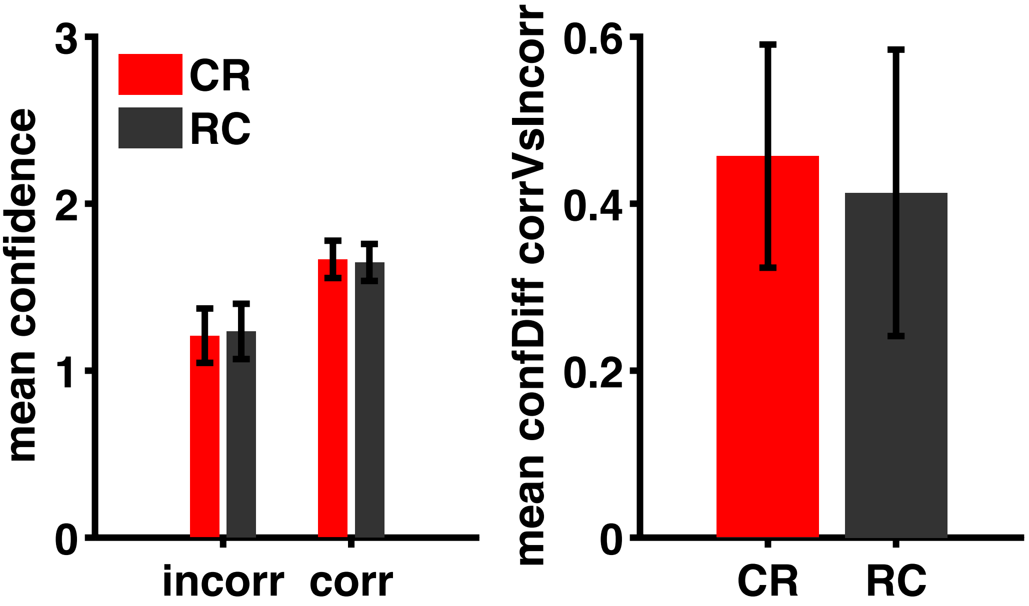


Figure S2: Confidence in correct vs. incorrect trials in choice-then-rating (CR) vs. rating-then-choice (RC) condition. *Left:* confidence in incorrect trials is slightly higher in the RC than in the CR condition. *Right:* Confidence-difference between incorrect and correct trials is higher in CR than in RC-condition. Although the differences are not significant trends indicate the involvement of an error-detection mechanism in post-decision-ratings (error bars represent SDs).

SUPPLEMENTAL TABLES

Table **S1**

Two-way ANOVA on RT2 in choice-then-rating and rating-then-choice trials.

Source	*Df*	*Error Df*	*F*	η_p^2	*p*
A Difficulty	3	78	2.76	.2	.076
B Response-condition	1	26	9.78	.27	.005
A * B (interaction)	3	78	6.97	.45	.001

Table **S2**

Two-way ANOVA on RT2 in choice-then-rating and rating-then-choice trials (Heereman et al., under review) (see Figure S1)

Source	*Df*	*Error Df*	*F*	η_p^2	*p*
A Difficulty	3	78	1.1	.32	.34
B Response-condition	1	26	45.7	.64	.001
A * B (interaction)	3	78	30.97	.65	.001

Table **S3**

Two-way mixed ANOVA on RT1 in choice/rating combined and rating-only trials.

Source	*Df*	*Error Df*	*F*	η_G^2	*p*
A Difficulty	3	156	122.04	.16	.001
B Response-condition	1	52	4.6	.08	.05
A * B (interaction)	3	156	2.3	.003	.1

Table **S4**

Two-way ANOVA on RT1 in choice-then-rating and rating-then-choice trials.

Source	*Df*	*Error Df*	*F*	η_p^2	*p*
A Difficulty	3	78	82.7	.86	.001
B Response-condition	1	26	19.5	.42	.001
A * B (interaction)	3	78	9.04	.44	.001

Table **S5**

Two-way ANOVA on RT1 in choice-only and rating-only trials.

Source	*Df*	*Error Df*	*F*	η_p^2	*p*
A Difficulty	3	78	110.6	.9	.001
B Response-condition	1	26	.5	.02	.5
A * B (interaction)	3	78	8.5	.42	.001

SUPPLEMENTAL REFERENCES

Bakeman, R. (2005). Recommended effect size statistics for repeated measures designs. *Behavior research methods, 37*(3), 379-384.

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial vision, 10*, 433-436.

Dakin, S., Hess, R., Ledgeway, T., & Achtman, R. (2002). What causes non-monotonic tuning of fMRI response to noisy images? *Current Biology, 12*(14), R476-R477.

Kleiner, M., Brainard, D., Pelli, D., Ingling, A., Murray, R., & Broussard, C. (2007). What's new in Psychtoolbox-3. *Perception, 36*(14), 14-14.

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial vision, 10*(4), 437-442.

Philiastides, M. G., Auksztulewicz, R., Heekeren, H. R., & Blankenburg, F. (2011). Causal role of dorsolateral prefrontal cortex in human perceptual decision making. *Curr Biol, 21*(11), 980-983. doi:10.1016/j.cub.2011.04.034

Troje, N. F., & Bülthoff, H. H. (1996). Face recognition under varying poses: The role of texture and shape. *Vision research, 36*(12), 1761-1771.