Master's thesis

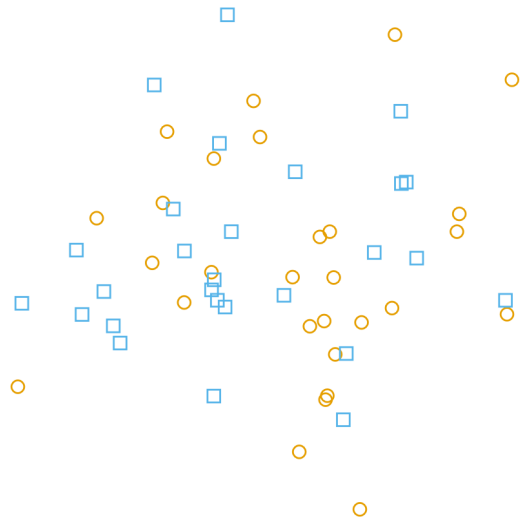
NTNU
Norwegian University of Science and Technology
Faculty of Social and Educational Sciences
Department of Psychology

Hans Fredrik Sunde

Individual Differences in Wishful Thinking

A Failed Signal Detection Experiment

Master's thesis in Psychology
Supervisor: Robert Biegler
May 2019



NTNU
Norwegian University of Science and Technology

Hans Fredrik Sunde

Individual Differences in Wishful Thinking

A Failed Signal Detection Experiment

Master's thesis in Psychology
Supervisor: Robert Biegler
May 2019

Norwegian University of Science and Technology
Faculty of Social and Educational Sciences
Department of Psychology

NTNU
Norwegian University of
Science and Technology

Individual Differences in Wishful Thinking:
A Failed Signal Detection Experiment

Hans Fredrik Sunde
Norwegian University of Science and Technology

Master Thesis
Supervisor: Robert Biegler

Table of Contents

Acknowledgements

I could not have done this project alone. Thanks to Kyrre Svarva and Sigurd Lundheim for helping with translations. Thanks to Eva J. B. Payne for helping me with my NSD application. Thanks to professor Mehmet Mehmetoglu for giving me a job as a teaching assistant in statistics, thus allowing me to develop my statistics and Stata software skills further. Thanks to my fellow students for being available for discussions, for helping with pilot testing, and for being great travel companions. Thanks to all participants who provided data for my experiment.

My supervisor, professor Robert Biegler, deserves a special thanks for his engagement and enthusiasm. Robert has always been available for discussions, and frequently stopped by the reading hall for a quick chat whenever I stayed at campus late. He has provided me with great opportunities for intellectual development, both when discussing the project and when wandering off topic.

Thanks to fellow students in *Psykologisk Tidsskrift* for making my time as a student fun, interesting, and rewarding. Thanks to my close friends for challenging me when possible and supporting me when needed. And finally, a huge thanks to my parents for their unconditional love and support.

Preface

Writing a master's thesis is an exercise in education, and with that comes a learning curve. I have learned a tremendous amount this last year, both in terms of knowledge and skills. From my newfound vantage point, I look back in hindsight at what I did further down on the learning curve. There is much I would have done differently, but unfortunately, many of the early decisions that went into the foundations of the project cannot be unmade. Some of this will be apparent from reading the thesis, some will not. Nevertheless, in terms of personal educational value, the project has been priceless. I am truly higher on the curve now than when I started, and, hopefully, will be able to climb it even higher with the tools I've attained from this project.

The project and research questions were developed jointly by my supervisor and me. My supervisor did the bulk of the experiment software programming, although I was able to learn a bit on the way. I carried out data collection and data analysis alone.

A note on language: I write "we" when describing decisions made jointly with my supervisor, and "I" when it comes to my own thoughts and decisions. I also occasionally use "we" to include you, the reader, such as in "as we can see in Figure 2".

Figures were made with the `plotplain` package in Stata, courtesy of Bischof (2017)

Sammendrag (*Summary in Norwegian*)

Bakgrunn: Ønsketenking er fremtredende i korrelasjonsstudier, men fraværende i eksperimenter hvor ønskelighet er eksperimentelt manipulert. En forklaring er at ønsketenking er en spesifikk form for interesse-sannsynlighet effekt (*stake-likelihood effect*), hvor individer overestimerer sannsynligheten for både positive og negative utfall hvis de har en interesse i utfallet. Dette kan komme av at emosjonell aktivering gjør utfallet i fokus mer tilgjengelig og på den måten fordreier forventinger. Hvis det er tilfelle, så burde mer aktiverbare mennesker være mer tilbøyelig til denne effekten.

Metode: Jeg rekrutterte 64 deltagere, og ga dem perseptuelle diskrimineringsoppgaver hvor en stimulus var assosiert med enten et positivt eller negativt utfall. Deltagere besvarte også et spørreskjema (*Emotional Reactivity Intensity Perseverance Scale*) og fullførte en læringsoppgave for å måle sensitivitet til positive og negative utfall (*reward sensitivity*).

Resultater: Den perseptuelle diskrimineringsoppgaven manglet konvergent validitet, så hoved-hypotesene kunne ikke bli testet.

Konklusjon: Mangel på konvergent validitet kom mest sannsynlig av en mislykket manipulasjon, som kan ha kommet av både metodologiske og teoretiske årsaker. Begge mulighetene diskuteres.

Abstract

Background: Wishful thinking is readily found in correlational studies, but eludes experiments where desirability is manipulated by experimenters. One explanation is that wishful thinking is a specific instance of the stake-likelihood effect, where individuals overestimate the likelihood of both positive and negative outcomes if they have a stake in the outcome. This may happen because arousal makes the focal outcomes more available, which in turn distorts expectations. If so, then more arousable people may be more susceptible to this effect.

Method: I recruited 64 participants and administered perceptual discrimination tasks where one stimulus was associated with either a positive or negative outcome. Participants also answered the Emotional Reactivity Intensity and Perseverance Scale and completed a learning task to measure reward sensitivity.

Results: The perceptual discrimination task lacked convergent validity, so the main hypotheses could not be tested.

Conclusion: Lack of convergent validity was most likely due to a failed manipulation, which could have resulted from both methodological and theoretical issues. Both possibilities are discussed.

Wishful thinking is the tendency to believe something to be true because one wants it to be true. The concept is common in folk psychology, where it is often used as a dismissive remark directed toward, say, religious or political opponents believing in an afterlife or denying climate change. Less politicised examples from everyday life are the buying of lottery tickets, the use of alternative medicine, and belief in a just world. Examples also exist with perception. A Norwegian hunter was in February 2019 charged with attempted murder after reportedly mistaking a Swedish jogger for a wild animal through his infrared scope ("Villsvinjeger tiltalt for drapsforsøk," 2019, March 22). Could his desire be responsible for his misperception?

On the one hand, wishful thinking seems like a problem in search of a solution. How can we make people more rational, and less influenced by motivation in their reasoning? On the other hand, wishful thinking offers a glimpse into the inner workings of the human mind, allowing us to paint a better picture of human nature. Regardless of which perspective you find more important, understanding wishful thinking is of great importance.

What is Wishful Thinking?

Wishful thinking occurs when preferences influence expectations. Technically it is a desirability bias, which is defined as when "the desirability (undesirability) of an outcome leads to an increase (decrease) in the extent to which it is expected to occur" (Krizan & Windschitl, 2007, p. 96). It is distinguished from mere overconfidence, which is overoptimism about one's own performance; wishful thinking is overoptimism about things outside the subject's control.

Wishful thinking is a specific form of motivated reasoning, a broad term used whenever motivational factors influence reasoning (Kunda, 1990). The term distinguishes motivated, or "hot" cognition from "cold" cognition, where normal information processing can appear to be influenced by motivations when it is not. For example, people think they are better than average drivers (Svenson, 1981). This is usually attributed to a desire to maintain a positive self-image, but it may just as well result from different opinions about what constitutes good driving, or access to more information about one's own driving decisions. In her extensive review, Kunda (1990) concludes that cold cognition alone cannot account for all results, and that motivational factors cause biased memory search and belief construction in normal cognition.

One important way motivation influences cognition is through motivation for accuracy. However, motivated reasoning is more commonly used to describe directional reasoning, where a specific conclusion, or one kind of conclusion, is *a priori* favoured over alternatives, and influences cognition in a way that increases the likelihood that such a conclusion is reached (Kunda, 1990). In principle, motivation can skew this process towards both desirable and undesirable conclusions, but both popular usage and research has usually focused on desirable conclusions. That is, the desirability bias.

One such area is studies on political predictions, where, in a typical study, people are asked for whom they will vote in an upcoming election and whom they honestly expect to win. In the 1932 US presidential election, most Roosevelt supporters expected Roosevelt would win, while most Hoover supporters expected Hoover would win (Hayes Jr, 1936). Modern studies reliably find similar overoptimism (Babad, 1997; Delavande & Manski, 2012; Krizan, Miller, & Johar, 2009), even among students of political science (Babad, 1995). The same pattern repeats when subjects predict the outcome of sports events, even when bets are used to incentivise accuracy (Babad & Katz, 1991). Studies like this demonstrate a clear link between preferences and expectations, but without the control of experimental settings, plausible alternative accounts cannot be ruled out. For example, if a person is convinced by a political argument, he or she may expect others to be swayed as well.

The preference-expectancy link comes up in experimental settings too. Experiments will commonly ask subjects about their stance on a given topic, say capital punishment, then see how their beliefs change in response to new information (Lord, Ross, & Lepper, 1979). Subjects with different prior beliefs reading the same information become on average even more polarised (Taber, Cann, & Kucsova, 2009). A more modern variant had subjects evaluate objectively unambiguous but subjectively ambiguous numerical evidence for or against a favoured political proposition, in this case whether gun control works (Kahan, Peters, Dawson, & Slovic, 2013). Not surprisingly, subjects reliably interpreted the evidence in favour of their preferred beliefs. More surprisingly, more numerate individuals were *more likely* to misinterpret the data if the correct interpretation went against their preferred beliefs, suggesting that smarter individuals engage in more motivated reasoning (see also Drummond & Fischhoff, 2017; Kahan, 2013; Kahan et al., 2012). In other words, motivated reasoning cannot be written off as a consequence of limited cognitive ability.

These studies are silent as to why one belief should be preferable to another. An experiment where desire was given a more explicit role involved parents who initially preferred home care to day care for their young children. A subset of the participants was

forced by external factors to choose day care. Participants read ambiguous empirical information about the advantages and disadvantages of the two options. Those who had to choose day care interpreted it in favour of day care (even though it went against their prior beliefs), while those who could use home care interpreted it in favour of home care (Bastardi, Uhlmann, & Ross, 2011).

The pattern here seems to be to readily accept congruent information, while scrutinizing incongruent information (Taber et al., 2009). Individuals are still constrained as to what sort of conclusion they can reach, what Kunda (1990) calls *reality constrains*, and Krizan and Windschitl (2007) call *verifiability constraints*. However, uncertainty allows plenty of wiggle room. When individuals want to reach a specific conclusion, they ask themselves "can I believe it?", and when they want to reject it, they ask "must I believe it?" (Gilovich, 1991, p. 81).

While the experiments and studies above show evidence of wishful thinking, they do not allow us to infer the mechanisms by which these effects arise. The biggest weakness is the lack of control over the content and strength of prior beliefs and preferences, which confounds possible moderators of the bias. Most studies, except Bastardi et al. (2011), also fail to explain why some beliefs are preferable to others, further clouding the search for underlying mechanisms. To better understand the mechanisms of wishful thinking, experiments must manipulate participants' desires.

Krizan and Windschitl (2007) did an extensive review of such experiments. There are not many studies like this, but a desirability bias was found in experiments using discrete outcome predictions as their dependent variable. Most of these used the Marks paradigm, where subjects are to guess whether a drawn card will be a picture card or not (Marks, 1951). The experimenters can manipulate whether picture cards are associated with gains or losses (e.g., the participant can win/lose money if a picture card is drawn), as well as the proportions of picture cards in the deck. Subjects were on average more likely to predict their preferred card than another card, even when incentivised for accuracy (Krizan & Windschitl, 2007).

However, it is impossible to derive the true subjective probability estimate of the outcome from discrete outcome predictions (just as it is impossible to tell the true probability that a political candidate would win based solely on the fact he or she won). The desirability bias was mostly evident when the probability of drawing a target card was about 50%. This can come from motivated reasoning having fewer reality constraints, but it also suggests the bias could be more of a tie-breaker in ambiguous circumstances than a substantial influence

on cognition. Desirability influenced participants' guesses, but not necessarily their subjective expectation (Windschitl, Smith, Rose, & Krizan, 2010).

Supporting the latter conclusion are the other experiments reviewed by Krizan and Windschitl (2007) which measured explicit likelihood judgements. For example, Bar-Hillel and Budescu (1995) showed participants a grid of 1000 white and pink cells where they could win (or lose) money if there were more white than pink cells. When asked how likely it was that they would win after seeing the grid, participants did overestimate the likelihood. However, when asked how likely it was that they would lose, they also overestimated that. Participants were overall no more likely to overestimate the likelihood of winning than the likelihood of losing. The other experiments reviewed by Krizan and Windschitl (2007) also failed to find a clean desirability bias.

The Stake-Likelihood Effect

Krizan and Windschitl (2007) are reluctant to say that the desirability bias does not exist, and instead propose possible mediators for when and how the effect should and should not materialise. Vosgerau (2010) presents an alternative account. He posits that overestimation of probabilities has more to do with the emotional impact of having a stake in the outcome than the outcome's desirability, and is caused by misattributing arousal for likelihood. Through a series of experiments, he demonstrates that overestimation of probabilities can be experimentally induced with arousal, regardless of whether the outcome in question was desired or not. That is, people judge both a desired outcome *and* a pessimistic outcome as more likely if they have a stake in the outcome. Wishful thinking may therefore be a specific instance of the stake-likelihood effect where the focal outcome is positive.

This is consistent with Bar-Hillel and Budescu (1995), who found that their null findings did not result from optimistic and pessimistic people cancelling each other out, but rather that people who overestimate the chance of winning also overestimate the chance of losing. The direction then is simply a result of which outcome is being considered or how the question is phrased. This also ties in with Kunda (1990), who concluded that directional motivated reasoning may result from people posing directional questions to themselves. If that is so, then experiments where the direction of the question is decided by the experimenter should not find the desirability bias per se, but may still engage many of the same mechanisms. What makes people ask certain directional questions in the first place is a question for another day, but may be related to individual differences in emotional affect

(Grafton, Ang, & MacLeod, 2012), and also the function of reason (Mercier & Sperber, 2011, 2017).

The stake-likelihood effect relates to various theories presenting feelings and affect as important proximate sources of information in decision-making, such as Damasio's (1994) somatic marker hypothesis or Schwarz' (2012) feelings-as-information theory. Newer theories of emotions stress the role of interpretation in the conscious experience of feelings, and are thus also treating feelings as sources of information (Barrett, 2017). Furthermore, studies have shown that feelings, especially level of arousal, can in some circumstances be misattributed, so-called excitation transfer (Zillmann, 1971). The go-to example is Dutton and Aron's (1974) famous study in which men misattributed arousal from being on a scary suspension bridge to feelings of attraction to the female experimenter. Another classic example is E. J. Johnson and Tversky (1983), who manipulated affect by having subjects read a brief distressing newspaper report, thereby increasing assessment of risk equally for both related and unrelated events compared to a control group. When participants read an uplifting story instead, the effect was reversed. Even though Vosgerau (2010) did not cite this paper, its effect is identical to his description of the stake-likelihood effect.

In sum, wishful thinking seems to be a stake-likelihood effect, where arousal stemming from having a stake in the outcome (i.e., preference) are misattributed for the expectation that the focal outcome – positive in the case of wishful thinking – is going to occur. This raises at least two questions: (1) are there individual differences in susceptibility to the stake-likelihood effect, and (2) does this susceptibility correlate with individual differences in arousability?

Finding and highlighting correlates of individual differences is a useful way of finding evidence of possible parameters in the underlying algorithm. Non-directional desirability effects may not be that interesting in and of themselves, but if arousability mediates susceptibility to the stake likelihood effect, and wishful thinking is a special case of the stake-likelihood effect, then it is possible that it mediates directional effects as well. Pinpointing the role of arousability and misattribution of arousal in both non-pathological wishful thinking and more pathological delusions matters because it invites research on possible interventions attempting to, say, lower or reattribute arousal. Much more work remains before such questions can be pursued, including finding whether more arousable people indeed are more prone to the stake likelihood effect. This is only possible with sufficient experimental control, which precludes focus on directional effects for now as they seem to disappear when desirability and focal outcome are manipulated by experimenters.

Even though Vosgerau (2010) found that misattributing arousal causes overestimation, it is not a given that more arousable people are more prone to this effect. Several issues come to mind. First, one study is not enough to confirm a finding, something the field of psychology has become painfully aware of in the last few years (Open Science Collaboration, 2015). Science depends on corroborating evidence from multiple methodologies. Second, even if the original finding is true (which we for the present purposes will assume), more arousable people are not necessarily more prone to *misattributing* the arousal. It is a reasonable hypothesis, though, because more arousable people presumably have more arousal to attribute, and hence more opportunities for misattribution. Other factors, such as whether more arousable people are also more easily bored, further complicate the issue. For now, that more arousable people should be more prone to the stake-likelihood effect remains a hypothesis in need of a test.

An Aside on the Function of Reason

It is epistemically irrational to behave as if something is true or more likely simply because it would be nice. A creature behaving like that would make poorer decisions, and that trait should therefore have been selected against. If such a creature does evolve, it must be because the creature gains benefits that outweigh the costs. Because humans seem to be such creatures, such benefits should exist. Two plausible situations come to mind where wishful thinking might lead to benefits. First, the desirability of an outcome coupled with the belief that one has control over said outcome naturally influences whether one believes that outcome is going to occur, and might even make it more likely if it makes people exert more effort (Krizan & Windschitl, 2007). This falls outside the usual definition of wishful thinking. The second situation is when social benefits can outweigh the costs of epistemic distortions. This is consistent with theories viewing reasoning as having primarily a social function, where having beliefs that are socially beneficial (e.g., sacred values) is more important than being epistemically rational (e.g., Haidt, 2012; Kurzban, 2011; Mercier & Sperber, 2011, 2017; Simler & Hanson, 2018; Trivers, 2011). This may be part of the explanation for why people ask themselves directional questions, and hence why wishful thinking is so ubiquitous in the naturalistic correlational studies reviewed above, but elusive in strict experimental settings. The few naturalistic experiments reviewed by Krizan and Windschitl (2007) had mixed results, but those finding a desirability bias induced motivation with an ingroup/outgroup distinction, consistent with a social function of wishful thinking. While this would be interesting to delve more deeply into, the present thesis will restrict itself to

mechanisms underlying non-social decontextualized tasks for reasons of experimental control (as discussed above). If the role of arousal and arousability is established and validated, future studies can see if the same parameter is relevant in social domains as well.

Measuring True Subjective Likelihoods

The experiments reviewed so far either used discrete outcome predictions or self-reported subjective likelihood estimates. I discussed the limitations of discrete predictions earlier, but the self-reported likelihoods also have issues. For example, people are infamously bad at introspection (Nisbett & Wilson, 1977), and their reports may engage secondary processes involved in producing a justifiable response instead of true subjective estimates (Mercier & Sperber, 2011). One way to circumvent this problem is to extract a measure of true subjective expectation from multiple discrete outcome predictions. To avoid confounds from secondary processes, one should look at earlier processes such as perception.

Studies find that motivation affects lower-level perceptual processes and the initial encoding of information. Tetlock (1985) demonstrated that motivation for accuracy did not have a retroactive effect on already encoded information, but did affect the encoding of new information. Balcetis and Dunning (2006) demonstrated that motivation affects information processing down to preconscious visual perception in a series of experiments. In one of them, both participants' conscious responses and their unconscious eye movements indicated that what they preferred to see affected how they perceived an ambiguous figure.

This also ties in with broad theories of brain function emphasising the role of expectation and prediction at all levels of processing, including perception (Clark, 2013; Hawkins & Blakeslee, 2004). The general idea here is that the mind approximates hierarchal Bayesian inference, with the continuity of cognition and perception operating by combining top-down prior expectations and bottom-up incoming information weighted by their relative subjective precision (Hohwy, 2017). What the stake-likelihood effect would do is to change the prior expectations. More specifically, arousal from the stake increases the search for and hence availability of information, which in turn changes the priors (E. J. Johnson & Tversky, 1983). We can also logically arrive at something approximating Kunda's (1990) reality constraints, in that incoming information will take precedence when its subjective precision is high compared to the prior expectations. Likewise, when the incoming information is relatively uncertain, the prior expectations take precedence.

By measuring perception under uncertainty instead of explicit predictions, we can isolate the effects from later secondary processes, as well as re-demonstrate that motivation

affects lower-level perceptual processes. Because we do not expect bias purely in the desirable direction (hence making the term desirability bias inappropriate), and because we are looking at early perceptual processes, I will henceforth use the term motivated perception.

E. J. Johnson and Tversky (1983) found that priming positive and negative affect caused a decrease and increase, respectively, in the risk assessment of undesirable events. They also show that this is independent of the semantic content that caused the mood. In other word, mood changes the availability of information used in risk assessment (i.e., prior expectations). In the stake-likelihood effect, the valence of the focal outcome accomplishes the same effect.

If the focal outcome primes the direction of memory search, and the degree of arousal determines the extent of that memory search, then more easily aroused individuals should in general have deeper memory search. A deeper memory search should result in more available information, and hence cause a biased prior with less subjective uncertainty. Furthermore, negative affect and positive affect are orthogonal (Watson, Clark, & Tellegen, 1988). It is therefore possible that individual differences in positive and negative affect selectively predicts how susceptible people are to the stake-likelihood effect when desirable and undesirable outcomes are considered, respectively.

Both positive and negative affect are multifaceted, and can be divided into distinct subcomponents, such as reactivity, intensity, and perseveration (Ripper, Boyes, Clarke, & Hasking, 2018). Emotional reactivity – how easily emotional reactions are triggered – seems synonymous with arousability in the sense used in the discussion above. To summarize, we hypothesise that specifically heightened emotional reactivity is associated with more motivated perception. Furthermore, differences in positive and negative reactivity may selectively influence motivated perception about desirable and undesirable outcomes, respectively.

Defining Motivated Perception Using Signal Detection Theory

Obtaining a continuous measure of perceptual bias is possible using signal detection theory, a mathematical tool designed to measure and distinguish *sensitivity* and *bias* in perceptual discrimination tasks (Macmillan & Creelman, 1991). Sensitivity refers to the ability to distinguish between two different stimuli (or a stimulus from noise) and bias refers to a tendency to see one or the other.

Bias is not necessarily irrational, as different outcomes can have different costs and benefits (different utility), thus making the optimal decision criterion different from the most

accurate decision criterion (Lynn & Barrett, 2014). For example, if correctly identifying an ambiguous image as containing a person wins you 100 dollars, while seeing a person when nobody is there (a false positive) only sets you back 50 dollars, then a decision criterion designed to earn the most money will be slightly biased towards seeing a person compared with the decision criterion designed for the highest accuracy. We can then define motivated perception mathematically: motivated perception is liberal bias in the absence of actual utility, or excessive bias compared to utility. From this definition it should be possible to measure motivated perception using perceptual discrimination tasks where one stimulus is preferable, but independent of perception or behaviour. Seeing one stimulus rather than the other does not offer more utility, but if the perceiver behaves as if it does, that is motivated perception.

When the utilities of two outcomes are imbalanced, as in the preceding example, the optimal decision criteria become more biased in more ambiguous, low-sensitivity situations. In other words, if there is no way to tell whether there is a person in a picture or not (sensitivity = 0), one should always see a person in the image. This is important because bias in and of itself does not tell us whether it comes from behaving as if utility is mistakenly considered or from any other cause. It should therefore be more informative to measure the difference in bias between unambiguous and ambiguous situations. If the bias is identical at various sensitivities, then we cannot tell for sure whether this is a result of flawed utility estimation. This can be incorporated by redefining motivated perception as the extent to which bias is more liberal in ambiguous situations compared to unambiguous situations. Because the direction of the stake-likelihood effect depends on the focal outcome, a positive outcome should lead to overestimating utility while negative outcomes should lead to underestimating utility.

Perceptual discrimination tasks are suitable because we can control the amount of noise and hence the freedom to interpret what is seen. In Bayesian terms, we can control the certainty of incoming information and hence how it should be weighed against prior expectations. By making one stimulus preferable, we should be able to measure bias and how that changes over various levels of noise. Furthermore, by either presenting a stimulus as desirable (say, associated with monetary gains) or as undesirable (say, associated with monetary loss), we should be able to see whether focal outcome or desirability is the key factor.

Because the present study attempts to measure an effect in a novel way, it is important to check for validity. While content validity is difficult to prove directly, one useful proxy is

convergent validity. Every measure will have variation, but it is impossible to distinguish noisy variation from meaningful variation without comparing it to something else (Carlson & Herdman, 2010). One way to solve this is by administering two different perceptual discrimination tasks and check whether the more biased individuals in one task are also more biased in the other task.

Reward Sensitivity and Utility Estimation

If motivated reasoning is about subjective utility functions, then individual differences in reward sensitivity should be involved too. One mechanism that could cause differences in reward sensitivity could be that gains and losses cause different levels of arousal, which ties in with emotional affect. One of the proximate mechanisms involved in affect is attentional biases (Grafton et al., 2012). Trait positive affect is associated with attentional bias towards positive information, while negative affect is associated with attentional bias toward negative information. Someone more attentive to positive information should presumably be more sensitive to positive outcomes such as gains, and someone more attentive to negative information should be more sensitive to negative outcomes such as losses. The subjective utility function may guide the attentional bias in the first place: If one person values gains more than another person, that person should presumably be more attentive to positive information and also be more aroused at the prospect of gains. A prediction would then be that individuals high in positive arousability should be more sensitive to positive rewards, and be more aroused by the prospect of potential gains, which in turn should lead to a bigger stake-likelihood effect when the focal outcome is positive. Likewise, individuals high in negative arousability should be more sensitive to negative rewards, and be more aroused by the prospect of potential losses, which in turn should lead to a bigger stake-likelihood effect when the focal outcome is negative.

Positive Illusions and Strategic Pessimism

There are theoretical reasons to expect people to be more biased when considering negative outcomes, especially if they have a stake in the outcome, because losses are weighed more than gains (Shepperd, Findley-Klein, Kwavnick, Walker, & Perez, 2000; Tversky & Kahneman, 1992; Weber, 1994). The standard explanation for so-called strategic pessimism is that overestimating the likelihood of negative outcomes makes people better prepared for the emotional impact associated with such outcomes. While this confuses function with mechanism (one would still need to explain why negative outcomes should produce a bigger emotional impact in the first place, see Appendix A), the proposed mechanism may still be

correct. The effect was found by Vosgerau (2010), and to a certain degree also Bar-Hillel and Budescu (1995), but the effect may be too small to reach significance in the present study. It is not integral to the research goals, though.

Strategic pessimism runs counter to Taylor and Brown's (1988) idea that humans entertain positive illusions for the benefit it brings to mental health (note the recurrent function-mechanism confusion). Nonetheless, positive illusions seem to be constrained to social beliefs (e.g., self-image) related to the social function discussed above, and is therefore of limited theoretical importance for the present study. Strategic pessimism on the other hand is more domain-general and is therefore relevant.

On Paying Participants

Paying participants based on performance is relatively common in economics, but psychologists tend to pay a flat fee instead, if any, and rely on hypothetical rewards instead (Hertwig & Ortmann, 2001). Psychological experiments have tended to find no substantial difference between hypothetical and real monetary rewards, but these have been limited to narrow domains such as temporal discounting (M. W. Johnson & Bickel, 2002; Locey, Jones, & Rachlin, 2011) or the framing effect (Kühberger, Schulte-Mecklenbeck, & Perner, 2002). Contrarywise, larger reviews from economics find divergent results, ranging from no differences to very large differences (Hertwig & Ortmann, 2001; Smith & Walker, 1993). These reviews also found that real monetary rewards resulted in less variation in performance, presumably because there was no noise from participants' different ability and willingness to entertain hypothetical rewards seriously. Because of the inconclusive findings and hence uncertainty of how valid hypothetical rewards will be in the stake-likelihood effect, and because real monetary rewards seem to result in less noisy data, using real monetary incentives should be the safer option for this study. This arguably has the added benefit of making participating in the experiment more fun and engaging compared to a flat fee, and should hence make recruitment easier.

Summary of the Hypotheses

The main goal of the present study is to validate a measure of biased perception in the form of two perceptual discrimination tasks, and to see whether it correlates with individual differences in emotional arousability. A secondary goal is to explore possible links between reward sensitivity, arousability and the stake-likelihood effect, and see how the effect differ with potential gains versus potential losses. The hypotheses and analysis plan were pre-

registered on the Open Science Framework (Sunde & Biegler, 2019), and is available at http://osf.io/vq93j/.

The first hypothesis is that bias difference in the two perceptual discrimination tasks should be positively correlated (convergent validity). Carlson and Herdman (2010) recommends a correlation of $r > .70$ as a benchmark for convergent validity, and anything between $r = .50$ and $r = .70$ as warranting further inspection. Anything below $r = .50$ is not measuring the same thing. The pre-registered analyses assume that this correlation is $r > .50$. If it is, the two scores will be averaged to test the remaining hypotheses. If convergent validity is not attained, then subsequent planned analyses will be uninformative.

I previously defined motivated perception as more liberal bias in ambiguous situations compared to unambiguous situations (i.e., bias difference). Because bias in the unambiguous situations will be subtracted from ambiguous situations, and because liberal bias is indicated by a negative value (see method), lower scores mean more motivated perception. Somewhat counterintuitively, then, variables though to predict more motivated perception should show a negative effect[1].

The second hypothesis – the main hypothesis – is that positive and negative reactivity should negatively correlate with bias difference. This directly tests whether more emotionally arousable people are more prone to motivated perception. This hypothesis can also be divided into three sub-hypotheses: (a) Positive reactivity should negatively correlate with bias difference in the gain group, independent of negative reactivity; (b) Negative reactivity should negatively correlate with bias difference in the loss group, independent of positive reactivity; and (c) if there is a difference between the gain group and loss group, we expect the loss group to have a larger bias difference (i.e., strategic pessimism).

The remaining hypotheses are more exploratory, but because we have a prior expectation of the direction of the effect, they were pre-registered as well. More complex relationships, such as mediation, may be explored depending on the results. A within-subject design would have been better at exploring these, but the study was designed with the main hypothesis in mind (see methods).

[1] This is so counterintuitive that the pre-registered hypotheses and analysis plan were specified wrong. For example, more emotional arousability is hypothesized to mean more motivated perception, which sounds like a positive effect but mathematically is a negative effect. When the hypotheses and analysis plan were written, we unfortunately wrote that we expected a positive effect whenever we expected a negative effect, and vice versa. This error was only made for the hypothesis were bias difference was the outcome variable (i.e., hypothesis 2, 5 and 6).

The third hypothesis is that positive reactivity should positively correlate with sensitivity to positive feedback, and the fourth hypothesis is that negative reactivity should positively correlate with sensitivity to negative feedback.

The fifth hypothesis is that sensitivity to positive feedback should negatively correlate with bias difference in the gain condition, while the sixth hypothesis is that sensitivity to negative feedback should negatively correlate with bias difference in the loss condition.

Method

Participants

I recruited 64 subjects with flyers on campus, word of mouth among recruited participants, and direct solicitation (Age[2]: $M = 22.7$, $SD = 2.2$, 63% female). They were most likely students from the Norwegian University of Science and Technology (I did not ask). Prospective participants were told that they could get money in the experiment, and the total amount depended on both random factors and their performance (Mean payment in Norwegian kroner[3] = 92.7, SD = 43.6). All participants signed a consent form approved by the Norwegian Centre for Research Data prior to the experiment and were given the opportunity to quit at any time (see Appendix G and Appendix H). Nine participants were excluded from all analyses unless otherwise stated based on criteria in the pre-registration, described in more detail below. Additional exclusions were made on a per-analysis basis and described where relevant.

Data was collected from the first day of February to the last working day of March 2019, which was the stopping rule specified in the pre-registration. The initial goal was a sample size of N > 100, but data collection got started late due to programming issues resulting in a lower than ideal sample size. As a result, the present study is underpowered, only being able to find correlations of $r > .35$ with a power of .80 (not taking multiple comparisons into account). Based on a review of effect sizes in individual difference research, Gignac and Szodorai (2016) recommends correlations of $r = .10$, $r = .20$, and $r = .30$ to be considered small, medium and large, respectively. The present study is thus only able to find big effects. Smaller effects of interest should be highlighted and discussed, but they will be difficult to conclusively distinguish from effects arising by chance.

[2] As required by the agreement with NSD, the data must properly anonymized before it can be made public. This includes removing age groups with only a single person. Five participants aged 26 or more were recoded to age 26, meaning the actual mean is slightly higher.

[3] Average conversion rate in data collection period: 1 USD = 8.59 NOK (Norges Bank, n.d.)

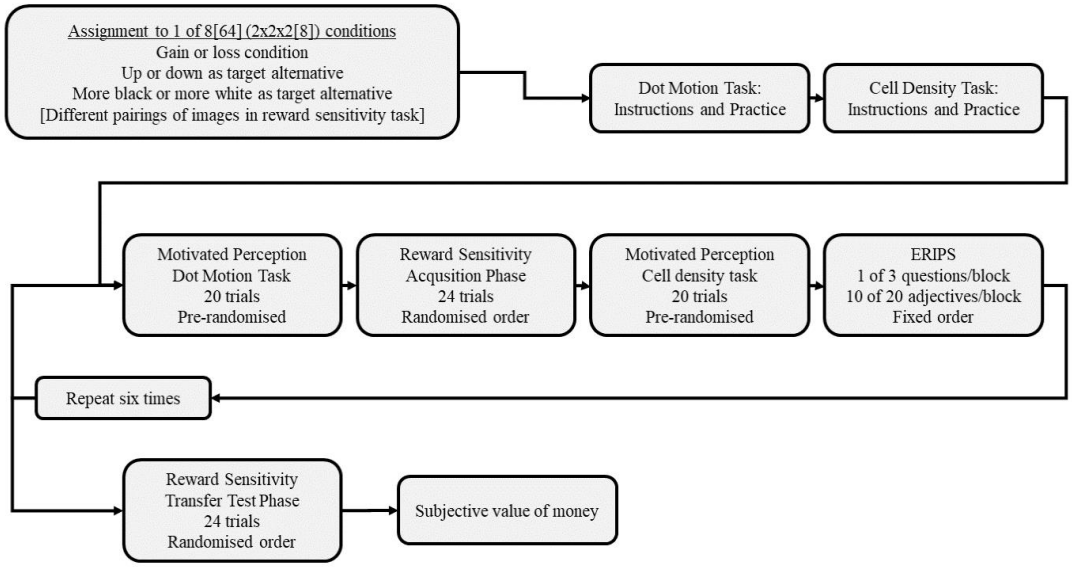


Figure 1 Outline of the experiment

Procedure

The entire experiment was administered via open-source software (PsychoPy, Pierce & MacAskill, 2018) on a laptop computer provided by the experimenter. The experimental materials will be made available at https://osf.io/vq93j/. Screenshots from the experiment, including instructions and stimuli, are available in Appendix E. The participants were tested one by one at their own convenience in various locations. Each session lasted approximately 40 minutes, including instructions and debriefing. Participants were assigned to conditions pseudorandomly, based on the order of participation. That is, the first participant was assigned the first condition, the second participant was assigned the second condition, et cetera, restarting the process when all conditions were assigned. Because the order of participation was outside my control (e.g., participants' own schedules, no-shows, etc.), no systematic bias should have played a role in which participants were assigned which condition[4].

The experiment consisted of alternating blocks of two perceptual discrimination tasks, a learning task, and a questionnaire. Age and sex were recorded at the beginning of the experiment, while a measure of participants' subjective value of money ended the experiment. The last measure was added as a possible covariate because how disposed

[4] Unfortunately, this deviates slightly from the procedure described in the pre-registration, which specified drawing randomly without replacement until all conditions were assigned, then restarting the process. That was written in an early draft of the registration, before later procedural changes made that impractical. The error was not spotted before submission.

someone is to motivated reasoning about monetary gains and losses may depend on their economic freedom. There is much variation in where students get money, not all of which are captured well by traditional measures of socioeconomic status (such as income, or parents' income). For example, students may or may not count the student loan as income, they may or may not receive financial support from their parents, and they may or may not count that support as income. Furthermore, students with the same income may have different levels of expenses (e.g., rent) that result in various levels of economic freedom. To circumvent these problems, we attempted to measure the subjective value of money directly by asking for how long a participant would generally be willing to work for 100 NOK. This also had the added benefit of retaining more anonymity.

Perceptual discrimination tasks. There were two different perceptual discrimination tasks, a dot motion task and a cell density task, where participants indicated which of two stimuli they thought they saw. They are described in more detail below. At the beginning of the experiment, participants were assigned to either a gain condition or a loss condition. In the gain condition, participants could potentially win money, while in the loss condition, participants were given money beforehand which they could potentially lose. The assignment to the gain or loss condition was consistent across both tasks. While a within-subject comparison of the gain and loss condition would be more informative, we do it between-subjects to shorten the already lengthy experiment, as well as to prevent making the goal of the experiment too obvious to participants. The difference between gain and loss condition is, after all, not the primary research question.

Participants were also assigned a target alternative in both tasks. Before the first block of each task, participants were informed that a random trial would be selected after the experiment, and they would win or lose money (50 NOK) if that trial had the target alternative, regardless of what they thought they saw. For example, a participant in the dot motion task would win or lose money if the motion happened to be mostly up, while another participant would win or lose if the motion happened to be mostly down. In the cell density task, the target alternatives were more white cells and more black cells. Assignment was consistent across trials for each participant.

In total, there were eight between-subject conditions, from the combinations of potential gain versus potential loss, dot motion target alternative, and cell density target alternative. Only the gain/loss condition is of relevance to the experiment, while the other 2x2 conditions are for counterbalancing only. Regardless of condition, participants were also given a monetary incentive to answer accurately. They were told they would receive 20 NOK

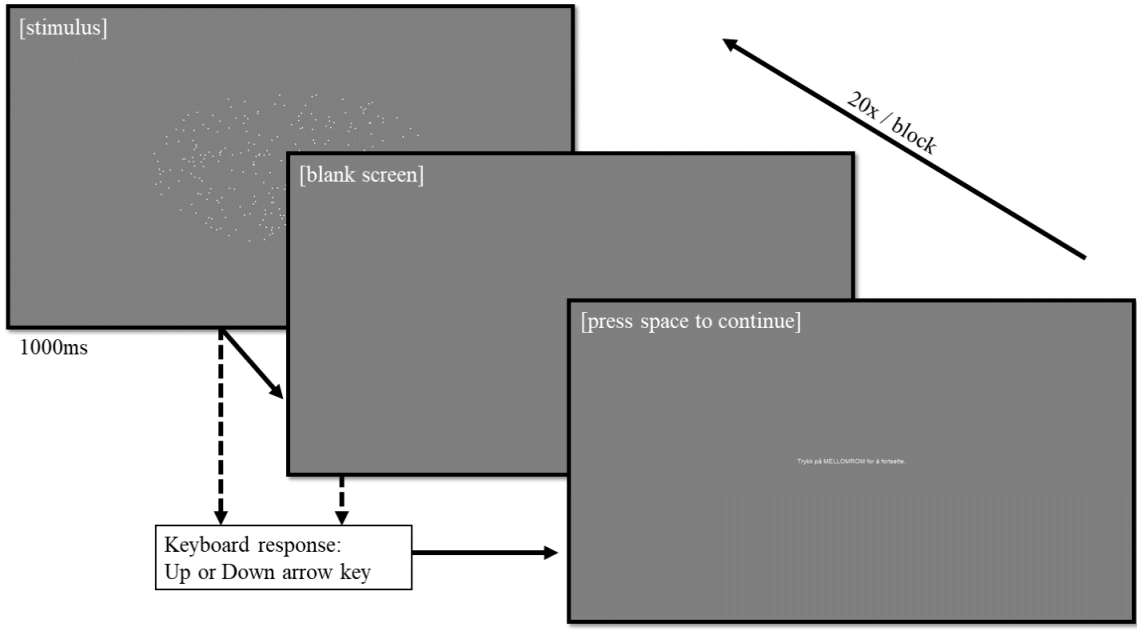


Figure 2 The dot motion task. 20 trials with 3 difficulty levels repeated in 6 blocks. Participants could respond as soon as the stimulus appeared. If participants did not answer for 5 seconds, a small reminder appeared on the blank screen reminding them which buttons they could press.

if they answered >64% correct on the dot motion task, and the same if they answered >64% correct on the cell density task, meaning a possible total of 40 NOK.

Every block started with an exclusion test, where participants were asked to remember what their target alternative was and press the corresponding key. Feedback was given immediately. They were informed that failing to answer correctly would make them either unable to win the 50 NOK (in the gain condition) or lose 50 NOK anyway (in the loss condition). Seven participants failed two or more exclusion tests in at least one of the tasks and were therefore excluded from analyses because they did not remember the target direction even when given feedback. Another two participants were excluded because they had accuracy scores below 50%. Both exclusion criteria were pre-registered.

Dot motion task. We presented participants with multiple trials of a dot motion task, each with a stimulus duration of up to 1000ms. The stimulus was dots moving mostly randomly across the screen, occasionally with a subset of dots moving systematically in one direction. Participants were to indicate whether they thought the net movement was up or down by pressing one of two keys. Participants could familiarise themselves with the task with 8 initial practice trials before testing began. During testing, there were three difficulty levels: easy, hard, and impossible. In the easy and hard conditions, there was an objective reality to classifying movement as up or down, with a coherence of .16 and .08 respectively.

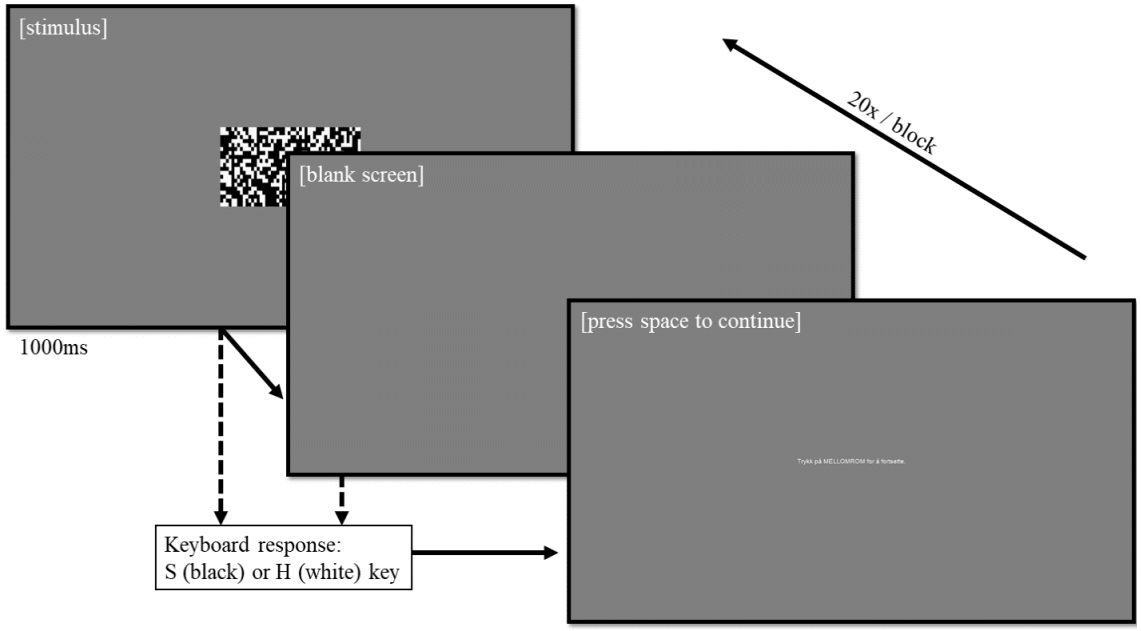


Figure 3 The cell density task. 20 trials with 3 difficulty levels repeated in 6 blocks. Participants could respond as soon as the stimulus appeared. If participants did not answer for 5 seconds, a small reminder appeared on the blank screen reminding them which buttons they could press.

In the impossible condition, the net movement was 0. The participants were only told that there would be easy and hard trials, but not that there would be impossible trials. The hard trials were included mainly to make the impossible trials less conspicuous and are not of direct relevance to the hypotheses. The task was separated into six blocks, and each block consisted of eight easy, four hard, and eight impossible trials (20 trials per block, 120 in total) in a pre-specified random order. The order was pre-specified so that a random trial more easily could be selected afterwards, and its net movement identified.

Cell density task. This task was adapted from the first experiment in Bar-Hillel and Budescu (1995), but changed substantially to be identically structured to the dot motion task. We presented participants with multiple trials with a 20x50 grid of randomly distributed black and white cells as stimulus. For each trial, the grid was shown for up to 1000ms, and participants indicated whether they thought there were more black cells than white cells or vice versa by pressing one of two keys. The grid was presented in five different proportions, analogous to the five conditions in the dot-motion task. Pilot testing indicated a general bias towards seeing more white cells, which we tried to compensate for by skewing the distribution of proportions towards black. That is, the impossible trials actually had a distribution of 51% black cells, while the other trials were centred around this. The easy trials had 47.5% and 54.5% black cells, and the hard trials had 49.5% and 52.5% black cells. The

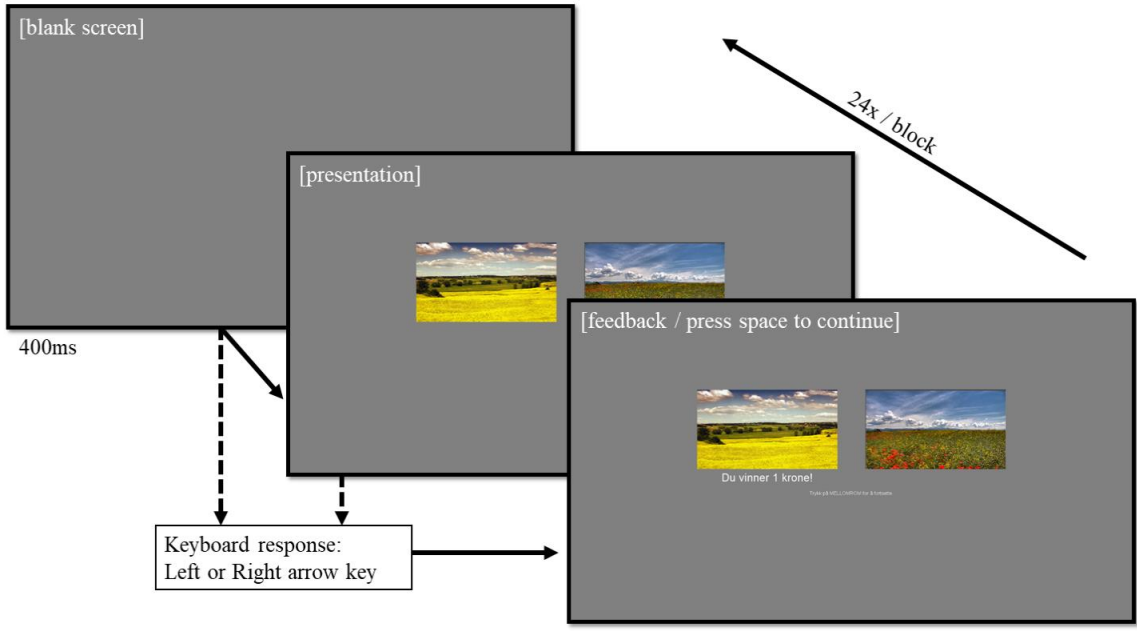


Figure 4 The acquisition phase of the learning task. There were four pairs of images.

numbers of blocks and trials were identical to the dot motion task, and the order was random but pre-specified.

Learning task. The learning task measures sensitivity to negative and positive outcomes (i.e., reward sensitivity), and was adapted from Gold et al. (2012). Contrary to the original study, we used real money as rewards and punishment. The participant could win or lose 1 NOK in each trial. The task consisted of two phases: the acquisition phase and the transfer test phase. The original study used 4 blocks of 40 trials (160 trials in total) for the acquisition phase, while we used 6 blocks of 24 trials (144 trials in total). We changed the number of blocks and number of trials per block so that the procedure would more readily fit in with the rest of the experiment. Fewer number of trials per block and fewer total trials should make the learning task more difficult. This was deliberate, as pilot testing suggested we would run up against ceiling effects. For the same reason, we also omitted practice trials.

In the acquisition block, participants were repeatedly presented with one of four pairs of landscape images (eight images in total) and asked to select the correct one. They had to learn which images were correct through trial and error, and were given immediate feedback on their selection. Two pairs involved potential gains, where participants could be rewarded if they chose the correct image ("You win 1 krone!"), while nothing usually happened if they chose the wrong image ("No change"). One pair rewarded the correct response 90% of the time and the wrong response 10% of the time. The other pair rewarded the correct response

80% of the time and the wrong response 20% of the time. The other two pairs involved potential losses, where nothing usually happened if participants chose the correct image ("No change!"), while they could be punished if they chose the wrong image ("You lose 1 krone!"). Here again, the two pairs were divided into a 90% condition and an 80% condition, specifying the probability of avoiding loss given the correct response. That is, correct responses were punished 10% or 20% of the time, and wrong responses were punished 90% or 80% of the time, respectively.

The order of the different pairs was random for each participant, with each pair being shown six times per block. There were 8! = 40320 possible combinations of the eight images, so we could not counterbalance all possible combinations. Instead, we only counterbalanced such that each image was assigned each possible role, but was only paired with one other image per role. While this is not perfect counterbalancing, it should add some noise if some images were easier to remember than others. Images can be assigned to eight different roles (2 roles times 4 pairs), meaning we had another eight between-subject conditions in addition to the 2x2x2 from the perceptual discrimination task. Again, most of these are for counterbalancing purposes. The assignment to conditions were again done based on the order of participation, so that which image had which role changed every eight participants. That the sample size matches the number of conditions (64) is coincidental.

The final block of the experiment was the transfer test phase, where we presented novel pairings of the eight images. The transfer test phase measured how well learning based on negative and positive outcomes transferred to new decisions. The participants were again instructed to select the image they thought was correct based on their earlier learning, and that they would be rewarded according to performance, but no feedback was given. Contrary to Gold et al.'s (2012) original procedure, we only showed the 12 of the 24 possible novel pairings that were informative about relative sensitivity to positive and negative outcomes[5]. For example, participants had to select between a picture that most likely would have gained them a reward and a picture that most likely would have avoided a loss. Six of the novel pairs are informative about positive reward sensitivity, while the other six are informative about negative reward sensitivity. Each novel pairing was shown twice, yielding a total of 24 trials. Whether the image was displayed on the left or right was counterbalanced across the two trials each pair was shown. Positive and negative sensitivity can be measured by the

[5] See Appendix B for an in-depth explanation. The other 12 pairings distinguish stimulus-response learning (model-free) from stimulus-response outcome learning (model-based), which is not relevant to the present study.

proportion of optimal responses in the corresponding pairs. The two scales have a possible range of 0 to 1, where a score of 0.5 indicates indifference and 1 indicates highly sensitive to the relevant feedback.

Emotional arousability. To measure emotional arousability, we administered the Emotional Reactivity Intensity Perseverance Scale (ERIPS, Ripper et al., 2018). While previous measures of emotional affect that supposedly measures reactivity exists, such as the Affect Intensity Measure (Larsen & Diener, 1987), there seems to be some confusion as to what it actually measures (Rubin, Hoyle, & Leary, 2012). The Positive Affect Negative Affect Schedule (PANAS, Watson et al., 1988) measures general affect, and is not specific enough to target arousability. The ERIPS solves this problem by using the same 10 positive and 10 negative adjectives as the PANAS, but rather than asking to what extent the subject generally feels a particular feeling, the ERIPS asks how the subject would compare themselves to the "average person" in terms of reactivity (how likely it is that they will experience this feeling), intensity (how intense it is), and perseveration (how long it persists). The ERIPS yields six different factors: positive reactivity, intensity, perseveration, and negative reactivity, intensity, and perseveration. The reactivity factors should, as mentioned, be equivalent to arousability. The scale was presented in six different blocks, each corresponding to one factor. The order of the blocks was equal for all participants, as was the adjectives. The order was as follows: positive reactivity, negative perseveration, positive intensity, negative reactivity, positive perseveration, negative intensity.

We found no existing Norwegian translations of the ERIPS, and prior translations of the PANAS were poorly documented or incomplete (e.g., some items were missing). We therefore had it re-translated and compared to existing translations by a third party. Because

Table 1

Means, standard deviations, correlations, and Cronbach's alpha for ERIPS (N=64)

	M	*SD*	1	2	3	4	5	6	7	8
1. Positive Reactivity	3.39	0.50	(.65)							
2. Positive Intensity	3.42	0.57	.54***	(.78)						
3. Positive Perseveration	3.23	0.55	.70***	.76***	(.80)					
4. Negative Reactivity	2.97	0.76	-.20	-.03	-.09	(:87)				
5. Negative Intensity	3.02	0.74	-.24*	.18	.00	.84***	(.87)			
6. Negative Perseveration	2.92	0.73	-.39**	.00	-.15	.77***	.84***	(.88)		
7. Average Positive Affect	3.35	0.48	.83***	.88***	.93***	-.12	-.02	-.19	(.89)	
8. Average Negative Affect	2.97	0.70	-.29	.05	-.09	.93***	.95***	.93***	-.12	(.95)

Note: Cronbach's alphas in the diagonal. Originally excluded participants are included here because the reasons for their exclusions should not affect their scores on the ERIPS.

*p<.1 | **p<.05 | ***p<.001

the prior translations were undocumented, we did not know the reasoning behind choices, and chose therefore differently where we thought we had good reasons to do so. The translated version was then translated back into English by another third party and compared with the original questionnaire to check that meaning was conserved. Ambiguous cases were discussed and, in some cases, changed. See Appendix F for details.

The correlations between the positive factors are roughly the same as in the original ERIPS paper (Ripper et al., 2018), but slightly higher for the negative factors (see Table 1). The alphas are somewhat lower than the original paper, but still mostly adequate. It is difficult to tell whether this results from actual poorer internal consistency or just lower sample size. The exception is positive reactivity, which has mediocre internal consistency ($\alpha = .65$). It is not clear why this is, but positive reactivity was the first block of questions, and participants did not have the opportunity to go back and change their answers. The low internal consistency may be because it took a few questions before participants understood the questions well. However, I cannot tell for sure because they were not counterbalanced. Because the sample size is relatively low, and because the sample size to item ratio is so low (1.07), a factor analysis would not be productive.

Why There is no Control Group

The experiment lacks a group with fixed payment regardless of performance. This is mainly due to budgetary concerns, seeing as the main research question is finding the correlation between bias and arousability, not the causation from stakes to bias. Holding the budget constant, adding a control group[6] would further reduce statistical power for the main research question.

Note, however, that by combining the various counterbalancing and experimental conditions, we can reach a lot of the same conclusions as we could have with a control group. If we compare the counterbalanced target alternative groups (say, “up” versus “down” in the dot motion task) and the physical biases differ, we know that it was because they were told to pay attention to different directions. We can therefore conclude whether asking participants to

[6] To disentangle the various plausible effects of stakes and direction priming, a basic control group would not suffice. We would need five different control groups with a fixed payment: One basic group where participants are simply asked to indicate what they saw, and four groups where they are told to pay attention to combinations of either “up” or “down”, and “black” or “white”. Additionally, they would need to be collected at the same time (to make sure we are sampling the same population), and for ethical reasons paid the average amount expected in the experimental conditions. While this may reveal useful information, it would drastically reduce the power for our main research question given a fixed budget and time window, and the control groups would only be informative if a correlation was found in the first place.

pay attention to a particular direction causes a bias in that direction without a control group, and whether more easily aroused individuals are more susceptible to this effect. If the physical biases did not differ across conditions, a control group would still not allow us to disentangle a true null finding from an invalid measurement.

Analysis

Data and Stata analysis scripts are available at https://osf.io/vq93j/. Before analyses could begin, the measures on the perceptual discrimination tasks had to be prepared. I calculated each participants' hit rate (the rate at which they correctly identified the target alternative) and false alarm rate (the rate at which they incorrectly identified the target alternative) separately for all conditions. These were then transformed to *z*-scores, which were used to calculate sensitivity (*d*') and bias (*c*). Sensitivity is the difference between the two z-scores, and can be interpreted similar to Cohen's *d*. Bias is the negative average, and can be interpreted as the deviation from the most accurate decision criterion assuming equal variation and base-rate (Macmillan & Creelman, 1991). Summary statistics are presented in Table 2

Because there were no objectively correct decisions in the impossible conditions, the corresponding sensitivity should in principle be 0 for all participants. There is random variation in the data around this value, but it should be of no consequence to the analyses because bias is mathematically independent from sensitivity (Macmillan & Creelman, 1991).

Table 2

Summary statistics of sensitivity and biases

	% correct		*d'*		*c*		physical[a] *c*		centred *c*	
	M	SD	M	SD	M	SD	M	SD	M	SD
Dot Motion Task										
Impossible	0.50	0.07	-0.01	0.36	0.01	0.49	-0.13	0.47	0.01	0.47
Hard	0.73	0.11	1.37	0.69	0.01	0.45	-0.13	0.44	0.01	0.44
Easy	0.80	0.12	1.93	0.92	0.07	0.45	-0.11	0.45	0.07	0.44
Bias Difference					-0.06	0.23			-0.06	0.23
Cell Density Task										
Impossible	0.50	0.07	0.00	0.41	-0.05	0.61	-0.41	0.44	-0.07	0.44
Hard	0.76	0.10	1.58	0.66	-0.01	0.53	-0.32	0.42	-0.03	0.41
Easy	0.80	0.09	1.98	0.58	-0.02	0.60	-0.42	0.42	-0.05	0.41
Bias Difference					-0.03	0.29			-0.02	0.28

a: Bias towards up in the dot motion task, bias towards white in the cell density task
Note: Notice how the means are similar but the standard deviations are smaller in the centred biases compared to the original biases. *d'* = Sensitivity, *c* = Bias

Table 3

Independents t-tests for whether biases must be centred

	c_{down}		c_{up}				
Dot Motion Task	M	SD	M	SD	*MD*	t(53)	*p*
Impossible	0.13	0.60	-0.12	0.31	-0.26	-1.99	.052
Hard	0.13	0.52	-0.12	0.34	-0.25	-2.11	.040
Easy	0.18	0.55	-0.04	0.30	-0.21	-1.78	.080
	c_{black}		c_{white}				
Cell Density Task	M	SD	M	SD	*MD*	t(53)	*p*
Impossible	0.34	0.36	-0.49	0.51	-0.83	-7.00	<.001
Hard	0.29	0.39	-0.35	0.45	-0.65	-5.74	<.001
Easy	0.38	0.40	-0.48	0.44	-0.85	-7.59	<.001

Bias can be both liberal and conservative, with negative values indicating liberal bias towards the target alternative (i.e., a low threshold for detecting the target).

Transformation of hit rates and false alarm rates to z-scores are not possible if the scores are 0 or 1. There were a few such instances in the data set, which had to be transformed first. There were 24 trials in total where the target alternative was shown in the easy dot motion trials. If a participant correctly identified all of them (a hit rate of 24/24 = 1), we transformed her hit rate to 23.5/24 = .98. Likewise, if she failed to identify any of them (a hit rate of 0/24 = 0), we transformed it to 0.5/24 = .02. This transformation was pre-registered and done to all conditions.

Next, I checked whether the distributions were bimodal or unimodal. A consistent physical bias towards seeing, say, dots moving up would skew the bias distribution of each target alternative in opposite directions, thus artificially inflating the variance. Those with up as their target alternative would appear to be more biased toward their target alternative, while those with down as their target alternative would appear to be less biased toward their target alternative. As specified in the preregistration, conditions showing significant bias differences ($p < .05$) between target alternatives were centred around the average physical bias (see Table 3).

The differences in the cell density task were all significant (all *p*s < .001) due to a large general bias toward seeing white (averages ranging from $c = -0.32$ to $c = -0.42$). The dot motion task showed some general bias toward seeing up (averages ranging from $c = -0.11$ to $c = -0.13$), but the difference between target alternatives was only significant in the hard condition ($p = .04$). The other two conditions were close (impossible: $p = .052$, easy: $p = .08$),

but because the threshold for transformation were set at $p < .05$, they were not centred for the preregistered analyses.

Last, I made the bias difference scores by subtracting the bias in the easy condition from the bias in the impossible condition for both the dot motion task and the cell density task (using the centred biases for the cell density score).

Results

The Perceptual Discrimination Tasks

The first hypothesis is that the two bias difference scores should measure the same thing (i.e., convergent validity). This is tested with a simple correlation. The preregistered threshold for moving on to subsequent analyses is $r > .50$, but it should ideally be higher. However, the correlation is only $r = .06$, 95% CIs [-.21, .32]. As we can tell by the confidence intervals, the correlation is significantly[7] lower than $r = .50$, the smallest effect size of interest (see Lakens, 2017). Also, as we can see in Figure 5, the low correlation seems to be accurate, and not caused by, say, outliers. In other words, the main measure that the planned analyses

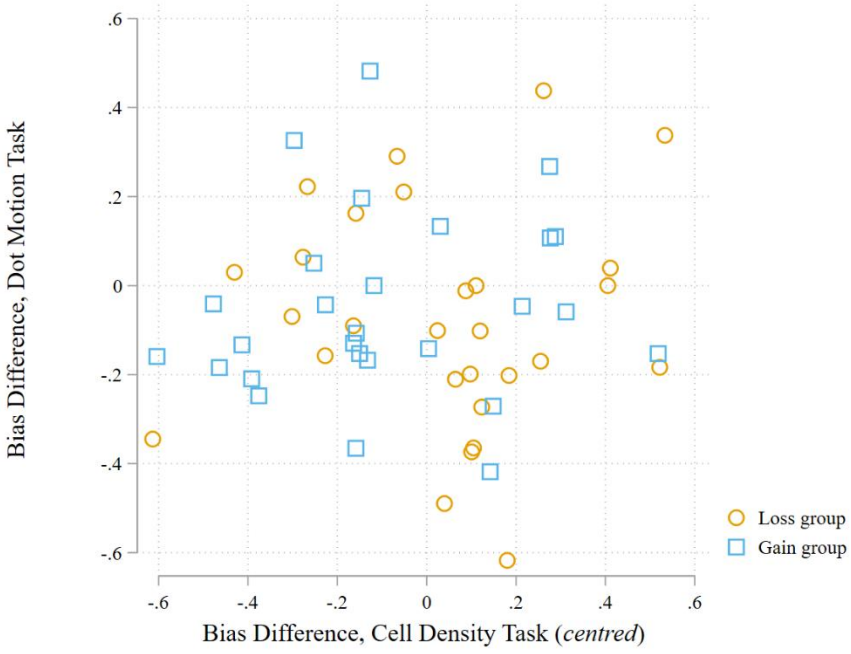


Figure 5 Scatterplot showing no relationship between the two bias difference scores, and hence complete lack of convergent validity. ($r = .06$, $p = .67$, 95% CIs [-.21, .32]).

[7] Using 95% confidence intervals for equivalence testing (testing whether an effect is within certain bounds) yields $\alpha = .025$, because one is technically testing against two values instead of one (an upper bound and a lower bound). For $\alpha = .05$, one should use 90% confidence intervals (Lakens, 2017).

depended on is invalid and must be scrapped. This is where confirmation ends and exploration begins.

The lack of correlation should not be a result of centring as the physical biases do not change drastically across conditions. To make sure, I reran the correlation using the bias difference scores based on the uncentred biases for both tasks (r = .07, 95% CIs [-.20, .33]). Next, I used bias difference scores that were based on centred biases for both tasks (r = .06, 95% CIs [-21, .32]). The lack of correlations is thus not a result of centring the data. For the sake of consistency, I only use the centred scores in the following exploration unless otherwise stated, but the results remain the same had I used the uncentred scores.

Do people become more biased with less sensitivity? While Figure 6 makes it look that way, this is mainly due to a few outliers. Testing for equality of variances yields no significant differences between the easy and impossible conditions (dot motion task: F(54, 54) = 1.16, p = .58; cell density task: F(54, 54) = 1.12, p =.69). Furthermore, the cell density task shows no significant difference in bias between the easy and impossible condition (Cohen's d = -0.06, t(54) = -0.65, p = .52), and the difference in the dot motion task, while significant, is small (Cohen's d = -0.14, t(54) = -2.10, p = .04). The lack of correlation, the equivalent variances, and the lack of substantial change in bias indicate that the bias difference scores only capture noise.

Perhaps the individual bias scores are more informative? As can be seen in the lower half of Table 4, biases in each condition within the same task correlate highly, meaning that people who were more biased in the easy condition were on average more biased in the hard and impossible conditions. Of more interest are the correlations between tasks. The

Table 4

Correlations between centred biases

	1	2	3	4	5	6
Dot Motion Task						
1. Impossible	—	.71***	.74***	-.08	-.15	-.17
2. Hard	.84***	—	.72***	.03	.03	.05
3. Easy	.88***	.83***	—	-.01	-.04	-.10
Cell Density Task						
4. Impossible	.25*	.25*	.28**	—	.52***	.71***
5. Hard	.22	.25*	.27**	.67***	—	.67***
6. Easy	.14	.21	.18	.78***	.74***	—

Note: Correlations below the diagonal are the original correlations (after pre-specified exclusions, N = 55). Above the diagonal are correlations after outliers outside 3 standard deviations were removed (N = 52).
*p<.1 | **p<.05 | ***p<.001

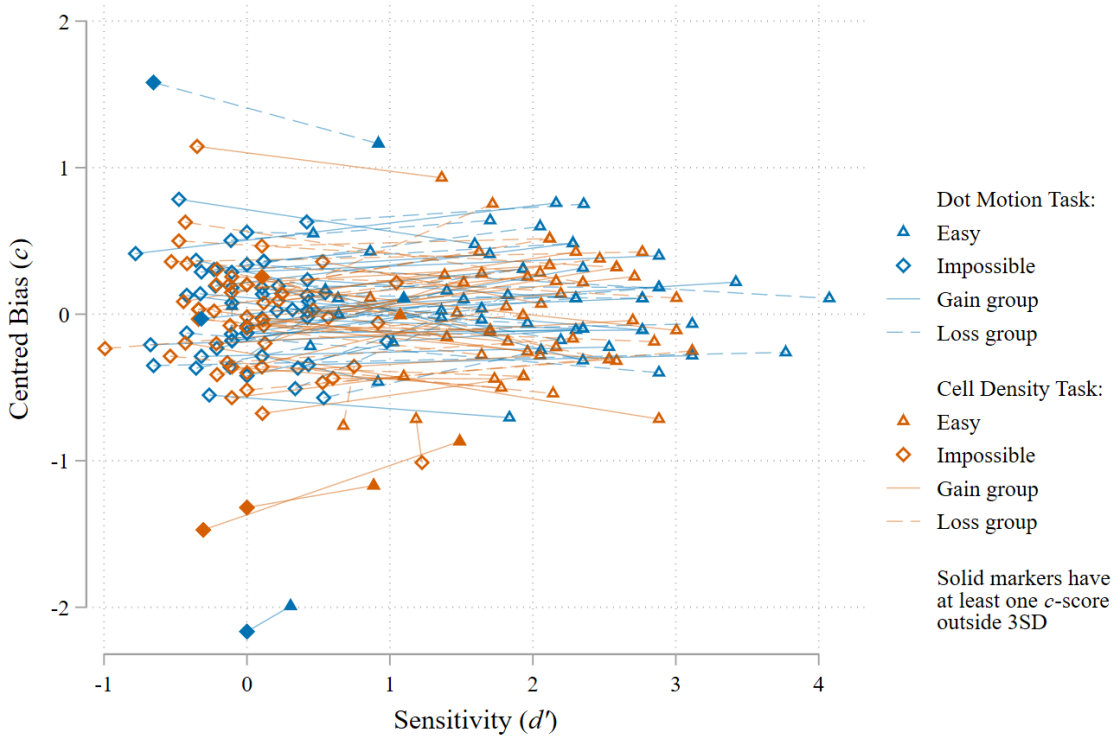


Figure 6 Relationship between sensitivity (*d'*) and bias (*c*). Note that negative *c* indicates liberal bias (i.e., more prone to seeing target alternative).

correlations are still too low to indicate any form of convergent validity, but they appear to have at least some things in common.

To investigate further, I produced scatterplots of the correlations, and noticed that the correlations seem to be heavily influenced by a few outliers (see Figure 7 for an example). If I exclude all outliers outside 3 standard deviations (3 exclusions), the correlation between the impossible conditions drops from $r = .25$ ($p = .06$) to $r = -.08$ ($p = .57$), as do the other correlations (see top half of Table 4). This exclusion criterion was not preregistered, but the excluded participants seem to have given the same response on almost all trials and not attempted to answer accurately. It is therefore safe to conclude that people who were more biased toward their target alternative in one task were *not* more biased in the other task. The tasks did not have things in common after all.

While the study would be underpowered for the main hypotheses, it does have enough power to find the correlations necessary to conclude convergent validity. Neither the bias difference scores nor the biases themselves correlate, meaning they do not measure the same thing. In principle, this can be the result of just one of the measures being faulty. However, it is difficult to imagine what could have gone awry with one of them (such as a failed manipulation) that would not also affect the other. It is therefore more likely that both

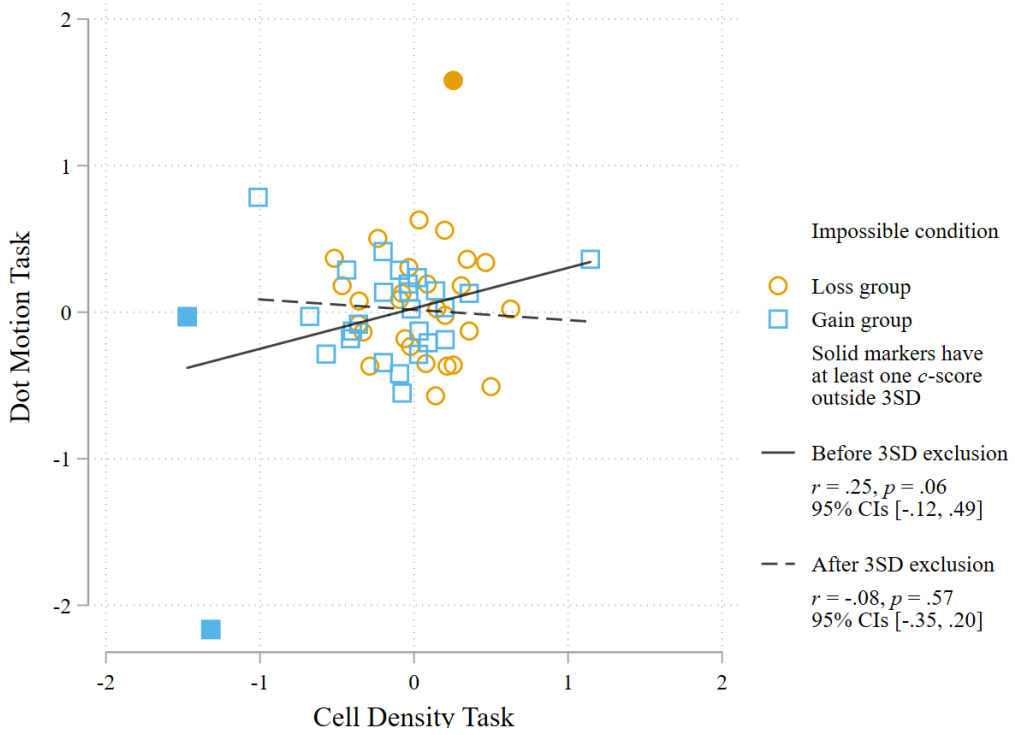


Figure 7 Scatterplot showing relationship between the impossible conditions in the dot motion task and cell density task (N=55). What little correlation there was is due to a few extreme outliers. For comparison, the standard deviation for both tasks is about 0.45.

measures only capture what for our purposes is noise, and that is why they do not correlate. It is therefore reasonable to conclude that the planned analyses for hypothesis 2, 5 and 6 will be uninformative whatever the results. They are therefore omitted from the main text. More manipulation checks and the originally planned analyses were carried out, but they are relegated to Appendix C.

The Learning Task

Figure 8 plots how performance progresses across blocks. There are several things to point out. First, most participants did very well in the positive outcome condition. In fact, the mode of the second block is already full score, with 21.8% of participants scoring perfectly. Participants did so well in fact, that 40% score perfectly in the transfer test phase, the intended measure to be analysed. The ceiling effects are not so bad for the negative outcome condition, where only 16% achieve a full score. In the preregistration, the proportion correct responses in block 6 was offered as an alternative measure if ceiling effects were a problem. Block 6 measures something slightly different (performance on a specific learning task, not how well it transfers to new decisions), so it is not a perfect substitute. However, here the ceiling effects are even worse: 45% of participants score perfectly in the positive condition and 20% score perfectly in the negative condition (with another 36% having only one error).

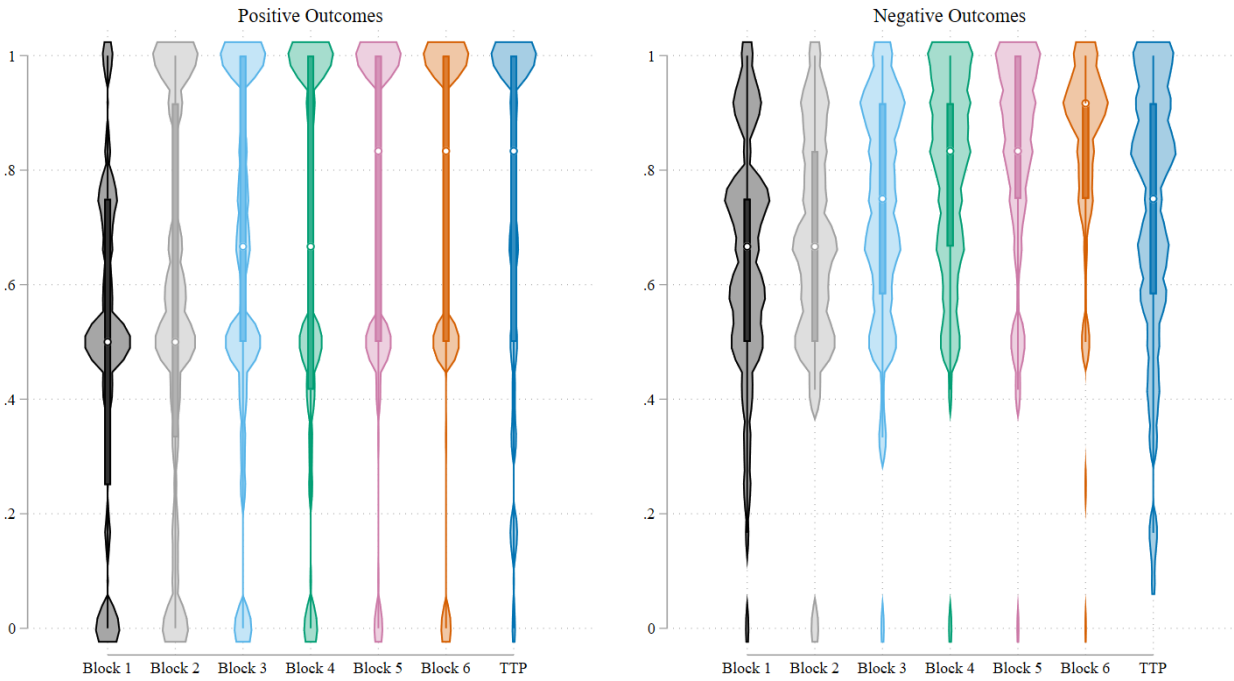


Figure 8 Performance across blocks in the learning task (N=55). The white dots represent the median, the surrounding bar represent the interquartile range, while the protruding lines represent the range of adjacent values. Note that the width of the distributions is determined by frequency compared to that block's mode and not absolute values. TTP = Transfer Test Phase, Y-axis: Proportion correct

We attempted, as mentioned earlier, to make the task more difficult by having fewer trials per block and omitting practice trials, but apparently it was still too easy. Delays in debugging while working in a new programming environment left not enough time to pilot test thoroughly enough to reduce the number of trials accordingly.

The second thing to point out is that the performance in the positive outcome condition seem to diverge into perfect scores and random scores. (The distribution actually seems to be trimodal: those who got it, those who did not get it, and those who *really* did not get it[8]). In fact, 27.3% of participants scored 50% correct in block 6, and another 11% of participants scored 0%. To check whether this had to do with participants' subjective value of money, I divided the participants into those who apparently did not understand the task (scores of >.4 and <.6) and those who did (scores of >.6) and ran a t-test. There was no significant relationship (d = -0.25, $t(46)$ = -0.83, p = .41), but the effect was in the expected direction in that those who understood the task were on average willing to work five minutes more for 100 NOK, and hence presumably valued money more.

[8] This may result from a reluctance to explore. Having previously received negative feedback, the "no change" option (the optimal option in the negative outcome condition but the sub-optimal option in the positive outcome condition) suddenly seem safe. This exemplifies why exclusions should be done with caution, because they may actually be informative about reward sensitivity.

Table 5

Means, standard deviations and correlations for reward sensitivity measures (N = 43)

	M	SD	1	2	3	4	5	6
1. Transfer Test Phase (Positive)	.81	.24	—					
2. Block 6 (Positive)	.83	.22	.81***	—				
3. Saturation Score (Positive)	3.05	2.42	.76***	.85***	—			
4. Transfer Test Phase (Negative)	.73	.21	-.17	-.20	-.27*	—		
5. Block 6 (Negative)	.87	.12	-.01	-.13	-.06	.35**	—	
6. Saturation Score (Negative)	2.33	1.57	-.17	-.05	-.01	.19	.41**	—

*p<.1 | **p<.05 | ***p<.001

In the preregistration, we specified that those who had an average below 60% correct in block 6 were to be excluded. This should include most of those who either did not understand the task or answered randomly nonetheless. Unfortunately, this numbers 12 people, dropping the sample size down to N = 43. This also means that the proportion of perfect scores is even higher.

Participants fulfilling the exclusion criterion might still be informative about reward sensitivity. While it is possible that they simply did not understand the task, it may also result from low sensitivity to rewards. Also, the criteria for the original nine exclusions, while pre-specified that they should be dropped from all analyses[9], have nothing to do with reward sensitivity. Because the sample size is so low, and because excluded participants may be informative nonetheless, I ran all analyses with the four different combinations of exclusion criteria. I report the analyses with the preregistered exclusion criteria in the text (both original and reward sensitivity exclusions [N = 43]), but report if any interesting deviations occur in the other analyses (no exclusions [N = 64], original exclusions only [N = 55], reward sensitivity exclusions only [N = 52]). Correlations and accompanying scatterplots are presented in Appendix D.

Because the original measures of sensitivity to positive outcomes have major ceiling effect problems, they are not very informative. A possible alternative measure is to see how quickly people saturate their scores. I gave each participant a score based on which block they first reached a perfect score (henceforth called the saturation score) and ran the analyses with this measure as well. Participants first reaching a perfect score in block 2 were given 6 points, first reaching a perfect score in block 3 meant 5 points, et cetera. If they scored

[9] In hindsight, specifying that they should be excluded from all analyses was a mistake. Truth be told, I did not expect that many to fail the exclusion tests, nor that many to fall below the accuracy thresholds. I chose to keep to the pre-specified criteria in the text and instead report interesting deviations, rather than argue for why the exclusion criteria did not matter after all.

between .60 and .99 in block 6, they were given 1 point, and the rest were given 0 points. I did the same for sensitivity to negative rewards, but ironically, this measure has minor floor effects because fewer people reached a perfect score. It may therefore be invalid, but I kept it for consistency's sake.

Summary statistics and intercorrelations between the different measures of reward sensitivity are presented in Table 5. For negative feedback, the correlation between block 6 and the transfer test phase is not high enough to suggest interchangeability. This is expected if they measure slightly different things. What is more surprising is the high correlations within the positive condition, but that may reflect ceiling effects rather than convergent validity.

The hypotheses were that positive emotional reactivity would predict sensitivity to positive outcomes and negative emotional reactivity would predict sensitivity to negative outcomes. They were tested with simple regressions as specified in the pre-registration. Results are displayed in Table 6.

Neither hypothesis gained direct empirical support from the analyses. All confidence intervals include β-values of either -.30 or .30. These are certainly bigger than the smallest effect size of interest, particularly because 75% of effects in individual difference research are lower (Gignac & Szodorai, 2016). The results are therefore inconclusive, because (1) the study is not powered enough to detect smaller effects if they exist, and (2) the study is not powered enough to distinguish true null effects from the smallest effect size of interest.

If we only look at effect sizes, we see that the ceiling effect-stricken positive transfer test phase shows no correlation at all. The impromptu saturation score on the other hand show a small effect in the hypothesised direction, but because of the low power, there is a 50% chance that we would find a similar or bigger effect if, in reality, there was no relationship (and an even bigger chance that this would happen in at least one of the tests). The relationship between negative reward sensitivity as measured by the transfer test phase and negative reactivity is of similar magnitude in the predicted direction, but the relationship is reversed when block 6 is used as the outcome variable. It is worth noting that most of the analyses in Table 6 do not have normally distributed errors, but judging by the scatterplots in Figure D1 (in Appendix D), no patterns of interest are missed.

Rerunning the analyses across the multiple exclusion criteria yields for the most part similar results. Ceiling effects have the unfortunate effect of giving non-perfect scores more weight (just like outliers). In fact, the positive transfer test phase correlated with positive reactivity as predicted before participants were excluded (β = .18, p = .19, 95% CIs [-.09, .45]), a non-significant medium effect, but the relationship disappears entirely when we

Table 6

Regressions between reward sensitivity and emotional reactivity (N=43)

				95% CIs for β		
	B	*SE B*	*p*	lower	β	upper
Transfer Test Phase (Positive)						
Positive Reactivity	-.006	.069	.93	-.33	-.01	.30
Block 6 (Positive)						
Positive Reactivity	.035	.065	.60	-.23	.08	.40
Saturation Score (Positive)						
Positive Reactivity	0.479	0.708	.50	-.21	.11	.42
Transfer Test Phase (Negative)						
Negative Reactivity	.040	.039	.31	-.15	.16	.47
Block 6 (Negative)						
Negative Reactivity	-.015	.022	.49	-.42	-.11	.21
Saturation Score (Negative)						
Negative Reactivity	-0.133	0.299	.66	-.38	-.07	.25

apply the specified exclusion criteria. Because we cannot discern whether their low scores are due to lack of understanding or lack of reward sensitivity, the results remain inconclusive.

The only thing of interest after exploring the relationship between reward sensitivity and emotional reactivity further is a negative relationship between the negative transfer test phase and *positive* reactivity (β = -.27, p = .05, 95% CIs [-.60, -.00]). This indicates that the more easily people react with positive affect, the less sensitive they are to negative feedback. Considering the low power and multiple comparisons being made, it is difficult to say whether this is a spurious effect or not. Also, while this finding remains significant across the different exclusion criteria, it disappears entirely if one uses block 6 as the outcome variable instead (β = .09, p = .56, 95% CIs [-.22, .40]). Whether this is because the effect is spurious, or whether it reflects genuine differences between the reward sensitivity measures is hard to say. Rerunning the analyses with sex, age, and the subjective value of money as covariates yielded nothing of interest (other than nudging the last effect to barely non-significant), and there were no interaction effects.

Discussion

The most important finding is that the two perceptual tasks showed no sign of convergent validity. There was no relationship between biases or bias differences in the two perceptual discrimination tasks. Even if arousability predicted bias in one task, it would not predict bias in the other task, hence making it impossible to conclude whether the hypotheses were supported. It is difficult to tell whether the manipulation failed because of theoretical or methodological issues, so I will discuss both possibilities starting with the latter.

Methodological Issues

First, the manipulation could have failed because the stakes were too low to incite adequate desire. A study that investigated real versus hypothetical rewards in temporal discounting found that real monetary rewards appeared to yield smaller effects, but that this was because real rewards were usually smaller compared to hypothetical rewards (M. W. Johnson & Bickel, 2002). Likewise, because budgetary concerns limit the size of the stake, it may not have been big enough for the desire to be sufficiently arousing. One way to check whether the stake induced arousal would have been an independent measure of arousal, such as skin conductance or heart rate, combined with a control group with no stake. Furthermore, we could then compare the physiological data with the ERIPS to see whether more arousable people indeed are more arousable. However, physiological measures are not flawless, many of them have validity concerns unless handled properly (e.g., Quintana & Heathers, 2014), and the equipment necessary would make the study both more expensive and make participation more tedious. This could be accomplished, but it would require a complete redesign of the study.

Second, the stake may have failed to induce arousal because it was not made salient enough by poor instructions. Poor instructions may also have added unnecessary amounts of noise to the data in general. Numerous people performed as if they answered randomly, both in the learning task and perceptual discrimination tasks, and many also failed to remember their target alternative, presumably an easy task. In total, 21 people – a third of all participants – met one or more of the pre-specified exclusion criteria. For example, in the positive learning task trials, half of those that did not score perfectly in block 6 had a score of 50%, which is what they would have gotten by chance alone. This is not just poor performance. It either indicates that participants wilfully ignored the instructions or that they did not understand them. Because it happened so frequently, the latter cause seems more likely. This is corroborated by anecdotal evidence: Some participants reported after the

experiment that they persisted in choosing the images they liked despite feedback as if it was a Rorschach test, others mentioned that they were not that interested in the monetary rewards. The subjective value of money did not predict who answered randomly.

The original instructions for the learning task were obtained via personal correspondence with Gold et al. (2012) and translated to Norwegian. One point that may have caused confusion is that participants were instructed that there were no absolute right or wrong answers. I fail to see how a more direct instruction to end up with as much money as possible would not yield equally valid measures. Those more sensitive to losses would presumably still focus more on avoiding losses than someone less sensitive to losses. The instructions in the perceptual discrimination tasks were our own, and hence any flaws with them are on our shoulders.

Third, much of the experiment, particularly the perceptual discrimination tasks, are highly decontextualized, which may have dampened the effect of the stake. While decontextualization allows for precise experimental control, it could also make the tasks less engaging and harder to understand for participants, particularly if the instructions are complicated and poorly communicated. Ecological validity may also be an issue. While the same cognitive mechanisms that operate "in the wild" should also operate in laboratory settings, the context-dependence of those mechanisms might complicate matters. For example, Cosmides (1989) famously demonstrated how performance on a simple logical problem dramatically improved when presented in an ecologically relevant context.

The solution to many of the potential issues outlined above would be to put more work into making the experiment more participant-friendly. This would involve more pilot testing with particular focus on instruction and comprehension. Increasing the stake may not be feasible, but it could be made more salient with, for example, pictures of money. More pilot testing would also allow finetuning the learning task difficulty and thus avoid ceiling effects. Gold et al. (2012) did not mention ceiling effects in their original study, so it is unclear whether they had similar issues. One plausible explanation is that the current participant population, students, may on average be more intelligent than the general population, and hence learn more quickly. Further pilot testing could have fine-tuned the difficulty, either by manipulating the number of trials, or by manipulating the probability of receiving the correct feedback.

One could also reframe this study *as* a pilot study. While the study was underpowered to test the main research questions, the study was not so for testing convergent validity. It is unlikely that the low correlation resulted from chance. Hence, the short data gathering period

and the resulting low sample size may actually have saved money and resources that would otherwise have been spent on gathering invalid data from even more participants.

A final methodological issue is that the experiment attempts to do too many things at once. In addition to looking for and correlating individual differences in the stake-likelihood effect with arousal, it also attempts to use a novel measure. Furthermore, the novel measure attempts to isolate the effect from later justificatory processes by looking at lower-level perceptual processes. In so doing, the study has inadvertently demonstrated the Duhem-Quine problem of falsification in that there is no logical way to tell what specifically has been falsified (Gershman, 2019; Kashyap & Sirola, 2018). Conceptual replications, as opposed to direct replications, are often criticised for changing too many parameters at once, precisely because null results would be difficult to interpret (Chambers, 2017). One counter-argument is that effects so fragile that they do not withstand changes to the experimental procedure are not of practical significance and hence not interesting. On the other hand, going from explicit likelihood judgements to perception may be too big of a leap, even though both fundamentally depend on expectations.

The analysis identified two major problems. The first and most important finding is the lack of convergent validity. A future study would be advised to, for example, use Vosgerau's (2010) original measures – explicit likelihood judgements – and correlate them with arousability. While such explicit measures have limits as discussed in the introduction, they are certainly more valid than the measure attempted here. The second problem is the dual issue of ceiling effects and misunderstandings in the learning task. Many participants either understood the task too well, or not at all. A future study should consider simplifying the original instructions, and finetune the difficulty level by changing the number of trials and trials per blocks. It is also possible that spreading fewer trials over more blocks made the task easier, as distributed practice seem to enhance learning in other studies (e.g., Karpicke & Roediger III, 2007). Future studies must take this into account.

Theoretical Issues

If we assume that the null results partly stem from desires having limited effect on perception, then there are three theoretical issues that need mentioning. The first is that the stake-likelihood effect may originally be a spurious finding. However, neither the current methods nor the resulting data are suited to address this, so I will not discuss this further.

The second issue is that the stake-likelihood effect could be limited to higher-level cognitive processes, and hence not affect perception. This is at odds with the continuity of

cognition and perception (Hohwy, 2017). Also, perception is readily influenced by expectations. For example, the "light-from-above" prior, which influences how shading affects perception of shape[10], can be changed with experience (Adams, Graf, & Ernst, 2004). Expectations can also be readily influenced by desires and motivations, evidence of which was discussed in the introduction. The influence of motivation and desires on perception should then follow. One possibility is that lower-level perceptual processes are more constrained by perceptual data and therefore more difficult to manipulate via top-down processes. Balcetis and Dunning (2006) successfully manipulated perception with desirability, but their perceptual stimuli were ambiguous in that they could be interpreted in two specific ways. This study, on the other hand, used noisy stimulus that were difficult to interpret either way. This would imply that, while an effect in principle could be found, it is more difficult to manipulate in lower-level processes than in higher-level processes. If this is the case, then methodological issues would only exacerbate the problem.

It could also be the case that arousals' supposed effect on memory search primarily occurs in higher-level processes, and thus have limited influence on lower-level processes. This relates to the third issue, which is that the mechanism may involve higher-level justificatory processes, precisely the ones we tried to isolate the effect from. If so, then the stake-likelihood effect may not reflect true prior expectations, but more be the result of processes producing justifiable responses (Mercier & Sperber, 2011, 2017). Related to this is the above critique of how the tasks may have been too decontextualized. Cosmides (1989) demonstrated how simple logical problems are solved more easily, not when they are presented in a more ecologically likely way, but in an ecologically *relevant* way. This shows that cognitive mechanisms are highly context dependent. If so, then the mechanisms involved in wishful thinking may only be triggered in specific contexts and may therefore be elusive in laboratory settings where those triggers are absent. A better approach may then be to focus on what specific function or functions wishful thinking might have, and from there hypothesise what should trigger the mechanisms. This assumes that wishful thinking is functional. On the other hand, if wishful thinking is cognition gone awry, asking what function or functions have gone awry might give the same benefit. (See also Appendix A)

[10] One fun example is to look up a satellite photo of a section of the Grand Canyon. If you rotate the photo so that the shadows are on the top, it looks like a deep canyon, while if you turn the picture around so that the shadows are at the bottom, it looks like an impressive mountain range.

Reward Sensitivity and Arousability

There were some methodological issues with the learning task as well, discussed above, but I was still able to test the hypotheses. Neither of the two hypotheses were supported by the analyses. However, the study was underpowered, and the methodological issues may have interfered. The analyses are therefore inconclusive. We can note, however, that if there is a relationship between arousability and reward sensitivity, it is likely to be small (provided the null findings are not exclusively attributable to methodological issues). Future theoretical work could take this into account.

Conclusion

The study is unable to address the original research questions properly due to an invalid measure and low power. It is therefore inconclusive. The measures of reward sensitivity gave more information, but the low power makes it difficult to discern interesting but non-significant effects from absent effects. The study highlights the importance of validity checks and pre-planning analyses. Without them, the analyses would be run uncritically, positive findings would be searched for wherever they may hide, and it would be impossible to tell whether null findings resulted from falsification or lack of validity (see also Appendix C). The study also highlights the importance of sufficient statistical power.

Unpredicted null results that cannot be attributed to chance are by necessity due to either errors in the theoretical reasoning or poor design or execution of methods. While it is easy to feel dejected by this realisation, one must remind oneself that the reason for investigating things empirically is because human reasoning is flawed. After all, if humans could reason perfectly, there would be no need to check the conclusions against reality in the first place. Proper scientific practice such as validity checks and pre-planned analyses prevented this study from demonstrating wishful thinking, not only in its results, but also in its interpretation.

List of Appendices

Appendix A: A Primer on Different Levels of Analysis

The present study attempted to examine some of the algorithmic underpinnings of motivated reasoning. More specifically, I investigated whether trait arousability is a relevant parameter in the motivated perception of ambiguous stimuli. Whenever discussing how the mind works, one must keep in mind the distinction between different levels of analysis. As with much of psychology, much research has investigated the mechanisms of motivated reasoning without giving proper thought to its function or, more importantly, without distinguishing properly between mechanism and function (case in point being the strategic pessimism and positive illusions mentioned in the introduction). This has arguably left the field in a bit of a mess (Muthukrishna & Henrich, 2019; Tooby & Cosmides, 1992). The research on reason – and psychology in general – needs clarity on this point. I therefore added this seemingly superfluous primer, as well as a short discussion on why it matters.

Ultimate cause (function), ontogenetic cause (development), and proximate cause (mechanism) must not be confused (Bateson & Laland, 2013; Tinbergen, 1963). The function is what something is for, just as a calculator is *for* calculating answers to mathematical problems. The function of something is important, in part, because it informs how we should expect the proximate mechanism of something to work. The proximate mechanical level can be further subdivided into algorithm and physical implementation (Marr, 1982). The former is the abstract rules that, in this case, a calculator executes to produce the outcome, while the latter is the actual mechanical workings of the circuit board. That is, how the different parts physically interact to produce the outcome. In psychology, the algorithmic level roughly describes cognitive science, while the implementational level describes neural science. A biological function can be solved by many different algorithms, and the same algorithm can be implemented physically in different ways, which in turn can be arrived at via many different developmental pathways. Confusing these levels leads to bad behavioural science.

Take Taylor and Brown's (1988) idea of positive illusions as an example. The general idea, which has become popular even outside psychological circles, is that people try to maintain a positive self-image even if it takes illusions to improve their mental health (usually conceptualised as self-esteem, cf. Kurzban, 2011). With the primer in mind, we see that this explanation confuses the mechanical role of emotions for its ultimate function. The explanation is equivalent to saying that people eat to stave off hunger: true, one might even say interesting, but not a very good explanation by itself. One would still need to ask why people feel hungry in the first place, or why a negative (or accurate) self-image results in lower self-esteem and why high self-esteem should be preferable. Such questions require

good, ultimate explanations (e.g., people need food to survive and reproduce). Sure, one could continue asking "why" to these explanation as well, but instead of taking you on the long and winding road of folk psychological concepts and intuitions, good ultimate explanations will quickly take you to the first principles of science (e.g., the need for survival and reproduction arises from the process of natural selection, which in turn arises because of the properties of a complex type of molecule, which in turn arises because of the physical characteristics of the universe, etc.,). This obviously does not mean that we should stop investigating proximate mechanisms such as positive illusions' effects on mental health or how hunger affects behaviour. After all, this study is primarily focused on a proximate mechanism. It means that these proximate explanations must both be properly distinguished from and preferably related to ultimate explanations.

Appendix B: The Reward Sensitivity Measure

The learning task consists of two parts: An acquisition phase where participants learn which picture to press in four different pairs, and a transfer test phase where participants are presented with the same pictures in novel pairs. The transfer test phase was originally designed to do two jobs: (1) Distinguish model-free stimulus-response learning from model-based stimulus-response-outcome learning, and (2), within model-based learning, distinguish the effects of positive outcomes from those of negative outcomes (Gold et al., 2012). To distinguish sensitivity to positive and negative outcomes, the original study relied mostly on one specific pairing (cell 7A / 1G in Figure B1 below). In the present study, we are interested in individual differences in the representation of expected utility and must therefore find a more fine-grained way to measure sensitivity to negative outcomes and positive outcomes. We can do this by using all the pairs that distinguish sensitivity to negative and positive outcomes.

Figure B1 shows a matrix that represents all possible pairings of the eight images. The original acquisition phase parings are in bold. Each cell shows the relative expected utility of selecting the picture represented by the column when paired with the picture represented by the row. For example, cell 2A (one of the original pairings) shows the expect utility to be +80 if one chooses the picture that is rewarding 90% of the time over the picture that is rewarding 10% of the time. This matrix assumes equal sensitivity to positive and negative outcomes.

Model-based, equal sensitivity

			A	B	C	D	E	F	G	H
		Condition	+90		+80		-80		-90	
	Condition	Rewarded?	90	10	80	20	-20	-80	-10	-90
1	+90	90		**-80**	-10	-70	-110	-170	-100	-180
2		10	**80**		70	10	-30	-90	-20	-100
3	+80	80	10	-70		**-60**	-100	-160	-90	-170
4		20	70	-10	**60**		-40	-100	-30	-110
5	-80	-20	110	30	100	40		**-60**	10	-70
6		-80	170	90	160	100	**60**		70	-10
7	-90	-10	100	20	90	30	-10	-70		**-80**
8		-90	180	100	170	110	70	10	**80**	

Figure B1: Expected utilities of selecting the picture represented by the columns. The pairs are color-coded so that green means positive utility, yellow indicates indifference, and red indicates negative utility. Cells in bold are the original acquisition phase pairs.

Figure B2 displays the same matrix twice, only here the relative expected utility assumes the actor to only be sensitive to positive outcomes in the top matrix, and only sensitive to negative outcomes in the bottom matrix. For some cells, the expected utilities are equal in both matrices. These are not informative for distinguishing sensitivity to positive and negative outcomes and was not shown to participants. Other cells, however, have different expected utilities, and are thus informative. One example is cell 7A. An actor only sensitive

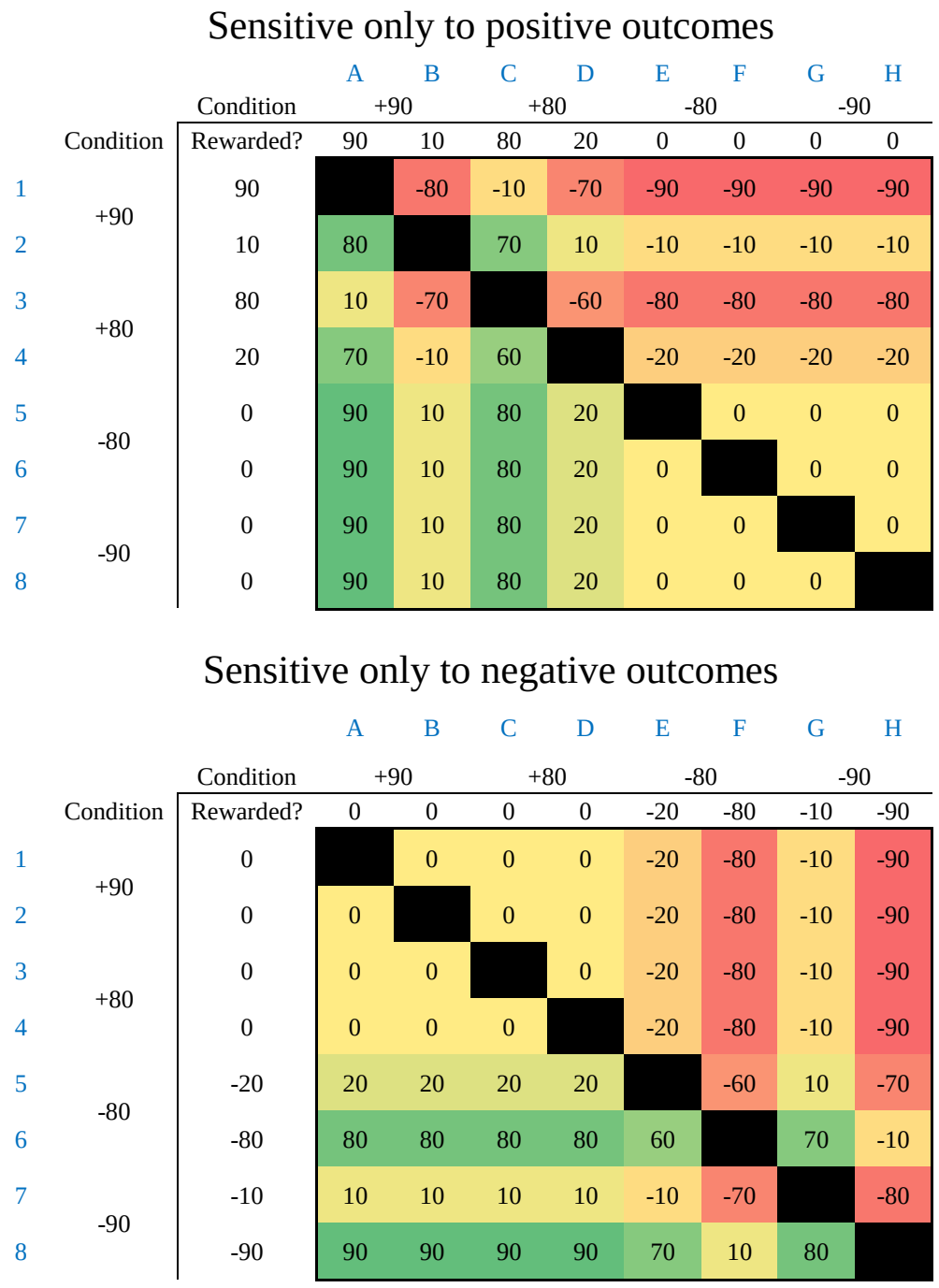


Figure B2: Expected utilities of selecting the picture represented by the columns for actors only sensitive to positive outcomes (top) and actors only sensitive to negative outcomes (bottom).

to positive outcomes should reliably choose the picture that is rewarding 90% of the time over the picture that avoids loss 90% of the time. An actor only sensitivity to negative outcomes, on the other hand, should be close to indifferent. Other cells, such as 7F, have the opposite prediction, where actors only sensitive to positive outcomes should be indifferent and actors only sensitive to negative outcomes should choose reliably.

To measure sensitivity to positive outcomes, then, we can sum the number of optimal responses in pairs where actors only sensitivity to negative outcomes should be indifferent, while to measure sensitivity to negative outcomes, we can sum the number of optimal responses in pairs where actors only sensitive to positive outcomes should be indifferent. Figure B3 shows which pairs are informative about positive and negative reward sensitivity. The blue cells are informative about sensitivity to positive outcomes, and the yellow cells are informative about sensitivity to negative outcomes. Because there are 6 pairs for each sensitivity and each pair is shown twice, each measure has a top score of 12 where 6 indicates indifference. The scores will be transformed to proportions before analysis.

This approach to analysis assumes that model-free learning is not fast enough to hide the effect of model-based learning. This is a reasonable assumption because Gold et al. (2012) managed to discriminate between the effects of two kinds of learning.

Pairs distinguishing positive and negative sensitivity

			A	B	C	D	E	F	G	H
		Condition	+90		+80		-80		-90	
	Condition	Rewarded?	90	10	80	20	-20	-80	-10	-90
1	+90	90				-70	-70		-80	
2		10			70			70		80
3	+80	80		-70			-60		-70	
4		20	70					60		70
5	-80	-20	70		60					70
6		-80		-70		-60			-70	
7	-90	-10	80		70			70		
8		-90		-80		-70	-70			

Figure B3: Blue cells are informative about sensitivity to positive outcomes and yellow cells are informative about sensitivity to negative outcomes. The values in each cell show the difference in relative expected utility between actors only sensitive to positive outcomes and actors only sensitive to negative outcomes.

Appendix C: What Might Have Been

This appendix serves two purposes: show the results from the planned analyses and further data exploration, and illustrate how low power and analytical flexibility can make anything significant (see also Simmons, Nelson, & Simonsohn, 2011). To appreciate the flexibility on offer, consider the number of possible ways to test the main hypothesis. First of all, there are multiple candidate outcome variables. In addition to the planned composite bias difference score, there are the bias difference scores from either task, as well as the six different bias scores. There are also different ways to measure and calculate bias, such as *c'* and $log(\beta)$, in addition to the standard measure of *c*. So far, we are up to 9 x 3 = 27 possible analyses.

The exclusion criteria considered in the thesis alone – pre-planned exclusions, 3SD exclusion, and no exclusions – brings the number of possible analyses up to 81, but an indeterminate number of other exclusion criteria are possible too. The possibilities gets further multiplied by considering (1) the order of exclusion, which affects the variance and hence standard deviations; (2) the use of covariates, such as sex, age, and the subjective value of money; (3) transformations and analytical choices if statistical assumptions are breached, et cetera. Furthermore, one could use different ERIPS factors as predictors. One could for example easily have argued that intensity is the relevant factor (and it may actually be, but that was not what we thought prior data collection). Reward sensitivity also offer similar amounts of analytic flexibility.

Not all these analyses are equally justifiable, and many will yield almost identical results. Nevertheless, there are more degrees of freedom than there are participants in the experiment: Some analyses will surely give "interesting" results. Note also that correcting for multiple comparisons becomes less straight-forward because there is an indeterminate number of possible comparisons. The normal correction – the Bonferroni correction – sets the α-level at .05/X, where X is the number of comparisons. This assumes that X is known, which works fine for a correlation table, but less so for general exploration. Instead, a good understanding of how extreme p-values become more likely with multiple comparisons must be combined with general caution in interpreting results.

My original plan was to examine multiple possible analyses and present a curated subset. However, the pre-planned analyses had interesting enough results, so I will only present those. Unless otherwise stated, the analyses use the centred biases with the pre-planned exclusions (N = 55). The data set and analysis script are available at https://osf.io/vq93j/.

Table C1

Correlations for c' *biases*

	1	2	3	4	5	6
Dot Motion Task						
1. Impossible	—	-.05	.13	-.04	-.03	.25*
2. Hard	.79***	—	.75***	-.13	.24*	.11
3. Easy	.83***	.97***	—	-.10	.01	.09
Cell Density Task						
4. Impossible	.53***	.57***	.58***	—	-.20	-.21
5. Hard	.64***	.76***	.73***	.39**	—	.67***
6. Easy	.57***	.53***	.58***	.25*	.79***	—

Note: Correlations below the diagonal are the original correlations (after pre-specified exclusions, N = 55). Above the diagonal are correlations after observations with at least one *c'* outside 3 standard deviations have been removed (N = 53).

*p<.1 | **p<.05 | ***p<.001

More Manipulation Checks

Perhaps the lack of convergent validity resulted from using the wrong bias measure? Unlike *c*, other measures are not mathematically independent of *d'*, but they may still give interesting results. One common alternative is *c'*, which is *c* divided by *d'* (Macmillan & Creelman, 1991). This obviously does not work if *d'* = 0, but by assuming *d'* = 0.1 for the relevant cases, a new bias measure could still be generated. Interestingly, this measure shows a relatively high degree of correlation between the bias difference scores (r = .42, p = .001, 95% CI [.18, .62]). What is more, the correlations between the bias scores are even higher, ranging from r = .53 to r = 76, as can be seen below the diagonal in Table C1. Could it be that the hypothesised effect required a different way to calculate bias, and that the measure shows some degree of convergent validity after all? The scatterplot in Figure C1 reveals the true cause of the correlations: dividing by sensitivity only increase the influence of outliers (especially because outliers tended to have low sensitivity). Again, removing the outlier completely removes the effect for the bias difference scores (r = -.05, p = .74, 95% CI [-.31, .23]). The bias scores too no longer correlate after removing outliers outside 3 standard deviations, as can be seen above the diagonal in Table C1. The way bias is calculated is therefore not at fault for the lack of convergent validity.

Hypothesis 2, 5, and 6 was deemed impossible to test because the dependent variable lacked convergent validity. This assumes that lack of convergent validity resulted from both bias difference measures being invalid. There is no good reason to expect only one of them to be faulty, but for the sake of argument, let us assume so. One way to check whether the

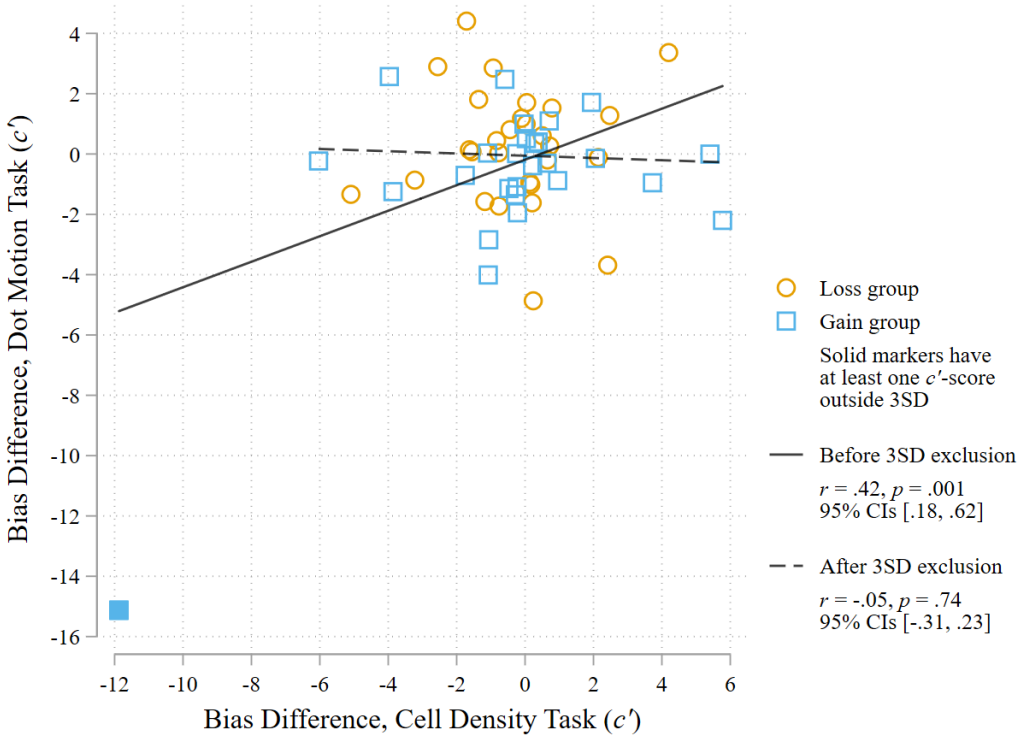


Figure C1 The correlation between bias difference scores based on *c'* is entirely due to one outlier.

manipulation worked would be to check for a difference in physical bias across target alternatives. As can be seen in Table C2, no condition in either task shows any significant change in physical bias across the target alternatives (all *p*s > .22), meaning those primed to see, say, black, where just as biased toward seeing black as those primed to see white. The difference is, incidentally, significant in the impossible cell density condition prior to any exclusions (*MD* = -0.26, *p* = .025). However, the overall pattern of differences suggests it likely is a false positive, especially because it only shows up before the pre-planned exclusions of those who did not remember their target alternative.

Table C2

Independent t-tests for whether physical bias differs between target alternatives

	physical c_{down}		physical c_{up}				
Dot Motion Task	M	SD	M	SD	*MD*	t(53)	*p*
Impossible	-0.13	0.60	-0.12	0.31	-0.01	-0.10	.92
Hard	-0.13	0.52	-0.12	0.34	-0.02	-0.14	.89
Easy	-0.18	0.55	-0.04	0.30	-0.14	-1.18	.24
	physical c_{black}		physical c_{white}				
Cell Density Task	M	SD	M	SD	*MD*	t(53)	*p*
Impossible	-0.34	0.36	-0.49	0.51	0.15	1.23	.22
Hard	-0.29	0.39	-0.35	0.45	0.06	0.54	.59
Easy	-0.38	0.40	-0.48	0.44	0.10	0.85	.40

Table C3

Independent t-tests for whether bias differ by desirability

	c_{loss}		c_{gain}				
Dot Motion Task	M	SD	M	SD	*MD*	t(53)	*p*
Impossible	0.08	0.44	-0.07	0.51	0.15	1.17	.25
Hard	0.08	0.41	-0.07	0.46	0.15	1.27	.21
Easy	0.16	0.34	-0.02	0.51	0.17	1.47	.15
	c_{loss}		c_{gain}				
Cell Density Task	M	SD	M	SD	*MD*	t(53)	*p*
Impossible	0.05	0.30	-0.20	0.52	0.25	2.18	.03
Hard	0.06	0.30	-0.13	0.49	0.19	1.73	.09
Easy	0.01	0.36	-0.11	0.46	0.12	1.08	.29

Perhaps participants were more influenced by the desirability rather than the focal outcome? This would be evidence for a general desirability bias rather than a stake-likelihood effect. As can be seen in Table C3, the mean differences are all in the correct direction to suggest a desirability bias (gain conditions have more liberal bias), and it is significant in the impossible cell density condition (*MD* = 0.25, *t*(53) = 2.18, *p* = .03). What is more, this does not appear to be the result of outliers, as can be seen by the distribution in Figure C2. While excluding the outlier outside 3 standard deviations nudges the difference to non-significant (*MD* = 0.20, *t*(52), *p* = .06), excluding outliers outside 2 standard deviations nudges it back again (*MD* = 0.17, *t*(49) = 2.10, *p* = .04).

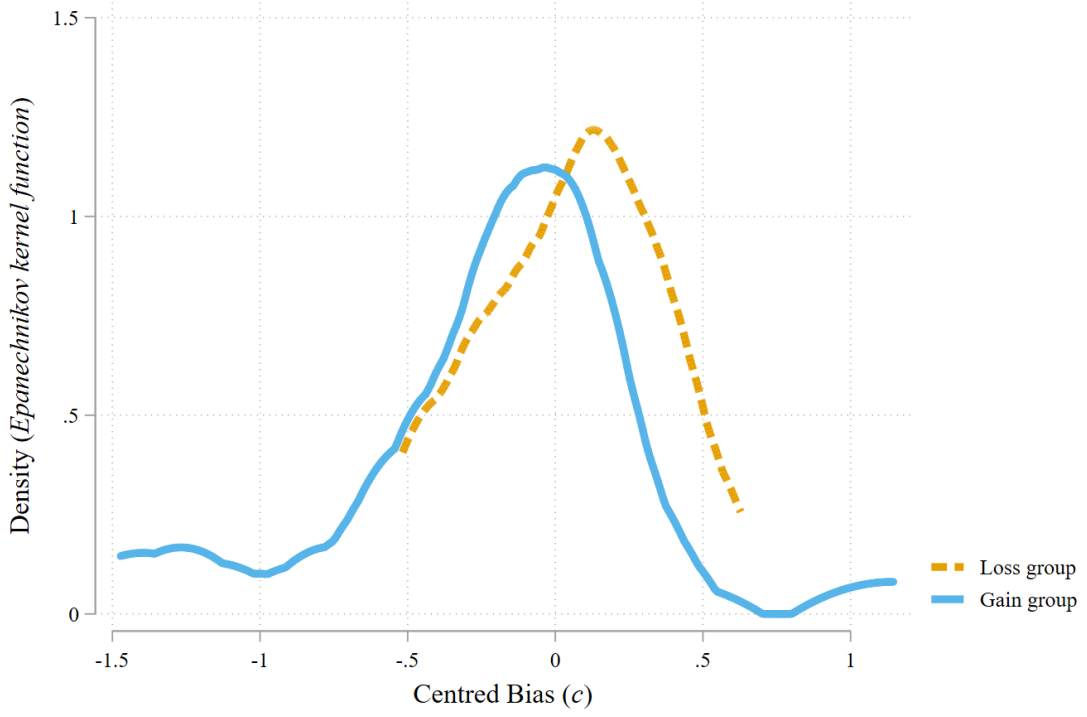


Figure C2 The distribution of biases in the impossible cell density condition shows that the difference in bias is not solely due to a few outliers. The same pattern is also evident in Figure 7.

Table C4

Moderated regression analysis of emotional reactivity and composite bias difference score (N = 55)

	B	*SE*	*t*	*p*	*β*
Main Effects[a]:					
1. Positive Reactivity	0.005	0.05	0.10	.92	0.01
2. Negative Reactivity	0.030	0.03	0.92	.36	0.13
3. Gain/Loss condition[b]	-0.049	0.05	-0.96	.34	-0.13
Constant	*-0.127*				
Interaction Effects[c]					
1. Positive Reactivity	-0.005	0.33	-0.01	.99	-0.01
2. Negative Reactivity	0.108	0.33	0.33	.75	0.46
1 x 2	-0.028	0.10	-0.29	.78	-0.42
3. Gain/Loss condition	1.962	1.50	1.31	.20	5.28
3 x 1	-0.604	0.43	-1.39	.17	-5.63
3 x 2	-0.863	0.45	1.90	.06	-7.15
3 x 1 x 2	0.260	0.13	1.96	.06	7.37
Constant	*-0.052*				

a: Main Effects: R^2 = .04, $F(3, 51) = 0.64$, p = .59
b: Loss condition = 0, Gain condition = 1
c: Interaction Effects: R^2 = .22, $F(7, 47) = 1.85$, p = .10
Note: Variables are *not* centred

The mean differences in the dot motion task are smaller but in the right direction. However, if this reflected genuine manipulation in both tasks, they should correlate, and they do not. It is therefore impossible to tell whether the effect is spurious or not.

The Pre-Planned Analyses

What would have happened if the planned analyses were run anyway? The plan was to use an average of the bias difference scores as the dependent variable (*M* = -0.04, *SD* = 0.19, *Min* = -0.47, *Max* = 0.45). Negative scores mean a more liberal bias in the impossible condition than in the easy condition, which is how I defined motivated perception in the introduction. In other words, a more negative score means more motivated perception.

Hypothesis 2. Hypothesis 2 – the main hypothesis – was that positive and negative reactivity should predict motivated perception. More specifically, (1) more positive reactivity should predict more motivated perception in the gain condition, (2) more negative reactivity should predict more motivated perception in the loss condition, and (3) people in the loss condition should on average show more motivated perception than those in the gain group. The original hypotheses stated that the effects should be independent of each other.

Hypothesis 2 was to be tested with a moderated regression analysis with a three-way interaction. The results are displayed in Table C4. There were no significant main effects, but some of the interaction effects are interesting. There was no significant interaction between positive and negative reactivity in the loss condition, but the third-order interaction with gain/loss condition is almost significant ($B = 0.26$, $SE = 0.13$, $p = .06$). Through linear combination, we find that the interaction between positive and negative reactivity is significant in the gain condition ($B = 0.232$, $SE = 0.09$, $p = .01$). This means that the effect of positive reactivity depends on the score of negative reactivity, and vice versa, in the gain condition. Some of the other interaction terms are also trending and therefore worth taking into account.

Figure C3 shows visual representations of the interaction effects. It is apparent that the effect is generally small in the loss condition, but that an interesting pattern emerges in the gain condition. We also see why there was no main effect: High and low scores have effects in the opposite directions that cancel each other out when averaged. In the gain condition, people who scored about the same on positive and negative reactivity showed on average no sign of motivated perception, unless their scores were extreme, in which case they had a more conservative bias in the impossible condition – the opposite of motivated perception. People who scored differently on positive and negative reactivity – that is, either high positive/low negative or low positive/high negative –showed on average more motivated perception. The effect of positive reactivity is significant ($p < .05$) for negative reactivity scores above 3.2, while the effect of negative reactivity is significant for positive reactivity scores below 2.3 and above 2.8.

The hypothesised effect in the gain condition was that positive reactivity would predict more motivated perception. This effect was found, but only for people scoring low on negative reactivity. For people scoring high on negative reactivity, the effect goes in the opposite direction. Because the effect was hypothesised to be independent of negative reactivity, the hypothesis is technically not supported. In hindsight, an interaction would not be an unreasonable expectation based on the underlying theory, and hence could easily be argued as in support of the hypothesis (or even initially hypothesised). However, what constituted support were specified prior to data collection, and an interaction that makes or breaks the effect was not part of it. In the loss condition, no effects of interest emerge. What little there is goes in the wrong direction, and is not close to statistical significance.

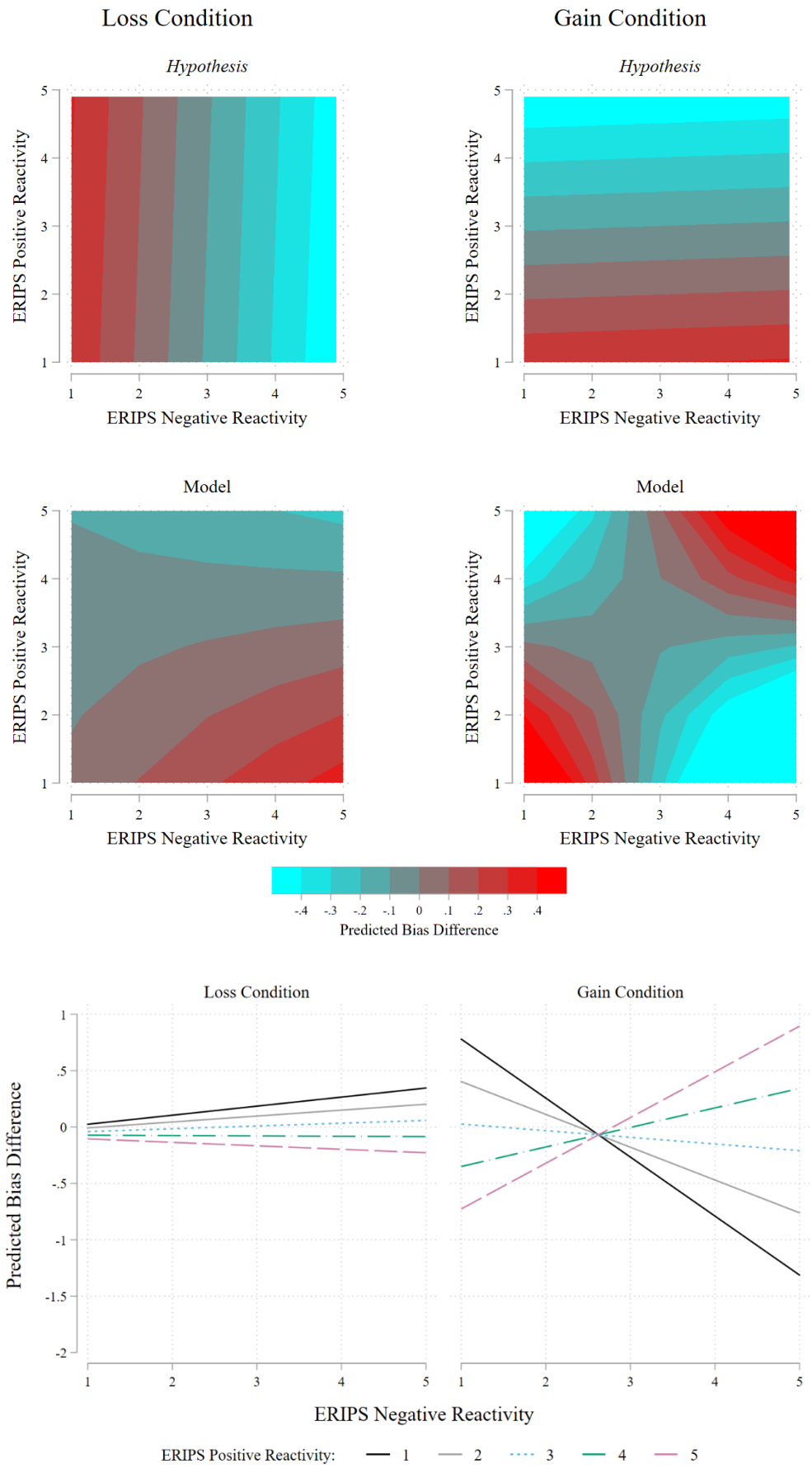


Figure C3 Predicted bias difference from emotional reactivity. Top row shows the hypothesized effects, while the middle and bottom rows show the results from the analysis. Negative scores mean individuals had a more liberal bias in the impossible condition compared to the easy condition (i.e., more motivated perception).

Table C5

Regression analysis of reward sensitivity and composite bias difference score (N = 43)

	B	*SE*	*t*	*p*	*β*
Main Effects[a]:					
1. Positive TTP	-0.001	0.12	0.01	.99	-0.00
2. Negative TTP	-0.278	0.14	-2.03	.05	-0.31
3. Gain/Loss condition[b]	-0.024	0.06	-0.44	.66	-0.07
Constant	*0.159*				
Interaction Effects[c]					
1. Positive TTP	1.470	0.93	1.59	.12	1.89
2. Negative TTP	1.427	1.06	1.34	.19	1.61
1 x 2	-2.055	1.22	-1.68	.10	-2.65
3. Gain/Loss condition	1.104	1.06	1.05	.30	3.05
3 x 1	-1.400	1.20	-1.17	.25	-3.30
3 x 2	-1.636	1.35	-1.21	.23	-3.64
3 x 1 x 2	2.016	1.53	1.31	.20	3.60
Constant	*-1.054*				

a: Main Effects: R^2 = .10, $F(3, 39)$ = 1.49, p = .23
b: Loss condition = 0, Gain condition = 1
c: Interaction Effects: R^2 = .17, $F(7, 35)$ = 1.05, p = .42
Note: Variables are *not* centred

Hypotheses 5 and 6. Hypothesis 5 was that more positive reward sensitivity should predict more motivated perception in the gain condition, and hypothesis 6 was the more negative reward sensitivity should predict more motivated perception in the loss condition. Again, both effects were hypothesised to be independent of each other. The hypotheses were originally planned to be tested separately with moderated regression analyses and gain/loss condition as moderator. I present a combined three-way interaction instead, partly for the sake of brevity, partly because we are exploring anyway, and partly because the original analyses yielded no more information than this analysis. The pre-planned exclusion criteria for reward sensitivity is applied, and the results are displayed in Table C5.

There was a main effect where more negative reward sensitivity predicted more motivated perception (B = -0.278, SE = 0.14, p = .05). This effect was predicted, but only for the loss group. Note that the first three coefficients in the moderation analysis describe the model for the loss group, while the remaining coefficients describe how the model changes for the gain group. The loss group has a trending interaction effect, meaning the effect of

negative reward sensitivity depends somewhat on the score of positive reward sensitivity, and vice versa (B = -2.005, SE = 1.22, p = .10). The other interaction effects are less reliable (i.e., more likely to arise by chance given no real effect), but the effect sizes suggest they cancel out the effects found in the loss condition.

Figure C4 shows a visual representation of the effects. It shows how the effects of positive and negative reward sensitivity are interdependent in the loss group, but independent in the gain group. However, the effect does not at any point become significant. The gain group shows only a small unmoderated effect similar to that of the main effect, where more negative reward sensitivity predicted more motivated perception. Linear combination shows the effect to be of similar magnitude at the mean[11] of positive reward sensitivity (B = -0.241, SE = 0.19, p = .20), but not significant. Linear combination can also be used to find the effect at mean positive reward sensitivity in the loss group, and it too is of similar magnitude (B = -0.237, SE = 0.24, p = .39).

Other Analyses of Interest

I also ran the analysis with the bias difference score from the dot motion task, which showed some evidence of motivated perception (i.e., on average more liberal bias in the impossible condition than in the easy condition). The results were for the most part similar, but generally larger. Some of the interactions changed slightly. For example, the interaction effect between positive and negative reward sensitivity, which was originally just present in the loss condition, remained present in the gain condition. Using this as an outcome variable would offer more interesting (not to mention significant) results than the planned analyses, which highlights how analytical flexibility allows anyone to find the effects they want (Simmons et al., 2011).

Rerunning the analyses with bias from the impossible cell density task, which showed some evidence of a desirability bias, yielded only null results.

Discussion

The preceding analyses are not cases of real effects "hiding" in the interactions. They are cases of complex models with low parameter to observation ratios being overfitted to noise. That is why the models – even though the effects are big enough to be interesting – do not reach significance: The likelihood of getting similar or bigger effects by chance is simply too high. This is also apparent if one checks the distribution of data points against the

[11] Positive TTP: M = 0.81, SD = 0.24, see also Table 5.

predicted values: A low sample size means the full range of data is not well represented, meaning the data is particularly sparse towards the edges of the distribution[12]. Adding interaction terms to a regression, then, allows the model to be fitted to the quirks and idiosyncrasies of the few edge cases. This is especially true for reward sensitivity, where the ceiling effects exacerbates the problem.

If the validity of the outcome variable was unknown, the effects could readily have been interpreted in light of theory. For example, both analyses suggest that potential gains and potential losses are processed somewhat differently. Arousability seem to be relevant for potential gains, where relative positive to negative arousability is more important than absolute arousability. In contrast, relative reward sensitivity, especially sensitivity to negative rewards, is more important when considering losses. I initially expected that I had to run many analyses to demonstrate Simmons et al. (2011), but even the planned analysis showed interesting results that are, if not a confirmation of the hypotheses, at least something to work with.

Because we can be quite certain that the outcome variable consists of nothing but random variation, the analyses serves as a demonstration for why it is important to have enough power for complex analyses. It also demonstrates the importance of validity checks and a detailed specification of what would constitute support for the hypotheses prior to analyses, as some effects are likely to appear anyway. Signal and noise can look quite similar when the quality of the input is unknown.

[12] This is obviously true no matter the sample size, but with low sample sizes, the data become sparse closer to the mean. With N = 50, you would expect 15 observations outside 1 standard deviation, while with N = 500, you would expect 150 observations outside 1 standard deviation (and 20 observations outside 2 standard deviations).

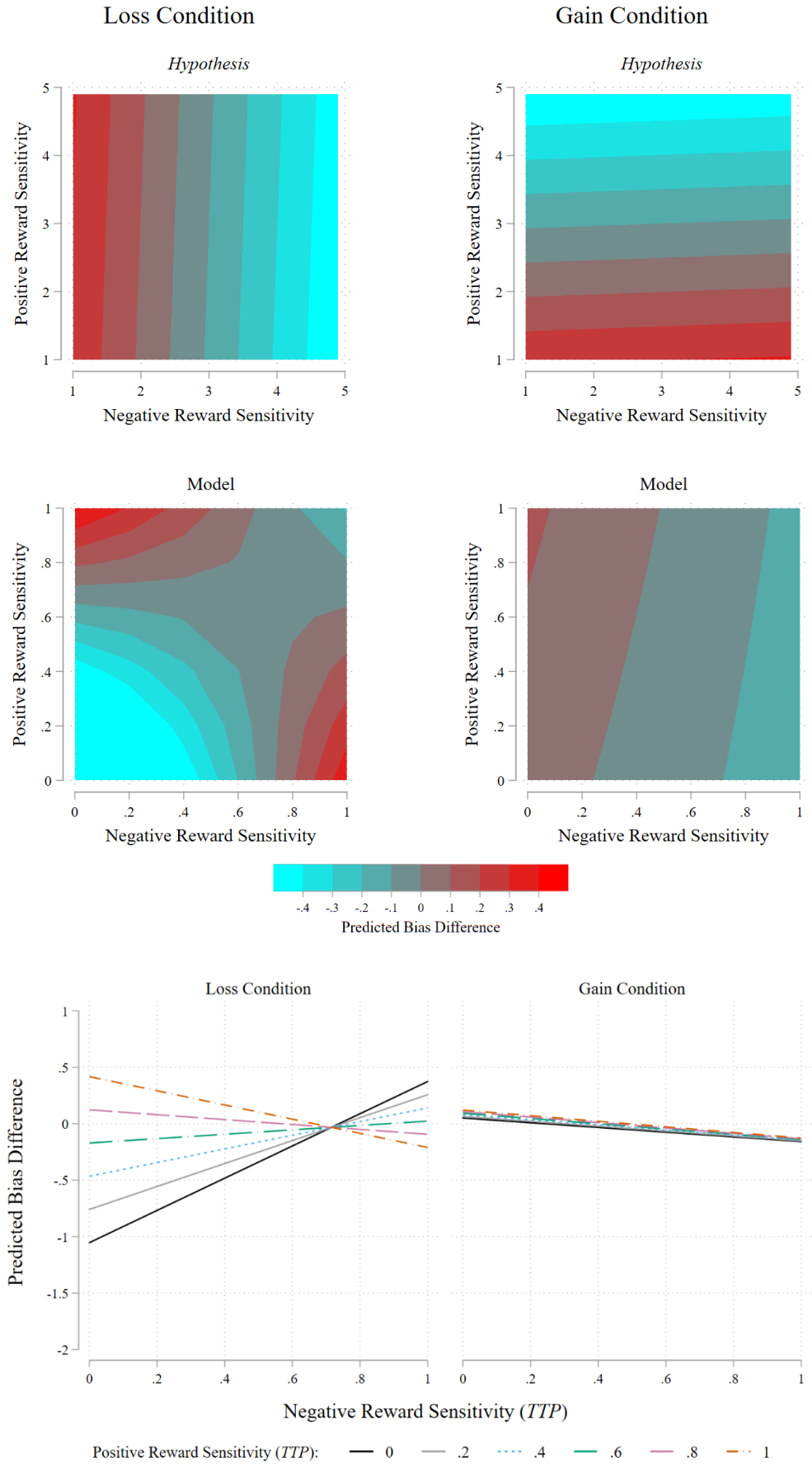


Figure C4 Predicted bias difference from rewards sensitivity. Top row shows the hypothesized effects, while the middle and bottom rows show the results from the analysis. Negative scores mean individuals had more liberal bias in the impossible condition compared to the easy condition (i.e., more motivated perception).

Appendix D: Correlations for Reward Sensitivity and Emotional Reactivity

Table D1

Correlations between emotional reactivity and reward sensitivity

	Positive Reactivity		Negative Reactivity	
	r	*p*	*r*	*p*
No exclusions (N=64)				
Transfer Test Phase (Positive)	.14	.28	.09	.46
Block 6 (Positive)	.17	.17	.10	.43
Saturation Score (Positive)	.12	.34	.11	.38
Transfer Test Phase (Negative)	**-.25**	**.044**	.07	.61
Block 6 (Negative)	.05	.71	-.04	.78
Saturation Score (Negative)	.10	.41	-.04	.74
Reward sensitivity exclusions only (N=52)				
Transfer Test Phase (Positive)	-.07	.63	.06	.69
Block 6 (Positive)	.01	.96	.10	.48
Saturation Score (Positive)	.04	.76	.06	.68
Transfer Test Phase (Negative)	**-.28**	**.047**	.18	.20
Block 6 (Negative)	.05	.72	-.05	.73
Saturation Score (Negative)	.17	.23	-.08	.58
Original exclusions only (N=55)				
Transfer Test Phase (Positive)	.18	.19	.10	.49
Block 6 (Positive)	.21	.12	.11	.42
Saturation Score (Positive)	.17	.21	.13	.34
Transfer Test Phase (Negative)	**-.27**	**.046**	.04	.75
Block 6 (Negative)	.06	.68	-.07	.62
Saturation Score (Negative)	.12	.37	-.03	.81
All exclusions (N=43)				
Transfer Test Phase (Positive)	-.01	.93	.07	.67
Block 6 (Positive)	.08	.60	.14	.39
Saturation Score (Positive)	.11	.50	.08	.59
Transfer Test Phase (Negative)	**-.30**	**.049**	.16	.31
Block 6 (Negative)	.09	.56	-.11	.49
Saturation Score (Negative)	.21	.17	-.07	.66

Note: Correlations with $p < .05$ are displayed in bold

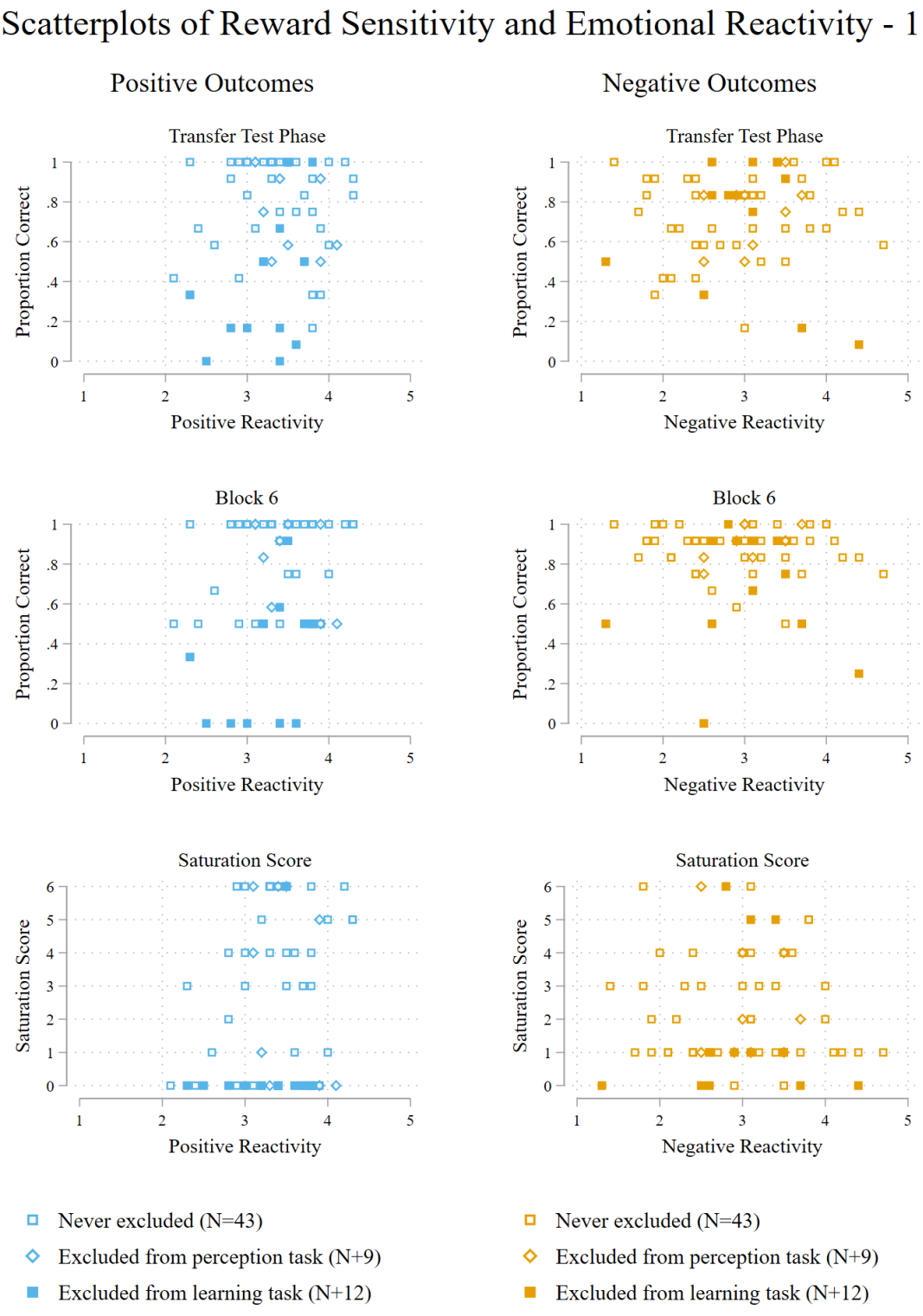


Figure D1 Scatterplots of relationship between reward sensitivity and emotional arousability

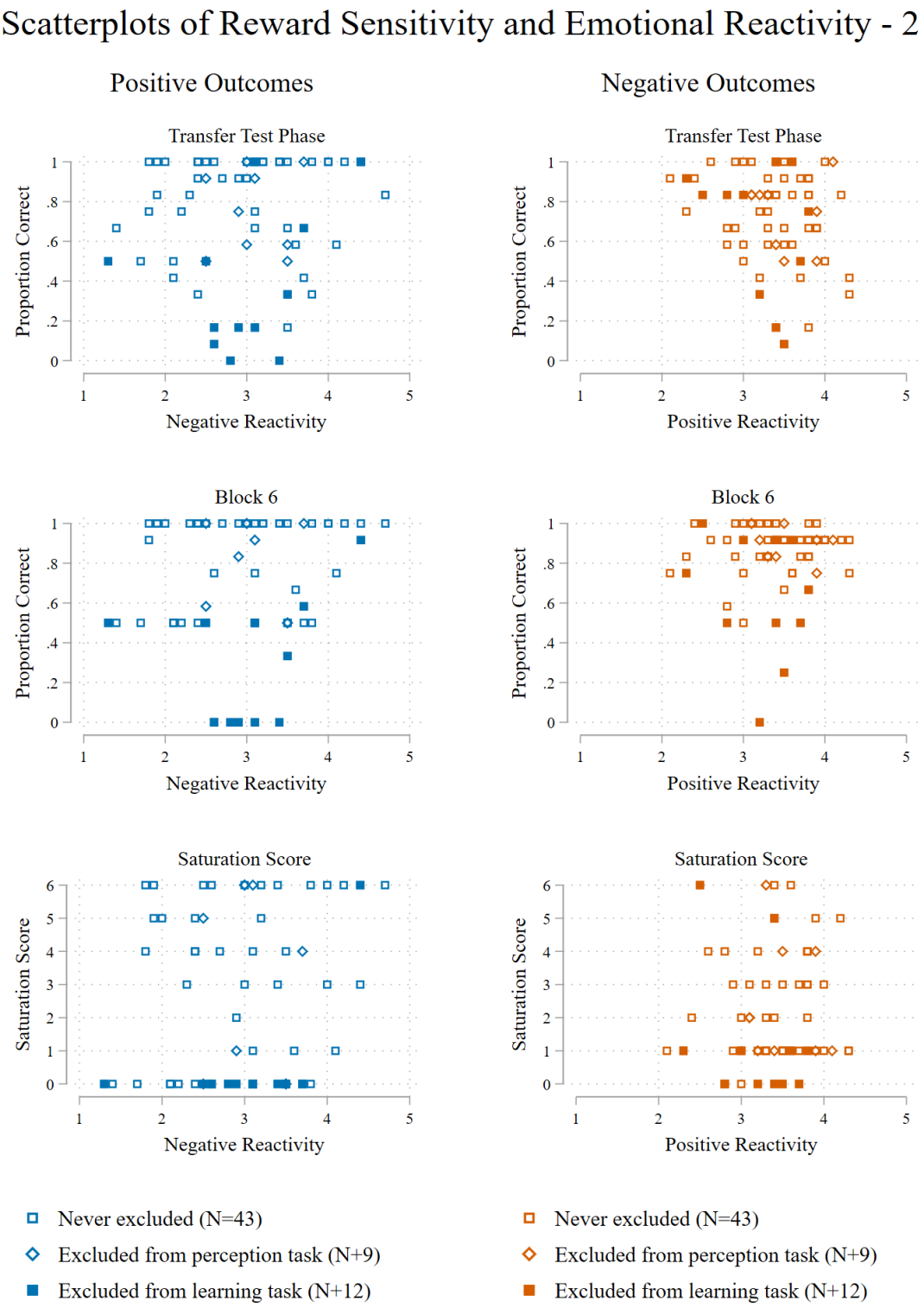


Figure D2 Scatterplots of relationship between reward sensitivity and emotional arousability

Appendix E: Screenshots from the Experiment

NB: Some images are cropped. The instructions were presented with more surrounding space. Some figures presented in the main text are repeated here.

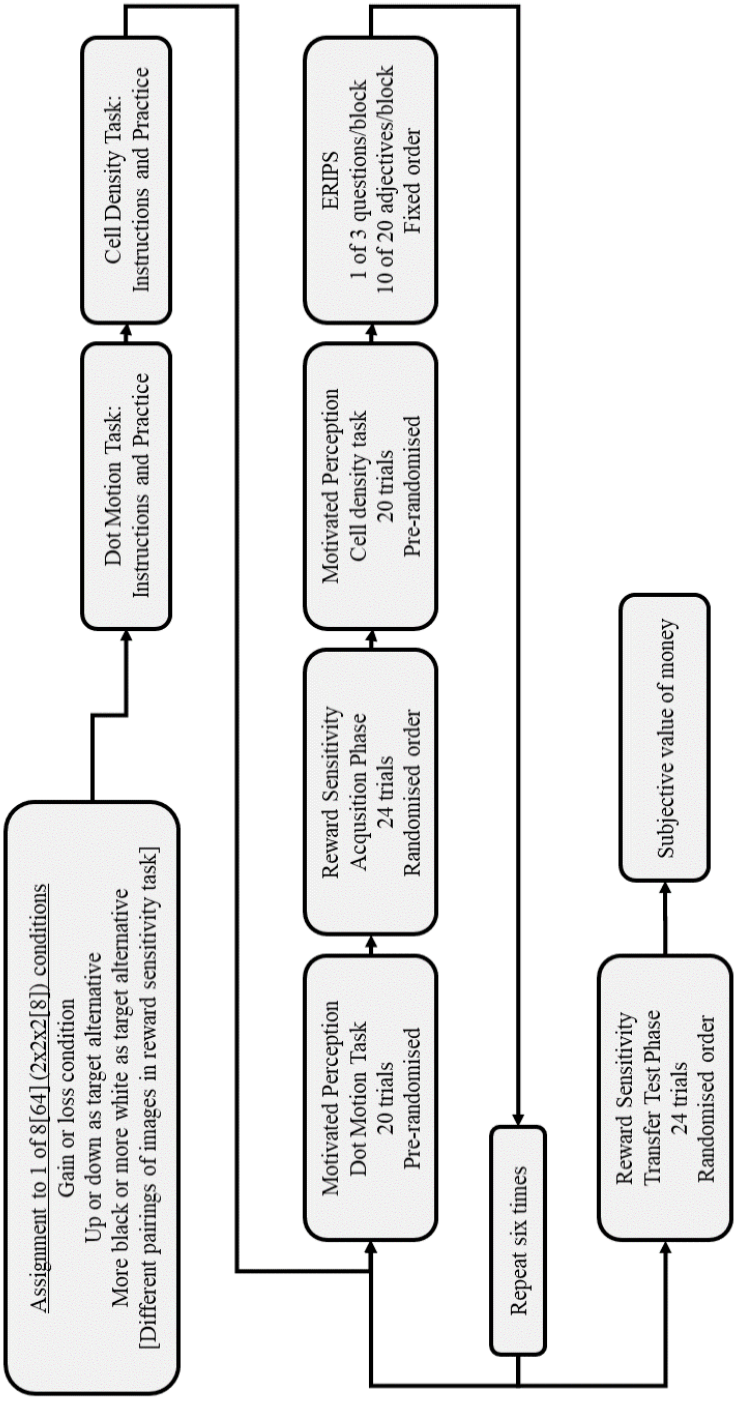


Figure E1 Outline of the experimental blocks

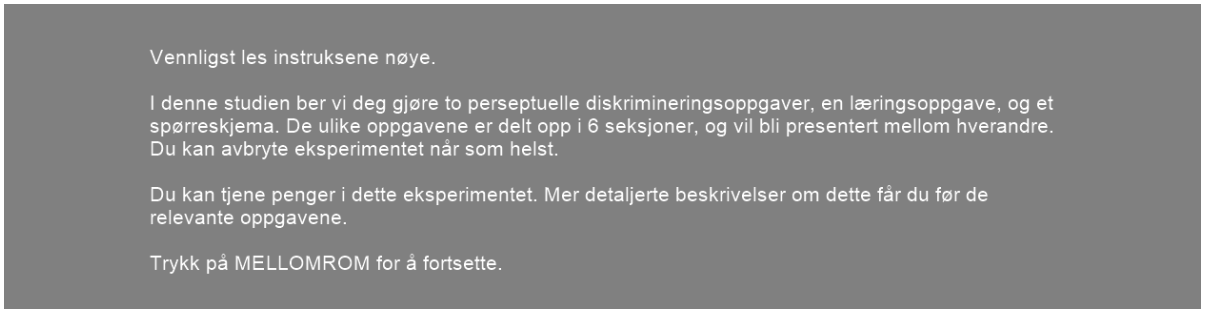


Figure E2 Initial instructions. Age and sex were entered before the experiment began together with the numbers specifying the conditions.

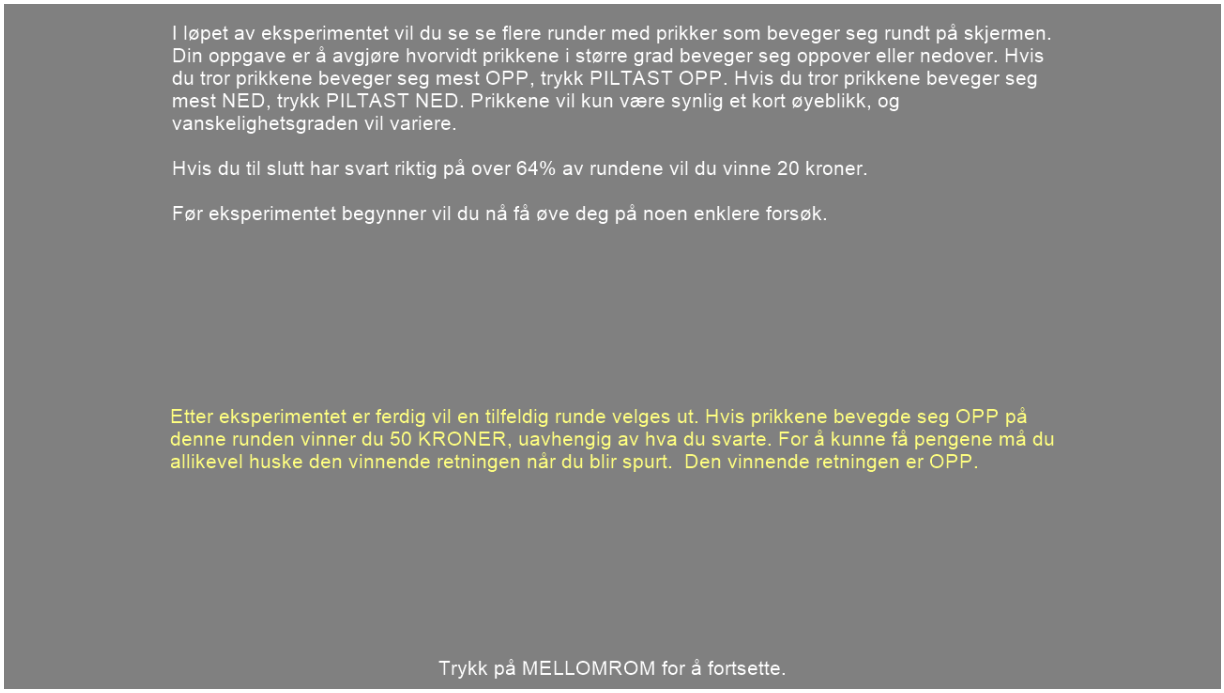


Figure E3 Instructions for the dot motion task. The yellow instructions in the middle changed depending on gain/loss condition and target alternative. This one is for the gain group.

Figure E4 Alternate dot motion task instructions for the loss group

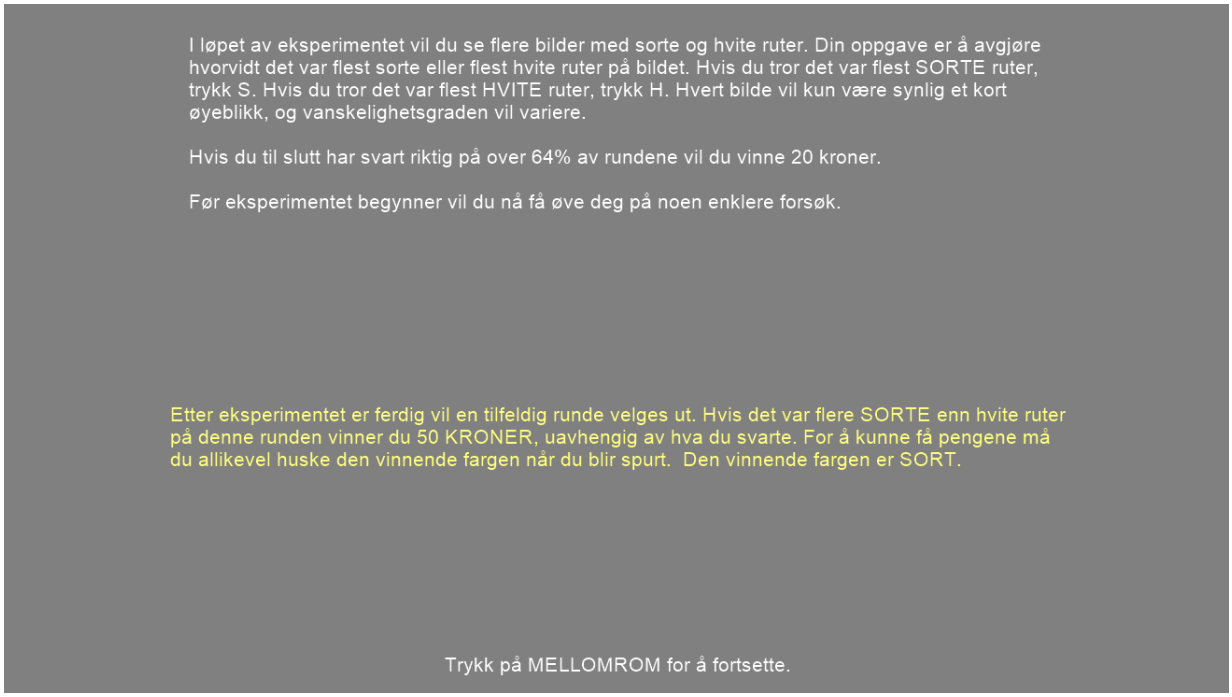


Figure E5 Instructions for the cell density task. The yellow instructions in the middle changed depending on gain/loss condition and target alternative. This one is for the gain group.

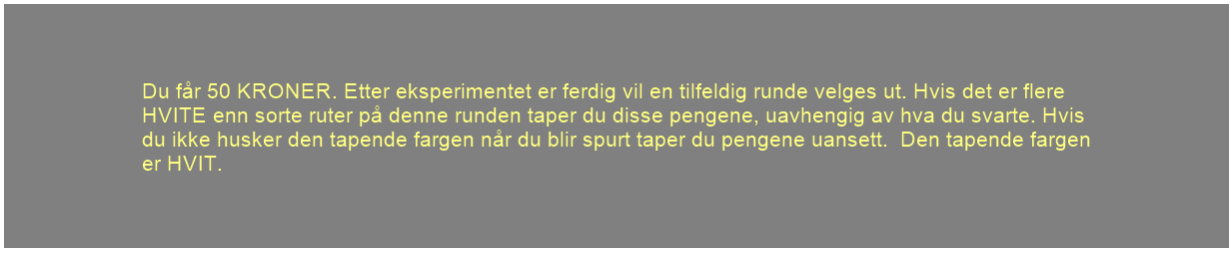


Figure E6 Alternate cell density task instructions for the loss group

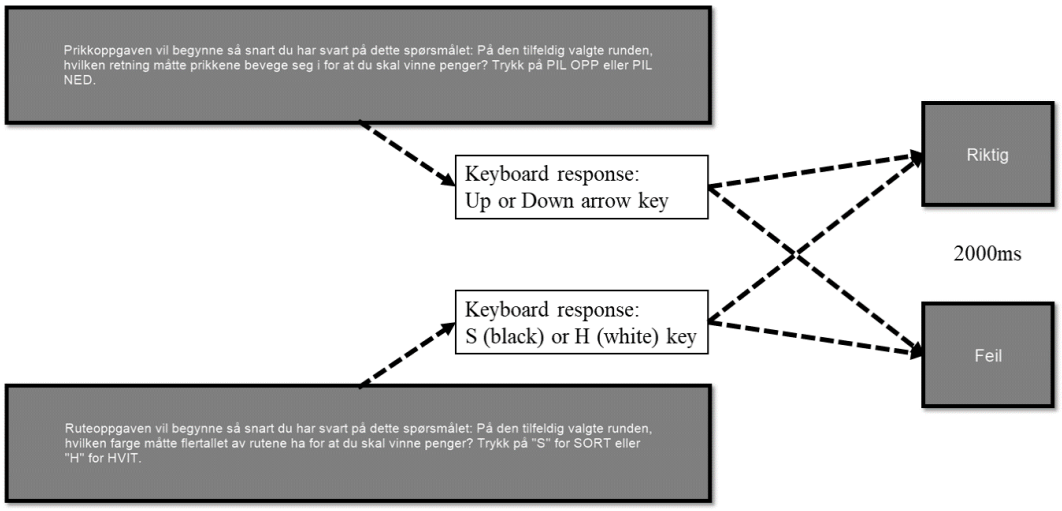


Figure E7 Exclusion test trials for both the dot motion task and the cell density task. The tasks started immediately after the feedback was shown.

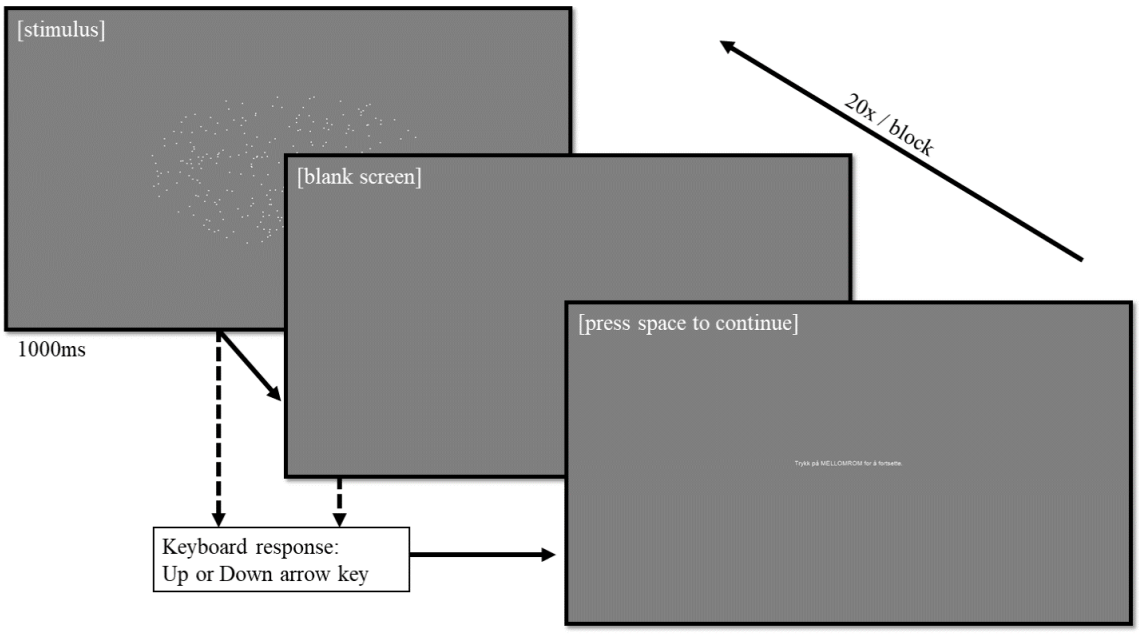


Figure E8 The dot motion task. 20 trials with 3 difficulty levels repeated in 6 blocks. Participants could respond as soon as the stimulus appeared. If participants did not answer for 5 seconds, a small reminder appeared on the blank screen reminding them which buttons they could press.

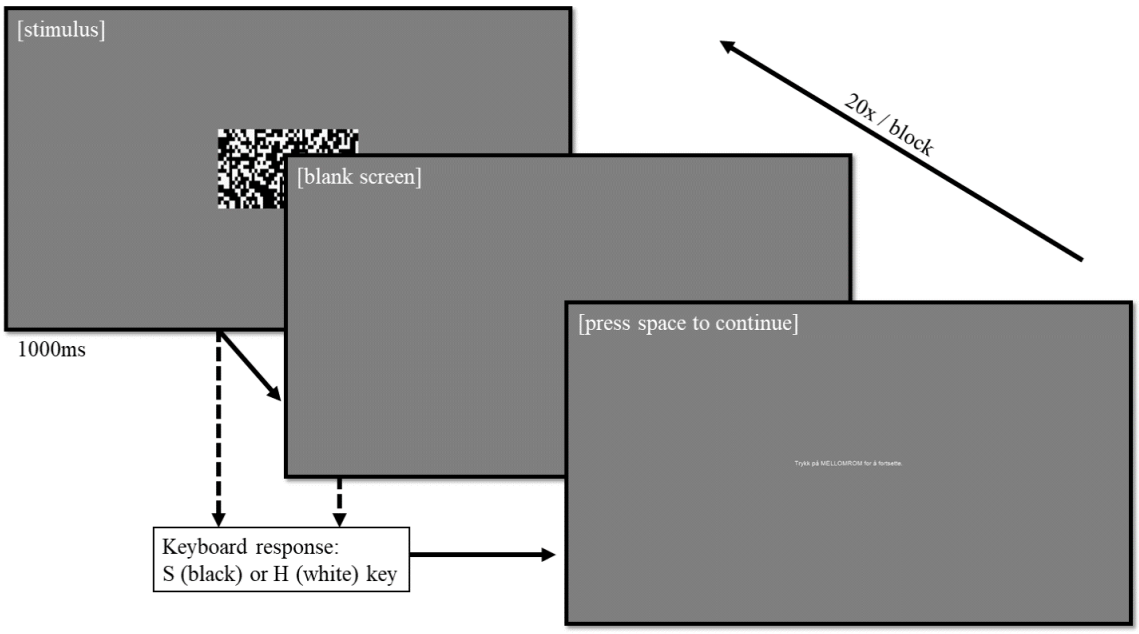


Figure E9 The cell density task. 20 trials with 3 difficulty levels repeated in 6 blocks. Participants could respond as soon as the stimulus appeared. If participants did not answer within 5 seconds, a small reminder appeared on the blank screen reminding them which buttons they could press.

Figure E10 Instructions for the first block in the acquisition phase of the learning task.

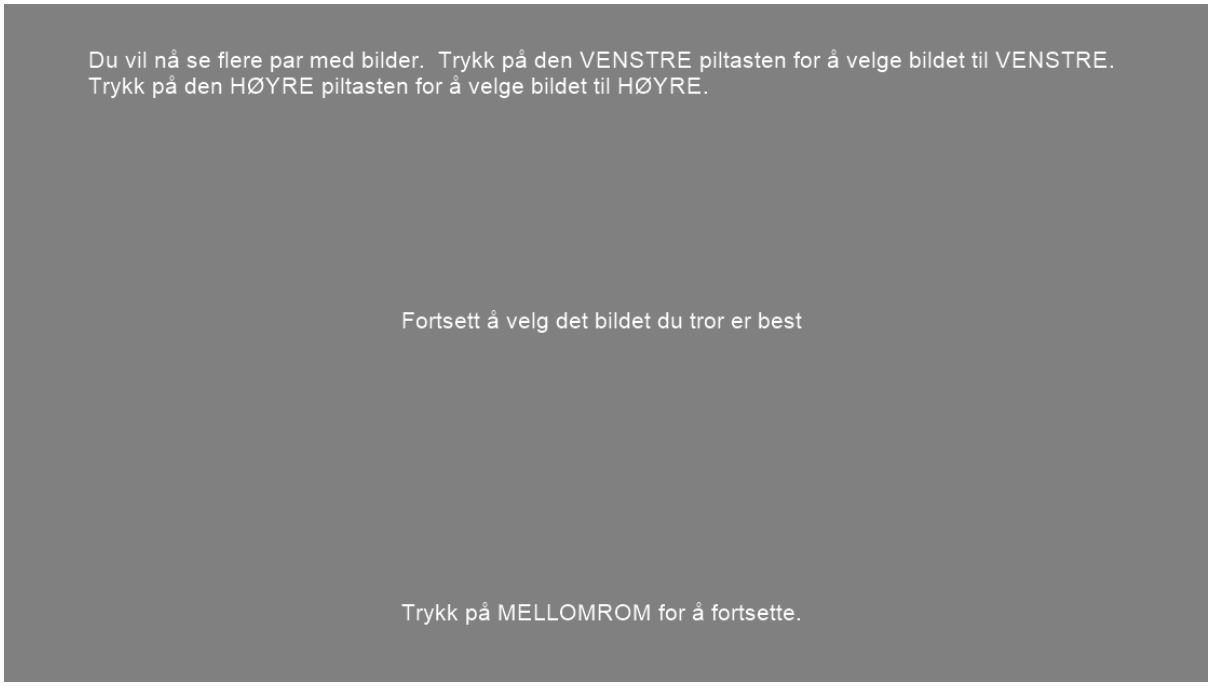


Figure E11 Instructions for the remaining blocks in the acquisition phase of the learning task.

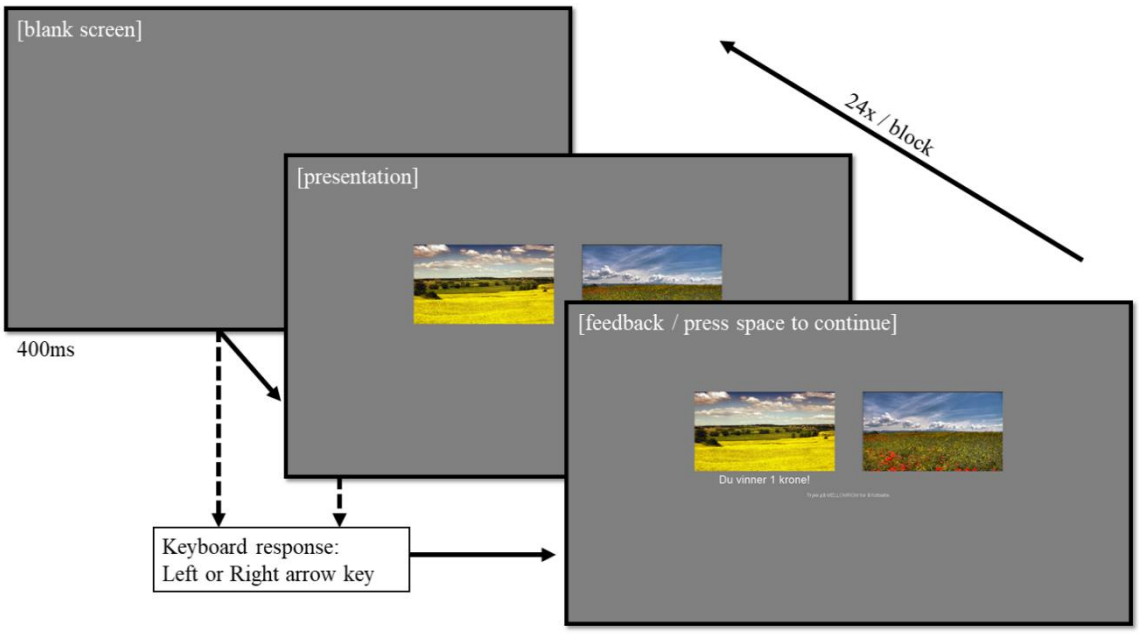


Figure E12 The acquisition phase of the learning task. There were four different pairs of pictures.

Figure E13 Instructions for the transfer test phase of the learning task.

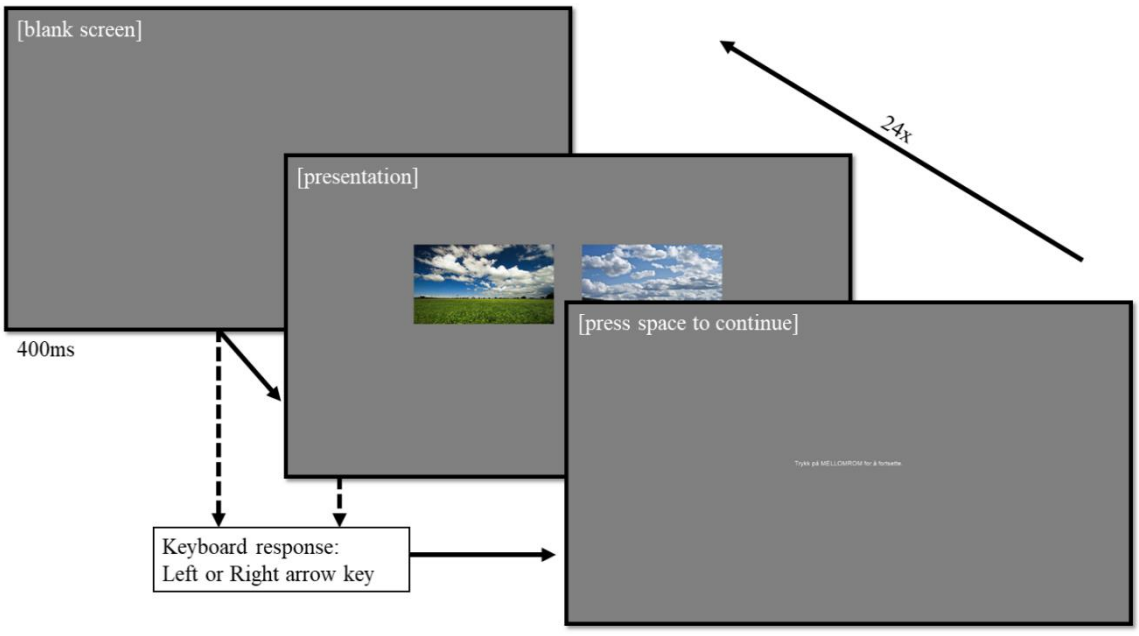


Figure E14 The transfer test phase of the learning task, with novel pairs and no feedback.

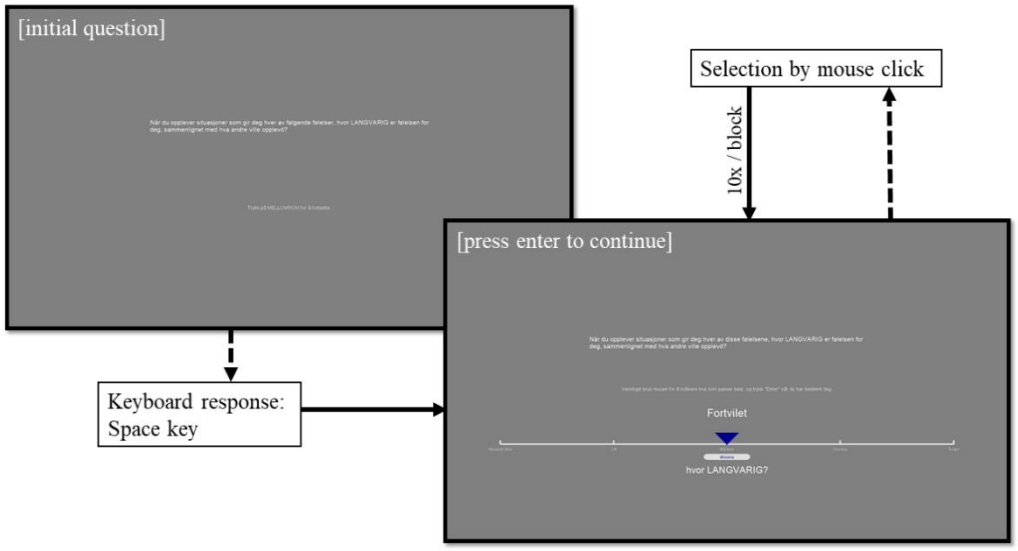


Figure E15 The ERIPS

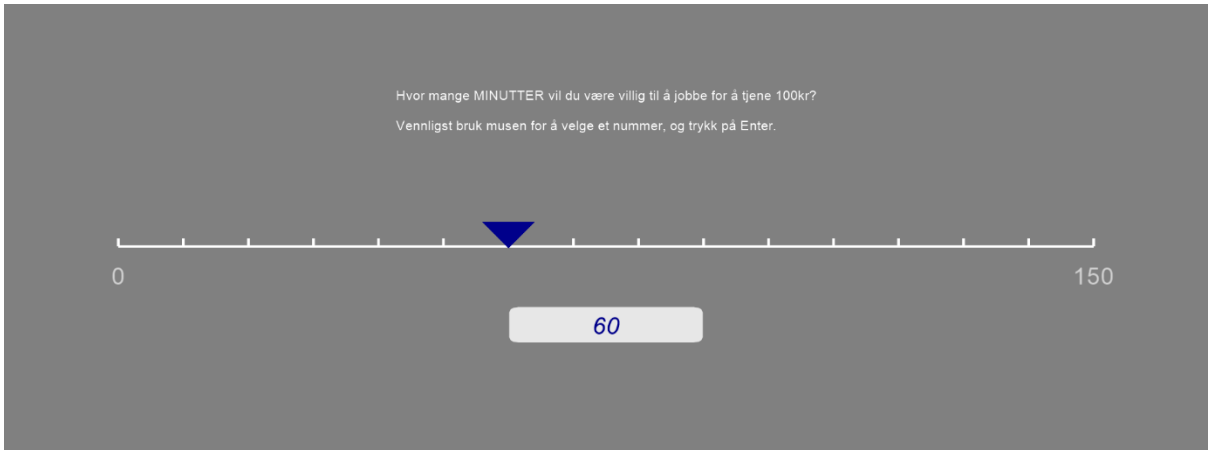


Figure E16 The final question about the subjective value of money.

After the final question, the program calculated their earnings and produced a text where participants learned their performance in the learning task, their accuracy scores, whether they had won/lost money, and how much money they ended up with in total.

Appendix F: Translation of ERIPS

This appendix consists of three parts: (1) The original translation report from Kyrre Svarva, (2) the translation back to English by Sigurd H. Lundheim, and (3) the final version.

Translation report from Kyrre Svarva

Chloe A. Rippera, Mark E. Boyesa, Patrick J.F. Clarke & Penelope A. Haskinga: **ERIPS** (Emotional Reactivity Intensity and Perseveration Scale), ref. Personality and Individual Differences 121 (2018) 93-99, basert på PANAS , ref. Watson, D., Clark, Lee Anna & Tellegen, A. (1988). Development and Validation of Brief Measures of Positive and Negative Affect: The PANAS Scales. Journal of Personality and Social Psychology, 1988, Vol. 54, No. 6, 1063-1070. Oversettelse ved Kyrre Svarva, SU-fakultetet, NTNU et al.

English intro text /instructions:	You have just completed a questionnaire in which you indicated how often you tend to have certain feelings or emotional experiences. However, individuals differ in the likelihood of experiencing specific feelings and the degree to which these feelings persist across time. In the following questionnaire you will be shown a list of feelings similar to those in the previous questionnaire but you are asked to make the following two different judgements concerning your tendency to experience such feelings
Kyrre's comment	I have reworded this somewhat, mostly in order to simplify the Norwegian version. Also, it does seem somewhat odd that the above texts specifies that the respondent is asked to make two judgements, while three questions are actually asked (likelihood, intensity and persistence).
Kyrres translation	[Du har nettopp besvart et spørreskjema der du oppga hvor ofte du har ulike følelser eller følelsesmessige opplevelser.] Vi mennesker varierer med tanke på hvor sannsynlig det er at vi opplever bestemte følelser, hvor sterke følelsene er og hvor lenge de varer. Hvor sannsynlig er det at du opplever følelsene på lista nedenfor, hvor intense er følelsene, og hvor lenge varer de for deg?

English reply alternatives:	Q. 1: 1 = Not at all likely, 2 = Slightly likely, 3 = Moderately likely, 4 = Very likely, 5 = Extremely likely Q. 2: 1 = Not at all intense, 2 = Slightly intense, 3 = Moderately intense, 4 = Very intense, 5 = Extremely intense Q. 3: 1 = Not at all persistent, 2 = Slightly persistent, 3 = Moderately persistent, 4 = Very persistent, 5 = Extremely persistent
Kyrre's comment	The use of "Moderately", "Very" and "Extremely" ("Middels" – "Svært" – "Ekstremt") makes the scaling appear a bit "stretched" at the right end – a more "even" scaling might perhaps have used "Moderately" – "Quite" – "Very" ("Middels" – "Ganske" – "Svært"). Another thing is that given the way the three questions are formulated, it would have made sense to anchor the scale at the midpoint to «the average person» that is mentioned in each of them, e.g. 1 = Far less, 2 = Somewhat less, 3 = About the same, 4 = Somewhat more, 5 = Far more (in Norwegian e.g. 1 = Mye mindre, 2 = Noe mindre, 3 = Omtrent det samme, 4 = Noe mer, 5 = Mye mer). I assume, however, this would be going too far from the original.
Kyrres translation	1 = Absolutt Ikke sannsynlig, 2 = Litt sannsynlig, 3 = Middels sannsynlig, 4 = Ganske sannsynlig, 5 = Svært sannsynlig 1 = Absolutt ikke intens, 2 = Litt intens, 3 = Middels intens, 4 = Ganske intens, 5 = Svært intens 1 = Absolutt ikke langvarig, 2 = Litt langvarig, 3 = Middels langvarig, 4 = Ganske langvarig, 5 = Svært langvarig

The three main questions:

1	English text	Emotional reactivity: When exposed to a situation that would make the “average” person experience this feeling, how likely is it that you will experience this particular feeling?
1	Kyrre’s comment	Rather than going for a "word-by-word" translation, I have attempted to write a sentence I believe a Norwegian speaker might have used in order to convey the intended meaning. Also, I have taken the liberty of skipping the ”heading” part of each question, since I think the intended meaning is conveyed well enough by the question itself.
1	Kyrre’s translation	I en situasjon der en gjennomsnittlig person ville oppleve hver av disse følelsene, hvor *sannsynlig* er det at *du selv* ville oppleve den?
2	English text	Emotional intensity: When you are experiencing a situation that does make you feel this way, how intense is the feeling compared to how other people feel?
2	Kyrre’s comment	Rather than going for a "word-by-word" translation, I have attempted to write a sentence I believe a Norwegian speaker might have used in order to convey the intended meaning.
2	Kyrre’s translation	Når du opplever situasjoner som gir deg hver av disse følelsene, hvor *intens* er følelsen for *deg*, sammenlignet med hva andre ville opplevd?
3	English text	Emotional perseveration: When you are experiencing a situation that does make you feel this way, how intense is the feeling compared to how other people feel?
3	Kyrre’s comment	Rather than going for a "word-by-word" translation, I have attempted to write a sentence I believe a Norwegian speaker might have used in order to convey the intended meaning.
3	Kyrre’s translation	Når du opplever situasjoner som gir deg hver av disse følelsene, hvor *langvarig* er følelsen for *deg*, sammenlignet med hva andre ville opplevd?

PANAS items:

Since the PANAS instrument has been available since 1988, I have assumed an “official”, approved-by-the-authors Norwegian translation may exist. I have attempted to search for this online, but so far in vain, although it is apparent that the instrument has been used in Norwegian settings, as evident from a number of articles/reports. However, none of the sources I have found mention whether their version was approved by the original authors / copyright holders. Also, unfortunately, few list the actual Norwegian items employd. So far, I have only found three (four) projects that have employed shortened versions of PANAS, and who list the Norwegian item words used in their reports:

- Hansen, Karl Petter Heie: "Hvilke sammenhenger er det mellom selvbestemt motivasjon, autonomistøttende treningsklima, selvoppfattet kompetanse, trening og vitalitet/velvære blant studenter?" Mastergradsoppgave, Pedagogisk forskningsinstitutt, UiO, 2009. https://www.duo.uio.no/handle/10852/31139
- Haslestad, Linn Cecilie & Nybakken, Camilla: "Hvordan påvirker formelle kompetansehevende tiltak arbeidstakers subjektive velvære, arbeidsmotivasjon og ytelse på arbeidsplassen, og i hvilken grad fungerer behovet for kompetanse som en mediator på dette forholdet?" Masteroppgave i strategi og kompetanseledelse ved Høgskolen i Buskerud avd. Hønefoss, 2013. http://docplayer.me/42383776-Linn-cecilie-haslestad-camilla-nybakken-5-15-2013.html.
- The norLAG project (NOVA/SSB et al., https://norlag.nova.no/, more specifically, https://blogg.hioa.no/norlag/files/2016/09/REV-okt-2014-NorLAGForsknInstrumentene-2012-fra-ial-endelig-1.pdf, pages 105-106 (same item selection and translation also used in Næss, Siri & Hansen, Thomas: "Naturelskere og naturbrukere", Tidsskrift for samfunnsforskning04 / 2012 (Volum 53), side 406-427. https://www.idunn.no/tfs/2012/04/naturelskere_og_naturbrukere

Note that the sources did not specify which of the original English items each of their items were intended to correspond to. Thus, some items were somewhat more difficult to place than others. It is also possible that some Norwegian item words have been intended to cover more than one of the original English PANAS items. Consequently, the placement of these words in the tables below is to some degree arbitrary (this applies to the words shaded in yellow).

Column legend:
A: English text as used by C. A Ripper et al. (2018). Items are listed here in sequence of their Table 2. Ripper's list is identical to original PANAS from 1988, however, the sequence is different.
B: Translation used by Hansen, Karl Petter Heie.
C: Translation used by Haslestad, Linn Cecilie & Nybakken, Camilla
D: Translation used in the norLAG project and by Næss, Siri & Hansen, Thomas
E: My (Kyrre's) translation.

PANAS positive items

	A	B	C	D	E
1	Interested	Interessert		Interessert	Interessert
2	Excited	Begeistret	Begeistret	Begeistret	Begeistret
3	Strong				Sterk
4	Enthusiastic	Entusiastisk	Entusiastisk	Oppglødd	Entusiastisk
5	Proud				Stolt
6	Alert	Årvåken	Oppvakt/klar	Årvåken	Årvåken
7	Inspired	Inspirert	Inspirert	Inspirert	Inspirert
8	Determined	Målbevisst	Målbevisst	Målbevisst	Målbevisst
9	Attentive				Oppmerksom
10	Active	Livlig	Livlig		Livlig

Comments from Kyrre:
N1: "Begeistret" (all columns) and "Oppglødd" (col. D) might both have been acceptable as translations of «Excited».
N2: "Oppvakt" (col. C) is not a very good translation of "Alert" (and it does not fit in elswhere) – it would rather mean "Bright" (in the sense of being intelligent". "Klar" (col. C) can, depending on context, mean "Ready" (… for something), but does not work well here, as it can also mean "Finished" (or even "Tired" in Trønder dialect).
N3: "Aktiv" may certainly be a good option here, but I agree with the listed authors that "Livlig" is better when the intention is to capture the sense, or feeling associated with being "active"

PANAS negative items

	A	B	C	D	E
1	Distressed	Fortvilet	Fortvilet		Fortvilet
2	Upset	Oppskaket	Oppskaket	Oppskaket	Oppskaket
3	Guilty				Skyldig
4	Scared			Skremt	Skremt
5	Hostile				Fiendtlig
6	Irritable	Irritert	Irritert	Irritert	Irritabel
7	Ashamed				Skamfull
8	Nervous	Nærvøs	Nervøs	Nervøs	Nervøs
9	Jittery		Redd	Redd	Skjelven
10	Afraid	Bekymret	Bekymret	Bekymret	Bekymret

Comments from Kyrre:
P1: The positive patt of the original PANAS contain three words signifying degrees of apprehensiveness or fear, "Scared", "Jittery" and "Afraid". Norwegian words employed in cols. B, C and D to correspond to one or more of

those include “Redd”, “Bekymret” and “Skremt”. Without access to documentation of the translation processes, it seems slightly strange that they have all used “Bekymret”, as this most often would be translated into “Worried”. Here are the Oxford and Merriam-Webster’s online dictionaries’ explanations of the words (slightly edited):

Scared (Oxford):
Fearful; frightened (‘she's scared stiff of her dad’; ‘I was scared I was going to kill myself’; ‘he's scared to come to you and ask for help’
Scared (Merrriam-Webster):
Thrown into or being in a state of fear, fright, or panic (scared of snakes;·scared to go out)

Jittery/Jitters/Jitter (Oxford):
1: (jitters) Feelings of extreme nervousness. (‘a bout of the jitters’)
2: Slight irregular movement, variation, or unsteadiness, especially in an electrical signal or electronic device (‘picture jitter’)
3: (jittery) Nervous or unable to relax (‘caffeine makes me jittery’)
Jittery/Jitters/Jitter (Merriam-Webster):
1: jitters (plural): a sense of panic or extreme nervousness (had a bad case of the jitters before his performance)
2: the state of mind or the movement of one that jitters
3: irregular random movement (as of a pointer or an image on a television screen); also : vibratory motion

Afraid (Oxford):
Feeling fear or anxiety; frightened. (‘I'm afraid of dogs’, ‘she tried to think about the future without feeling afraid’)
1: Worried that something undesirable will occur or be done (‘she was afraid that he would be angry’)
2: Unwilling or reluctant to do something for fear of the consequences (‘I'm often afraid to go out on the streets’)
3: Anxious about the well-being or safety of (‘William was suddenly afraid for her’)
Afraid (Merriam-Webster):
1: filled with fear or apprehension (afraid of machines; was afraid for his job)
2: filled with concern or regret over an unwanted situation (I'm afraid I won't be able to go).
3 : having a dislike for something (She's not afraid of hard work. [=she's not unwilling to work hard])

The two main online Norwegian dictionaries are Språkrådets/UiBs Bokmålsordboka/Nynorskordboka (here shortened to UiB) and Det norske akademis ordbok, naob.no (NAOB).

“Scared” clearly corresponds most closely with “Skremt” (“I was scared by something” = “Jeg ble skremt av noe”).
Skremme/skremt (UiB):
1: gjøre redd (“du skremmer ikke meg”) [make afraid] (…)
Skremme/skremt (NAOB):
1: gjøre (plutselig) redd, forskrekket; inngi frykt; forskrekke (“hesten ble skremt av sin egen skygge”) [make (suddenly) afraid, instil fear, frighten] (…)

“Jittery”, Norwegian dictionary entries:
Skjelven (UiB):
som skjelver (skjelve = dirre, skake [shiver, shake]) (“bli skjelven” = “bli urolig, redd” / “være skjelven på hendene”, “i stemmen”)
Skjelven (NAOB):
som dirrer, rister, skjelvende, engstelig og redd (for å handle) [shivering, shaking, anxious and afraid (to do something)]

In Norwegian, a person in a relatively high state of fear might say, “Jeg var helt skjelven” (= “I was totally shaking [with fear]”). Since “skjelven” as well as “Jittery” is based on the concept of being in a state marked by shivering/shaking, and since both words carry the meaning of doing this due to fear/apprehensiveness, I have ended up with “skjelven” as my translation.

“Redd”, Norwegian dictionary entries:

Redd (UiB):
1: engstelig, skremt [anxious, scared] ("jeg er redd du ikke vil lykkes" = I'm afraid you won't succeed" / "bli redd" = "become afraid") (…)
Redd (NAOB):
1: som føler (og viser) frykt; engstelig; skremt [who feels/shows fear, anxious, frightened] ("ikke vær redd, jeg skal passe på deg" = "don't be afraid, I will look after you")

The above listed sources all employ the word "Bekymret":
Bekymret (UiB):
1: engste, uroe [anxious, uneasy/worried] ("situasjonen bekymrer meg" = I am worried by the situation")
Bekymret (NAOB):
nervøs, engstelig, urolig (for) ("ha et bekymret uttrykk i ansiktet" = "have a worried facial expression"; "hvor har du vært? Jeg har vært så bekymret for deg!" = "where have you been? I have worried so much for you")

Being "Redd" clearly means to be "Afraid" or "Frightened". Comparing this with "Bekymret", which was used by all of the above listed authors, I find that "Redd" corresponds rather better with "Afraid" than "Bekymret".

However, we may also evaluate the various words with respect to their strenght of apprehension/fear (i.e. emotional strength implied by the word in everyday parlance). The following is based on the dictionary examples above and my own, highly subjective opinion:

Scared: strong
Jittery: medium to strong
Afraid: weak to medium

Skremt: medium to strong
Skjelven: strong
Redd: Undifferentiated, weak to strong
Bekymret: weak to medium

If we want three words that altogether imply all degrees of apprehension/fear, we might drop the most undifferentiated option ("Red"), and use the weaker "Bekymret" instead. Also, considerations of this type may be one reason why "Bekymret" was included by the listed authos. Consequently, I have landed on drpping "Redd" in favour of "Bekymret", in spite of the latter's relative lack of correspondence with the English terms.

Lastly, note that during my searches for information on the PANAS, I came across an article by Thompson, E. R. (2007), who has created a10-item (5+5) English-language version of the PANAS based on a multinational sample (I-PANAS-SF; although not with any Norwegians or Scandinavians) (see http://journals.sagepub.com/doi/10.1177/0022022106297301). He employs the following items: Upset, Hostile, Alert, Ashamed, Inspired, Nervous, Determined, Attentive, Afraid, and Active. In the event that you can do with ashortened PANAS version, this might be a way to go.

On the next page, you will find the questionnaire in one-column setup, as it will have to be in SelectSurvey (and presumably in many other online survey systems). For a printed questionnaire form, a two-column setup within each of the three main questions may be attempted.

ERIPS *(as originally translated by Kyrra Svarva)*

Vi mennesker varierer med tanke på hvor sannsynlig det er at vi opplever bestemte følelser, hvor sterke følelsene er og hvor lenge de varer. Hvor sannsynlig er det at du opplever følelsene på lista nedenfor, hvor intense er følelsene, og hvor lenge varer de for deg?

1. I en situasjon der en gjennomsnittlig person ville oppleve hver av disse følelsene, hvor *sannsynlig* er det at *du selv* ville oppleve den? *NB: Ett kryss for hver følelse.*

	Absolutt ikke	*Litt sann-synlig*	*Middels sannsynlig*	*Ganske sannsynlig*	*Svært sannsynlig*
1. ..... Interessert	☐	☐	☐	☐	☐
2. .....Begeistret	☐	☐	☐	☐	☐
3. .....Sterk	☐	☐	☐	☐	☐
4. ..... Entusiastisk	☐	☐	☐	☐	☐
5. ..... Stolt	☐	☐	☐	☐	☐
6. ..... Årvåken	☐	☐	☐	☐	☐
7. ..... Inspirert	☐	☐	☐	☐	☐
8. ..... Målbevisst	☐	☐	☐	☐	☐
9. .....Oppmerksom	☐	☐	☐	☐	☐
10. .....Livlig	☐	☐	☐	☐	☐
11. ..... Fortvilet	☐	☐	☐	☐	☐
12. ..... Oppskaket	☐	☐	☐	☐	☐
13. .....Skyldig	☐	☐	☐	☐	☐
14. ..... Skremt	☐	☐	☐	☐	☐
15. .....Fiendtlig	☐	☐	☐	☐	☐
16. .....Irritabel	☐	☐	☐	☐	☐
17. .....Skamfull	☐	☐	☐	☐	☐
18. .....Nervøs	☐	☐	☐	☐	☐
19. .....Skjelven	☐	☐	☐	☐	☐
20. ..... Bekymret	☐	☐	☐	☐	☐

2.Når du opplever en situasjon som gir deg hver av disse følelsene, hvor *intens* er følelsen for *deg*, sammenlignet med hva andre ville opplevd?

[*Gjenta PANAS-testleddene som ovenfor.*]

3.Når du opplever situasjoner som gir deg hver av disse følelsene, hvor *langvarig* er følelsen for *deg*, sammenlignet med hva andre ville opplevd?

[*Gjenta PANAS-testleddene som ovenfor.*]

Backtranslation by Sigurd H. Lundheim

ERIPS *(as translated back to English by Sigurd Lundheim)*

Vi mennesker varierer med tanke på hvor sannsynlig det er at vi opplever bestemte følelser, hvor sterke følelsene er og hvor lenge de varer. Hvor sannsynlig er det at du opplever følelsene på lista nedenfor, hvor intense er følelsene, og hvor lenge varer de for deg?

We humans vary considering how probable it is that we experience certain emotions, how strong these emotions are and how long they last. How probable is it that you experience the emotions on the list below, how intense are they and how long do they last?

1. I en situasjon der en gjennomsnittlig person ville oppleve hver av disse følelsene, hvor *sannsynlig* er det at *du selv* ville oppleve den? *NB: Ett kryss for hver følelse.*

   In a situation where an average person would experience each of these emotions, how probable is it that you yourself would experience the emotion?

	Absolutt ikke	*Litt sann-synlig*	*Middels sannsynlig*	*Ganske sannsynlig*	*Svært sannsynlig*
1. Interessert/**interested**	☐	☐	☐	☐	☐
2. Begeistret/**extatic**	☐	☐	☐	☐	☐
3. Sterk/**strong**	☐	☐	☐	☐	☐
4. Entusiastisk/**enthusiastic**	☐	☐	☐	☐	☐
5. Stolt/**proud**/	☐	☐	☐	☐	☐
6. Årvåken/**aware**	☐	☐	☐	☐	☐
7. Inspirert/**inspired**.....	☐	☐	☐	☐	☐
8. Målbevisst/**goal oriented**	☐	☐	☐	☐	☐
9. Oppmerksom/**alert**	☐	☐	☐	☐	☐
10. Livlig/**lively**	☐	☐	☐	☐	☐
11. Fortvilet/**despair**.....	☐	☐	☐	☐	☐
12. Oppskaket/**aroused**	☐	☐	☐	☐	☐
13. Skyldig/**guilty**	☐	☐	☐	☐	☐
14. Skremt/**scared**.....	☐	☐	☐	☐	☐

15. Fiendtlig/**hostile**	☐	☐	☐	☐	☐
16. Irritabel/**irritable**	☐	☐	☐	☐	☐
17. Skamfull/**shameful**	☐	☐	☐	☐	☐
18. Nervøs/**nervous**	☐	☐	☐	☐	☐
19. Skjelven/**shaken**.....	☐	☐	☐	☐	☐
20. Bekymret/**worried**	☐	☐	☐	☐	☐

2. Når du opplever en situasjon som gir deg hver av disse følelsene, hvor *intens* er følelsen for *deg*, sammenlignet med hva andre ville opplevd?

   When you experience a situation that gives you each of these emotions, how intens is the emotion for you, compared to what others would experience?

[*Gjenta PANAS-testleddene som ovenfor.*]

3. Når du opplever situasjoner som gir deg hver av disse følelsene, hvor *langvarig* er følelsen for *deg*, sammenlignet med hva andre ville opplevd?
   When you experience situations that gives you each of these emotions, how long lasting is the emotion for you compared to what others would experience?

[*Gjenta PANAS-testleddene som ovenfor.*]

Final version

ERIPS *(Final version. The introductory remark on top was dropped from the experiment as it was deemed superflous)*

Vi mennesker varierer med tanke på hvor sannsynlig det er at vi opplever bestemte følelser, hvor sterke følelsene er og hvor lenge de varer. Hvor sannsynlig er det at du opplever følelsene på lista nedenfor, hvor intense er følelsene, og hvor lenge varer de for deg?

1. I en situasjon der en gjennomsnittlig person ville oppleve hver av disse følelsene, hvor *sannsynlig* er det at *du selv* ville oppleve den?

	Absolutt ikke 1	*Litt sann-synlig* 2	*Middels sannsynlig* 3	*Ganske sannsynlig* 4	*Svært sannsynlig* 5
1. Interessert	☐	☐	☐	☐	☐
2. Begeistret	☐	☐	☐	☐	☐
3. Sterk	☐	☐	☐	☐	☐
4. Entusiastisk	☐	☐	☐	☐	☐
5. Stolt	☐	☐	☐	☐	☐
6. Våken	☐	☐	☐	☐	☐
7. Inspirert	☐	☐	☐	☐	☐
8. Bestemt	☐	☐	☐	☐	☐
9. Oppmerksom	☐	☐	☐	☐	☐
10. Livlig	☐	☐	☐	☐	☐
11. Fortvilet	☐	☐	☐	☐	☐
12. Opprørt	☐	☐	☐	☐	☐
13. Skyldig	☐	☐	☐	☐	☐
14. Skremt	☐	☐	☐	☐	☐
15. Fiendtlig	☐	☐	☐	☐	☐
16. Irritabel	☐	☐	☐	☐	☐
17. Skamfull	☐	☐	☐	☐	☐
18. Nervøs	☐	☐	☐	☐	☐
19. Bekymret	☐	☐	☐	☐	☐
20. Redd	☐	☐	☐	☐	☐

2. Når du opplever en situasjon som gir deg hver av disse følelsene, hvor *intens* er følelsen for *deg*, sammenlignet med hva andre ville opplevd?

[*Gjenta PANAS-testleddene som ovenfor.*]

3. Når du opplever situasjoner som gir deg hver av disse følelsene, hvor *langvarig* er følelsen for *deg*, sammenlignet med hva andre ville opplevd?

[*Gjenta PANAS-testleddene som ovenfor.*]

Appendix G: Approval from NSD

NSD NORSK SENTER FOR FORSKNINGSDATA

NSD sin vurdering

Prosjekttittel

Effekten av insentiver på persepsjon og læring

Referansenummer

356170

Registrert

02.10.2018 av Hans Fredrik Sunde - hansfsu@stud.ntnu.no

Behandlingsansvarlig institusjon

NTNU Norges teknisk-naturvitenskapelige universitet / Fakultet for samfunns- og utdanningsvitenskap (SU) / Institutt for psykologi

Prosjektansvarlig (vitenskapelig ansatt/veileder eller stipendiat)

Robert Biegler, robert.biegler@ntnu.no, tlf: 73590469

Type prosjekt

Studentprosjekt, masterstudium

Kontaktinformasjon, student

Hans Fredrik Sunde, hansfsu@stud.ntnu.no, tlf: 99588824

Prosjektperiode

20.09.2018 - 01.05.2019

Status

18.12.2018 - Vurdert

Vurdering (2)

18.12.2018 - Vurdert

NSD har vurdert endringen registrert 17.12.2018.

Det er vår vurdering at behandlingen av personopplysninger i prosjektet vil være i samsvar med personvernlovgivningen så fremt den gjennomføres i tråd med det som er dokumentert i meldeskjemaet med vedlegg den 18.12.2018. Behandlingen kan fortsette.

OPPFØLGING AV PROSJEKTET
NSD vil følge opp ved planlagt avslutning for å avklare om behandlingen av personopplysningene er avsluttet.

Lykke til med prosjektet!

Kontaktperson hos NSD: Eva J B Payne
Tlf. Personverntjenester: 55 58 21 17 (tast 1)

11.12.2018 - Vurdert

Det er vår vurdering at behandlingen vil være i samsvar med personvernlovgivningen, så fremt den gjennomføres i tråd med det som er dokumentert i meldeskjemaet den 11.12.2018 med vedlegg, samt i meldingsdialogen mellom innmelder og NSD. Behandlingen kan starte.

MELD ENDRINGER
Dersom behandlingen av personopplysninger endrer seg, kan det være nødvendig å melde dette til NSD ved å oppdatere meldeskjemaet. På våre nettsider informerer vi om hvilke endringer som må meldes. Vent på svar før endringen gjennomføres.

TYPE OPPLYSNINGER OG VARIGHET
Prosjektet vil behandle særlige kategorier av personopplysninger om helseforhold og alminnelige personopplysninger frem til 01.05.2019.

LOVLIG GRUNNLAG
Prosjektet vil innhente samtykke fra de registrerte til behandlingen av personopplysninger. Vår vurdering er at prosjektet legger opp til et samtykke i samsvar med kravene i art. 4 nr. 11 og art. 7, ved at det er en frivillig, spesifikk, informert og utvetydig bekreftelse, som kan dokumenteres, og som den registrerte kan trekke tilbake.

Lovlig grunnlag for behandlingen vil dermed være den registrertes uttrykkelige samtykke, jf. personvernforordningen art. 6 nr. 1 a), jf. art. 9 nr. 2 bokstav a, jf. personopplysningsloven § 10, jf. § 9 (2).

PERSONVERNPRINSIPPER
NSD vurderer at den planlagte behandlingen av personopplysninger vil følge prinsippene i personvernforordningen:

- om lovlighet, rettferdighet og åpenhet (art. 5.1 a), ved at de registrerte får tilfredsstillende informasjon omog samtykker til behandlingen
- formålsbegrensning (art. 5.1 b), ved at personopplysninger samles inn for spesifikke, uttrykkelig angitte ogberettigede formål, og ikke viderebehandles til nye uforenlige formål

- dataminimering (art. 5.1 c), ved at det kun behandles opplysninger som er adekvate, relevante ognødvendige for formålet med prosjektet
- lagringsbegrensning (art. 5.1 e), ved at personopplysningene ikke lagres lengre enn nødvendig for åoppfylle formålet

DE REGISTRERTES RETTIGHETER
Så lenge de registrerte kan identifiseres i datamaterialet vil de ha følgende rettigheter: åpenhet (art. 12), informasjon (art. 13), innsyn (art. 15), retting (art. 16), sletting (art. 17), begrensning (art. 18), underretning (art. 19), dataportabilitet (art. 20).

NSD vurderer at informasjonen som de registrerte vil motta oppfyller lovens krav til form og innhold, jf. art.
12.1 og art. 13.

Vi minner om at hvis en registrert tar kontakt om sine rettigheter, har behandlingsansvarlig institusjon plikt til å svare innen en måned.

FØLG DIN INSTITUSJONS RETNINGSLINJER
NSD legger til grunn at behandlingen oppfyller kravene i personvernforordningen om riktighet (art. 5.1 d), integritet og konfidensialitet (art. 5.1. f) og sikkerhet (art. 32).

For å forsikre dere om at kravene oppfylles, må prosjektansvarlig følge interne retningslinjer/rådføre seg med behandlingsansvarlig institusjon.

OPPFØLGING AV PROSJEKTET
NSD vil følge opp planlagt avslutning for å avklare om behandlingen av personopplysningene er avsluttet.

Lykke til med prosjektet!

Kontaktperson hos NSD: Eva J B Payne
Tlf. Personverntjenester: 55 58 21 17 (tast 1)

Appendix H: Consent Form

Vil du delta i forskningsprosjekt om «effekten av insentiver på persepsjon og læring»?

Dette er et spørsmål til deg om å delta i et forskningsprosjekt hvor formålet er å se hvordan insentiver påvirker persepsjon og læring. I dette skrivet gir vi deg informasjon om målene for prosjektet og hva deltakelse vil innebære for deg.

Formål

Studien er en del av et masterprosjekt i psykologi. Formålet med prosjektet er å se hvordan insentiver i form av tap og gevinst av penger påvirker persepsjon og læring. Videre ønsker vi å se om dette har sammenheng mellom forskjeller i måten folk opplever følelser på. I tillegg til en masteroppgave skal resultatene publiseres i en forskningsartikkel. Da vil datasettet bli lagt åpent tilgjengelig for andre forskere som vil ønsker reanalysere dataene. Prosjektet skal etter planen avsluttes sommeren 2019.

Hvem er ansvarlig for forskningsprosjektet?

Professor Robert Biegler ved Institutt for psykologi, NTNU er ansvarlig for prosjektet.

Hva innebærer det for deg å delta?

Hvis du velger å delta i prosjektet, innebærer det å fullføre et eksperiment på en datamaskin. Instrukser vil bli gitt på skjermen. Det vil ta deg omtrent 30 minutter å fullføre. Eksperimentet inneholder perseptuelle oppgaver og læringsoppgaver, samt spørsmål om hvordan du opplever ulike følelser. I tillegg registrerer vi kjønn og alder. Dine svar blir registrert elektronisk.

I eksperimentet vil det være mulig å tjene til seg penger. Det endelige beløpet vil for de fleste ligge mellom 100,- og 200,- kroner, men det er mulig å ende opp med både mindre og mer. Det er også en teoretisk mulighet for å ende opp med et negativt beløp, men vi vil selvfølgelig ikke kreve penger fra deg. Det endelige beløpet vil bestemmes like mye av tilfeldige faktorer som prestasjonen din, og vil komme tydelig frem av instruksene på skjermen. For å anonymisere datamaterialet vil vi kaste terning etter eksperimentet, og legge til antall øyne på det foreløpige beløpet du har endt opp med. Det beløpet vil bli overført til din bankkonto i løpet av noen dager.

Det er frivillig å delta

Det er frivillig å delta i prosjektet. Hvis du velger å delta, kan du når som helst trekke samtykke tilbake uten å oppgi noen grunn. Det vil ikke ha noen negative konsekvenser for deg hvis du ikke vil delta eller velger å trekke deg underveis, men du vil ikke få utbetalt penger.

Ditt personvern – hvordan vi oppbevarer og bruker dine opplysninger

Vi vil bare bruke opplysningene om deg til formålene vi har fortalt om i dette skrivet. Vi behandler opplysningene konfidensielt og i samsvar med personvernregelverket. Så lenge du kan identifiseres i datamaterialet vil kun studenten og veilederen ha tilgang til resultatene fra ditt eksperiment. Dine resultater vil bli anonymisert kort tid etter du har fullført

eksperimentet, og det vil da ikke være mulig å identifisere enkeltpersoner. Samtykkeerklæringen med navn og kontonummer vil oppbevares innelåst, og makuleres etter prosjektet er ferdig.

Banktransaksjonen vil bli loggført i bankens arkiver, og er utenfor vår kontroll.

Dine rettigheter
Så lenge du kan identifiseres i datamaterialet, har du rett til:

- innsyn i hvilke personopplysninger som er registrert om deg,
- å få rettet personopplysninger om deg,
- få slettet personopplysninger om deg,
- få utlevert en kopi av dine personopplysninger (dataportabilitet), og
- å sende klage til personvernombudet eller Datatilsynet om behandlingen av dine personopplysninger.

Hva gir oss rett til å behandle personopplysninger om deg?
Vi behandler opplysninger om deg basert på ditt samtykke. På oppdrag fra institutt for psykologi, NTNU har NSD – Norsk senter for forskningsdata AS vurdert at behandlingen av personopplysninger i dette prosjektet er i samsvar med personvernregelverket.

Hvor kan jeg finne ut mer?
Hvis du har spørsmål til studien, eller ønsker å benytte deg av dine rettigheter, ta kontakt med:

- Institutt for psykologi, NTNU, ved professor Robert Biegler (robert.biegler@ntnu.no)
- Vårt personvernombud: Thomas Helgesen
- NSD – Norsk senter for forskningsdata AS, på epost (personvernombudet@nsd.no) eller telefon: 55 58 21 17.

Med vennlig hilsen

Robert Biegler	Hans Fredrik Sunde
Prosjektansvarlig	Student
(Forsker/veileder)	

Samtykkeerklæring

Jeg har mottatt og forstått informasjon om prosjektet «effekten av insentiver på persepsjon og læring», og har fått anledning til å stille spørsmål. Jeg samtykker til å delta i dette eksperimentet.

Jeg samtykker også til at mine opplysninger behandles frem til prosjektet er avsluttet, ca. sommeren 2019, og forstår at banktransaksjonen vil loggføres utover dette.

Kontonummer (*skriv tydelig!*)

Dato og prosjektdeltakers signatur

Fylles inn av leder for eksperimentet:
Beløp til utbetaling: _____ *Utbetalt?:* ____

References

Adams, W. J., Graf, E. W., & Ernst, M. O. (2004). Experience can change the 'light-from-above' prior. *Nature Neuroscience, 7*(10), 1057-1058. https://doi.org/10.1038/nn1312

Babad, E. (1995). Can accurate knowledge reduce wishful thinking in voters' predictions of election outcomes? *The Journal of Psychology, 129*(3), 285-300. https://doi.org/10.1080/00223980.1995.9914966

Babad, E. (1997). Wishful thinking among voters: Motivational and cognitive influences. *International Journal of Public Opinion Research, 9*(2), 105-125. https://doi.org/10.1093/ijpor/9.2.105

Babad, E., & Katz, Y. (1991). Wishful thinking—against all odds. *Journal of Applied Social Psychology, 21*(23), 1921-1938. https://doi.org/10.1111/j.1559-1816.1991.tb00514.x

Balcetis, E., & Dunning, D. (2006). See what you want to see: motivational influences on visual perception. *Journal of Personality and Social Psychology, 91*(4), 612-625. https://doi.org/10.1037/0022-3514.91.4.612

Bar-Hillel, M., & Budescu, D. (1995). The elusive wishful thinking effect. *Thinking & Reasoning, 1*(1), 71-103. https://doi.org/10.1080/13546789508256906

Barrett, L. F. (2017). *How emotions are made: The secret life of the brain*. London: Macmillan.

Bastardi, A., Uhlmann, E. L., & Ross, L. (2011). Wishful thinking: Belief, desire, and the motivated evaluation of scientific evidence. *Psychological Science, 22*(6), 731-732. https://doi.org/10.1177/0956797611406447

Bateson, P., & Laland, K. N. (2013). Tinbergen's four questions: an appreciation and an update. *Trends in Ecology and Evolution, 28*(12), 712-718. https://doi.org/10.1016/j.tree.2013.09.013

Bischof, D. (2017). New graphic schemes for Stata: plotplain and plottig. *Stata Journal, 17*(3), 748-759. https://doi.org/10.1177%2F1536867X1701700313

Carlson, K. D., & Herdman, A. O. (2010). Understanding the impact of convergent validity on research results. *Organizational Research Methods, 15*(1), 17-32. https://doi.org/10.1177/1094428110392383

Chambers, C. (2017). *The seven deadly sins of psychology: A manifesto for reforming the culture of scientific practice*. Princeton: Princeton University Press.

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences, 36*(3), 181-204. https://doi.org/10.1017/S0140525X12000477

Cosmides, L. (1989). The logic of social exchange: Has natural selection shaped how humans reason? Studies with the Wason selection task. *Cognition, 31*(3), 187-276. https://doi.org/10.1016/0010-0277(89)90023-1

Damasio, A. (1994). *Descartes' error: Emotion, reason, and the human brain*. London: Vintage.

Delavande, A., & Manski, C. F. (2012). Candidate preferences and expectations of election outcomes. *Proceedings of the National Academy of Sciences, 109*(10), 3711-3715. https://doi.org/10.1073/pnas.1200861109

Drummond, C., & Fischhoff, B. (2017). Individuals with greater science literacy and education have more polarized beliefs on controversial science topics. *Proceedings of the National Academy of Sciences, 114*(36), 9587-9592. https://doi.org/10.1073/pnas.1704882114

Dutton, D. G., & Aron, A. P. (1974). Some evidence for heightened sexual attraction under conditions of high anxiety. *Journal of Personality and Social Psychology, 30*(4), 510-517. https://doi.org/10.1037/h0037031

Gershman, S. J. (2019). How to never be wrong. *Psychonomic Bulletin Review, 26*(1), 13-28. https://doi.org/10.3758/s13423-018-1488-8

Gignac, G. E., & Szodorai, E. T. (2016). Effect size guidelines for individual differences researchers. *Personality and Individual Differences, 102*, 74-78. https://doi.org/10.1016/j.paid.2016.06.069

Gilovich, T. (1991). *How we know what isn't so: The fallibility of human reason in everyday life*. New York: Free Press.

Gold, J. M., Waltz, J. A., Matveeva, T. M., Kasanova, Z., Strauss, G. P., Herbener, E. S., . . . Frank, M. J. (2012). Negative symptoms and the failure to represent the expected reward value of actions. *Archives of General Psychiatry, 69*(2), 129-138. https://doi.org/10.1001/archgenpsychiatry.2011.1269

Grafton, B., Ang, C., & MacLeod, C. (2012). Always look on the bright side of life: The attentional basis of positive affectivity. *European Journal of Personality, 26*(2), 133-144. https://doi.org/10.1002/per.1842

Haidt, J. (2012). *The righteous mind: Why good people are divided by politics and religion*. London: Penguin Books.

Hawkins, J., & Blakeslee, S. (2004). *On intelligence*. New York: St. Martins Press.

Hayes Jr, S. P. (1936). The predictive ability of voters. *The Journal of Social Psychology, 7*, 183-191. https://doi.org/10.1080/00224545.1936.9921660

Hertwig, R., & Ortmann, A. (2001). Experimental practices in economics: A methodological challenge for psychologists? *Behavioral and Brain Sciences, 24*(3), 383-403.

Hohwy, J. (2017). Priors in perception: Top-down modulation, Bayesian perceptual learning rate, and prediction error minimization. *Consciousness and Cognition, 47*, 75-85. https://doi.org/10.1016/j.concog.2016.09.004

Johnson, E. J., & Tversky, A. (1983). Affect, generalization, and the perception of risk. *Journal of Personality and Social Psychology, 45*(1), 20-31. https://doi.org/10.1037/0022-3514.45.1.20

Johnson, M. W., & Bickel, W. K. (2002). Within-subject comparison of real and hypothetical money rewards in delay discounting. *Journal of the Experimental Analysis of Behavior, 77*(2), 129-146. https://doi.org/10.1901/jeab.2002.77-129

Kahan, D. M. (2013). Ideology, motivated reasoning, and cognitive reflection. *Judgment and Decision Making, 8*(4), 407-424. https://doi.org/10.2139/ssrn.2182588

Kahan, D. M., Peters, E., Dawson, E. C., & Slovic, P. (2013). Motivated numeracy and enlightened self-government. *Behavioural Public Policy, 1*(1), 54-86. https://doi.org/10.1017/bpp.2016.2

Kahan, D. M., Peters, E., Wittlin, M., Slovic, P., Ouellette, L. L., Braman, D., & Mandel, G. (2012). The polarizing impact of science literacy and numeracy on perceived climate change risks. *Nature Climate Change, 2*, 732-735. https://doi.org/10.1038/nclimate1547

Karpicke, J. D., & Roediger III, H. L. (2007). Expanding retrieval practice promotes short-term retention, but equally spaced retrieval enhances long-term retention. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33*(4), 704-719. https://doi,.org/10.1037/0278-7393.33.4.704

Kashyap, A., & Sirola, V. S. (2018). The Duhem-Quine problem for equiprobable conjuncts. *Studies in History and Philosophy of Science, In Press*. https://doi.org/10.1016/j.shpsa.2018.09.002

Krizan, Z., Miller, J. C., & Johar, O. (2009). Wishful thinking in the 2008 U.S. presidential election. *Psychological Science, 21*(1), 140-146. https://doi.org/10.1177/0956797609356421

Krizan, Z., & Windschitl, P. D. (2007). The influence of outcome desirability on optimism. *Psychological Bulletin, 133*(1), 95-121. https://doi.org/10.1037/0033-2909.133.1.95

Kühberger, A., Schulte-Mecklenbeck, M., & Perner, J. (2002). Framing decisions: Hypothetical and real. *Organizational Behavior and Human Decision Processes, 89*(2), 1162-1175. https://doi.org/10.1016/S0749-5978(02)00021-3

Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin, 108*(3), 480-498. https://doi.org/10.1037/0033-2909.108.3.480

Kurzban, R. (2011). *Why everyone (else) is a hypocrite: Evolution and the modular mind*. Princeton: Princeton University Press.

Lakens, D. (2017). Equivalence tests: A practical primer for t tests, correlations, and meta-analyses. *Social Psychological and Personality Science, 8*(4), 355-362. https://doi.org/10.1177/1948550617697177

Larsen, R. J., & Diener, E. (1987). Affect intensity as an individual difference characteristic: A review. *Journal of Research in Personality, 21*(1), 1-39. https://doi.org/10.1016/0092-6566(87)90023-7

Locey, M. L., Jones, B. A., & Rachlin, H. (2011). Real and hypothetical rewards. *Judgment and Decision Making, 6*(6), 552-564.

Lord, C. G., Ross, L., & Lepper, M. R. (1979). Biased assimilation and attitude polarization: The effects of prior theories on subsequently considered evidence. *Journal of Personality and Social Psychology, 37*(11), 2098-2109. https://doi.org/10.1037/0022-3514.37.11.2098

Lynn, S. K., & Barrett, L. F. (2014). "Utilizing" signal detection theory. *Psychological Science, 25*(9), 1663-1673. https://doi.org/10.1177/0956797614541991

Macmillan, N. A., & Creelman, C. D. (1991). *Detection theory: A user's guide*. Cambridge: Cambridge University Press.

Marks, R. W. (1951). The effect of probability, desirability, and “privilege” on the stated expectations of children. *Journal of Personality, 19*(3), 332-351. https://doi.org/10.1111/j.1467-6494.1951.tb01107.x

Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information*. Cambridge: The MIT Press.

Mercier, H., & Sperber, D. (2011). Why do humans reason? Arguments for an argumentative theory. *Behavioral and Brain Sciences, 34*(2), 57-74. https://doi.org/10.1017/S0140525X10000968

Mercier, H., & Sperber, D. (2017). *The enigma of reason: A new theory of human understanding*. London: Allan Lane.

Muthukrishna, M., & Henrich, J. (2019). A problem in theory. *Nature Human Behaviour, 3*(3), 221-229. https://doi.org/10.1038/s41562-018-0522-1

Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review, 84*(3), 231-259. https://doi.org/10.1037/0033-295X.84.3.231

Norges Bank. (n.d.). Exchange rate for USD. Retrieved from https://www.norges-bank.no/en/Statistics/exchange_rates/currency/USD/

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science, 349*(6251), aac4716. https://doi.org/10.1126/science.aac4716

Pierce, J. W., & MacAskill, M. R. (2018). *Building experiments in PsychoPy*. London: SAGE Publications Ltd.

Quintana, D. S., & Heathers, J. A. (2014). Considerations in the assessment of heart rate variability in biobehavioral research. *Frontiers in Psychology, 5*, 805. https://doi.org/10.3389/fpsyg.2014.00805

Ripper, C. A., Boyes, M. E., Clarke, P. J. F., & Hasking, P. A. (2018). Emotional reactivity, intensity, and perseveration: Independent dimensions of trait affect and associations with depression, anxiety, and stress symptoms. *Personality and Individual Differences, 121*, 93-99. https://doi.org/10.1016/j.paid.2017.09.032

Rubin, D. C., Hoyle, R. H., & Leary, M. R. (2012). Differential predictability of four dimensions of affect intensity. *Cognition & Emotion, 26*(1), 25-41. https://doi.org/10.1080/02699931.2011.561564

Schwarz, N. (2012). Feelings-as-information theory. In P. A. M. V. Lange, A. W. Kruglanski, & E. T. Higgins (Eds.), *Handbook of Theories of Social Psychology* (Vol. 1, pp. 289-308). London: SAGE Publications Ltd.

Shepperd, J. A., Findley-Klein, C., Kwavnick, K. D., Walker, D., & Perez, S. (2000). Bracing for loss. *Journal of Personality and Social Psychology, 78*(4), 620-634. https://doi.org/10.1037/0022-3514.78.4.620

Simler, K., & Hanson, R. (2018). *The elephant in the brain: Hidden motives in everyday life*. New York: Oxford University Press.

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science, 22*(11), 1359-1366. https://doi.org/10.1177/0956797611417632

Smith, V. L., & Walker, J. M. (1993). Monetary rewards and decision cost in experimental economics. *Economic Inquiry, 31*(2), 245-261. https://doi.org/10.1111/j.1465-7295.1993.tb00881.x

Sunde, H. F., & Biegler, R. (2019). *Motivated perception and arousability [Registered report]*. Retrieved from https://doi.org/10.17605/OSF.IO/QPTYB

Svenson, O. (1981). Are we all less risky and more skillful than our fellow drivers? *Acta Psychologica, 47*(2), 143-148. https://doi.org/10.1016/0001-6918(81)90005-6

Taber, C. S., Cann, D., & Kucsova, S. (2009). The motivated processing of political arguments. *Political Behavior, 31*(2), 137-155. https://doi.org/10.1007/s11109-008-9075-8

Taylor, S. E., & Brown, J. D. (1988). Illusion and well-being: A social psychological perspective on mental health. *Psychological Bulletin*, 193-210. https://doi.org/10.1037/0033-2909.103.2.193

Tetlock, P. E. (1985). Accountability: A social check on the fundamental attribution error. *Social Psychology Quarterly, 48*(3), 227-236. https://doi.org/10.2307/3033683

Tinbergen, N. (1963). On aims and methods of Ethology. *Zeitschrift für Tierpsychologie, 20*(4), 410-433. https://doi.org/10.1111/j.1439-0310.1963.tb01161.x

Tooby, J., & Cosmides, L. (1992). The psychological foundations of culture. In J. Tooby, L. Cosmides, & J. H. Barkow (Eds.), *The adapted mind: Evolutionary psychology and the generation of culture* (pp. 19-136). Oxford: Oxford University Press.

Trivers, R. L. (2011). *Deceit and self-deception: Fooling yourself the better to fool others*. London: Penguin Books.

Tversky, A., & Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty, 5*(4), 297-323. https://doi.org/10.1007/bf00122574

Villsvinjeger tiltalt for drapsforsøk. (2019, March 22). *NRK*. Retrieved from https://www.nrk.no/buskerud/villsvinjeger-tiltalt-for-drapsforsok-1.14453421

Vosgerau, J. (2010). How prevalent is wishful thinking? Misattribution of arousal causes optimism and pessimism in subjective probabilities. *Journal of Experimental Psychology: General, 139*(1), 32-48. https://doi.org/10.1037/a0018144

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: the PANAS scales. *Journal of Personality and Social Psychology, 54*(6), 1063-1070. https://doi.org/10.1037/0022-3514.54.6.1063

Weber, E. U. (1994). From subjective probabilities to decision weights: The effect of asymmetric loss functions on the evaluation of uncertain outcomes and events. *Psychological Bulletin, 115*(2), 228-242. https://doi.org/10.1037/0033-2909.115.2.228

Windschitl, P. D., Smith, A. R., Rose, J. P., & Krizan, Z. (2010). The desirability bias in predictions: Going optimistic without leaving realism. *Organizational Behavior and Human Decision Processes, 111*(1), 33-47. https://doi.org/10.1016/j.obhdp.2009.08.003

Zillmann, D. (1971). Excitation transfer in communication-mediated aggressive behavior. *Journal of Experimental Social Psychology, 7*(4), 419-434. https://doi.org/10.1016/0022-1031(71)90075-8

NTNU
Norwegian University of
Science and Technology