Utrecht University

Interpreting Classification Images of the Self in the Context of Predictive Coding

---

Désirée Lawson

Esmée Verburgt

Lana Hoekstra

Research Project in Social Neuroscience

Supervised by Loek Brinkman, Department of Psychology

Utrecht University, Utrecht 2019

Abstract

According to the predictive coding theory, the perception of the environment, others and ourselves is influenced by prior knowledge, also referred to as predictions. The objective sensory input is combined with these predictions into a weighted average, which forms our eventual subjective perception. These different internalized predictions explain interindividual differences within perception. The reverse correlation technique strives to visualize this implicit prediction also known as “prior” within an image. Within the first part of our research we strived to validate whether this technique is a valid tool to visualize the implicit self-image through a recognition task, which might eventually be used within therapeutic settings. Results show that participants can significantly recognize their implicit self-image, which eventually indicates that reverse correlation visualizes self-image. The second part of our research investigated whether this visualization represents the prior within the predictive coding theory. Our results suggest that this is indeed the case. Ultimately, we propose some adjustments within our research to even further validate this notion, but consider reverse correlation to be a valuable method to do research within the theory of predictive coding.

The perception that we have of the world, the people around us and ourselves is based on our beliefs, opinions, memories and knowledge that we have built up in the course of our lives. How we perceive the things that we encounter and how we react to them, is influenced by previous experiences and the knowledge that we gained through these experiences. This knowledge has an influence on how we perceive ourselves as well. To apprehend how the perception of ourselves and others is constituted, it is important to know how our predictions of the world influence this perception. One of the theories that provide a framework for the process in which individuals perceive the environment is the predictive coding theory.

The predictive coding theory provides a general mechanism to account for action, perception and learning, based on predictions, input and error-signals (Friston, 2010). The theory states that all self-organizing systems try to minimize surprise. We lessen surprise, or the error-signal, by having an accurate a priori prediction of the upcoming visual input. This could be done by either adjusting the internal prediction that we hold, or we could change the visual input by selectively focussing on the input that we predicted (Friston, 2010). One of the main assumptions within the predictive coding theory is that we hold these internalized predictions, also called 'priors', and that they influence our eventual perception.

The theory suggests that our perception works according to mechanisms that continuously anticipate on what might happen in the foreseeable future. An example is that people are able to infer characteristics and expected behaviour within the social domain from faces, such as traits, even when these are not observable. This is because people match the objective input that they receive to the mental representation that they already have (Brinkman, Todorov & Dotsch, 2017). When our prior does not match with the sensory input, a prediction error occurs between them. This error is believed to be forwarded through the brain, while backward connections holding the prediction try to minimize this error. Thus the prediction error could be considered as the difference between our prior and the sensory input, or the amount of surprise. Minimizing surprisal is akin to minimizing the prediction error.

The predictive coding theory can be explained through a Bayesian description since it combines our prior knowledge, predictions, and the sensory information into a combined subjective perception. Bayesian statistics could be used to calculate how the two sources of

information can be integrated. Our general model of the world is optimized through sensory inputs by calculating a weighted average. The weighted average is a combination of the input and the prior, and is called the 'posterior'. The average is weighted because it reflects the importance and strength of our prior and the visual input. Within figure 1, an example of a Bayesian probability distribution with a prediction error and a statistical weighted average is shown. The height within the graph shows us the strength of either the prior or the input. The strength of the prior could be influenced by our experience, or lack of it, and the certainty of our input could be compromised by for example darkness. When we for example experience something for the first time, we have no existing prior. This causes the average to be pulled towards the sensory input. It is hypothesized that our perception works through conceptualizing new inferences until our prediction is in line with our sensory input. Our prior beliefs about the world will be updated since the error is fed forward within our system (Clark, 2012).

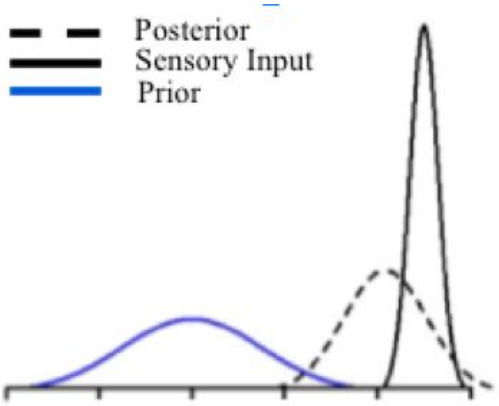


Figure 1. The Bayesian distribution. This distribution represents the certainty of the prior, posterior and the visual input. The prior is the prediction of what we expect to see. When we have high predictions, the blue line is high and narrow. When we are not sure of our predictions, the blue line is low and wide. The sensory input is the visual environment that there actually is. In the dark, this visual input is less certain than in day light. A combination of the two, the posterior, is what we actually perceive. When our prior is strong, we perceive what we expect to perceive. When the prior is uncertain, we are less driven by our expectations of the environment.

Ultimately, the prior highly influences how we perceive the social environment. Since it is also generated from individual experiences, these priors account for interindividual differences in how we eventually perceive the sensory input. This again explains why people tend to behave distinctly upon sensory input. Therefore it is of great importance to apprehend these individual priors to shed light on subjective perception of other individuals and of ourselves.

Recently, a new technique was developed that is presumed to provide us a visualization of these individual subjective predictions. The reverse correlation technique aims to reflect the content of our priors within an image. This image determines which facial characteristics are implicitly used to define our internalized prediction, which might reflect the prior (Brinkman, Todorov & Dotsch, 2017). Participants who perform the reverse correlation task are forced to choose between two pictures. These pictures consist of a base face with a random grey noise overlay. Participants are asked to determine which one of the faces best represents the variable or prediction, that the researcher is interested in, for example traits or etnic characteristics (Dotsch & Todorov, 2012; Dotsch, Wigboldus, Langner & Knippenberg, 2008). By averaging the pictures of faces, a classification image (CI) can be created. This classification image visualizes the facial characteristics that people focus on when they think of the characteristics that were tested for. For instance, Dotsch and Todorov (2012) found that the mouth, eyes and hair regions were important characteristics when participants had to identify trustworthy and untrustworthy faces. Additionally, reverse correlation has been previously applied in studying in-group projections (Imhoff, Dotsch, Bianchi, Banse & Wigboldus 2011), romantic partners (Karremans, Dotsch & Corneille 2011), bodies (Lick, Carpinella, Preciado, Spunt & Johnson, 2013) and presidential candidates (Young, Ratner & Fazio 2014).

The most recent development in using reverse correlation regards creating a mental image of the self. This self representation is among other things, influenced by self esteem, and has an effect on to which extent the CI resembles the subject (Shorten, 2017). Brinkman and Kennis (2017) used reverse correlation to visualize the implicit actual, and ideal self-image of participants. When participants are instructed to continuously pick the image within the RC task that represents their actual self-image, a CI could be computed that represents the unconscious image that participants have of themselves. Since the classification image conveys information about for example ethnicity, gender or traits, it could also enclose important information about someone's physiological and psychological characteristics. This could be applied in therapeutic settings by confronting people with the mental image they have of themselves in comparison to how they would like to look or how other people perceive them.

Brinkman and Kennis previously performed a study in which they aimed at validating reverse correlation regarding the self-image. They validated reverse correlation by producing the actual- and ideal self-image of the participants and compared the two CIs to visualize the

discrepancy. However, before these CIs can have therapeutic implications for people with a distorted self-image, it is necessary that participants can identify themselves correctly within their CI, as they must perceive their CI to truly reflect themselves. If this is not the case, the positive clinical implications that the image might have, can only be regarded as a placebo. Then when confronted with their actual self-image, participants might change the mental image that they have of themselves. Therefore, participants were asked to perform a recognition task in which they had to identify their own computed CI among five other random CIs. Brinkman and Kennis found that in 60% of the cases, participants are able to correctly identify themselves within the first three guesses. This study is replicated by Brinkman, Verspui and Yilmaz (2018), who found that 87% of the participants are able to correctly identify their own CI within the first three guesses. This result is very high considering that the authors accounted for gender differences within the recognition task by presenting five other random CIs of the same gender. This makes the recognition task even harder since participants are not able to exclude CIs based on the opposite gender from the task.

To validate the notion that participants are able to recognize their own classification image above average even further, the first part of our research will therefore focus on replicating the results of the recognition task as done by Brinkman and Kennis. Because the original study by Brinkman and Kennis resulted in a great effect, we expect that our participants are able to recognize themselves above the expected 1/6th chance. If these results can be replicated, we can state that reverse correlation is a valid and reliable method that can be used to capture the self-image of individuals, and possibly to confront patients that struggle with own self-image, such as anorexia nervosa (Esposito, Cieri, di Giannantonio & Tartaro, 2018), with their actual self-image. This self-image is supposed to represent the prior in the predictive coding theory. However, the theory that the visualization of the CI reflects the content of the prior is a mere theoretical assumption and not yet empirically tested (Brinkman, Todorov & Dotsch, 2017. Within the second part of our research, we set out to empirically test the assumption that a classification image represents the prior and that the predictive coding framework is applicable in the interpretation of the CIs.

We suggest that when subjects create their own CI exclusively from memory, it will differ from when they are simultaneously offered the actual visual input of their own face by means of a mirror. This could imply that if the CIs are distinguishable, the CI purely created from memory

could entails the actual prior. To test this assumption, the second part of this experiment will consist of three reverse correlation tasks. The first reverse correlation task will be performed as in Brinkman and Kennis's (2017) experiment where participants have to choose the image that best represent themselves. The CI that follows is assumed to reflect the implicit self-image, that is the prior one holds in their mind's eye. For the second reverse correlation task, participants simultaneously look in a mirror, where they are asked to choose the face that best represents their subjective self-image. The combination of the implicit self-image and the sensory input as perceived in the mirror, is believed to reflect the posterior. The last reverse correlation task will be a CI made by an objective researcher from a photograph of the participant. We assume this CI to be a representation of the objective visual input, since the objective researcher holds minimal prior information about the participant. Eventually, our supervisor, Loek Brinkman, should be able to label the three CIs in the right order as shown in figure 2. In summery, if our hypothesis is true, then we are able to distinguish the 3 CIs: the self CI will represent the prior, the mirror CI will represent the posterior, and the objective CI will be likely to represent the visual input. Eventually, we can conclude that the CI from memory in reverse correlation conveys the prior.

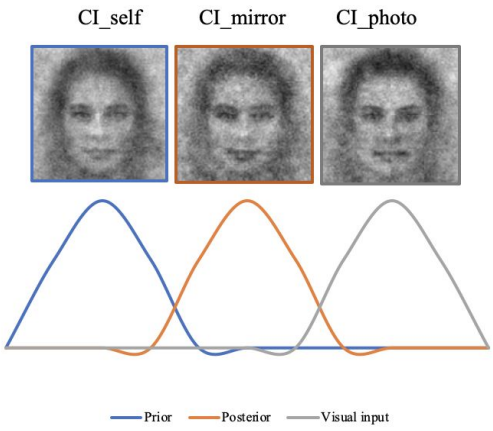


Figure 2. Hypothesis, the sorting task explained in a Bayesian distribution. The Bayesian distribution represent the probability of the prior, posterior and visual input. The CIs that follow from the first reverse correlation task each represent a different line in the distribution. The CI made from the first reverse correlation task, CI_self, represents the prior. The CI made by the objective researcher from photograph, CI_photo, represents the visual input. The CI made by the participant while looking in the mirror represents a combination of CI_self and CI_photo and thus is the posterior. Note that the weight of the probability of the prior, posterior and objective representation do not amount to exact numbers, as this distribution figure is a means of illustrating and an assessment of the combination of the self and photo CI to form the mirror CI.

.

Additionally, we perform two exploratory analyses and one control analysis. It might be interesting to analyse if participants are able to recognize the assumed posterior and visual input CI in the two remaining recognition tasks. We hypothesize that participants are able to recognize themselves within the prior CI, because this is the image that should theoretically represent their mental self-image, which conveys how they think they look themselves. Since the classification images that represent the posterior and the visual input convey the mental self-image to a lesser degree, we expect that participants are worse in recognizing themselves in these CIs. Secondly, we perform a control analysis to determine whether the objective CI really represents the photograph. We expect that the objective CI does indeed represent the photograph and that Loek Brinkman will be able to label the objective CI correctly above chance level. Lastly, we will perform an analysis of the sorting task which only includes the participants that correctly identify their prior CI within the first trial. We will do so because incorrect recognition may imply that participants did not perform the reverse correlation task well, and that their potential prior and possibly the mirror CI do not reflect them correctly, which will make it hard for Loek Brinkman to lay them in the right order.

Method

Participants

The recruitment of participants has happened by means of social communication, Facebook and the University of Utrecht. The amount of needed participants for the recognition task was calculated with the aid of G*Power 3.1. Because we conduct two different types of experiments, we also performed two power analyses. The power analysis that we conducted following the replication of Brinkman, Yilmaz and Verspui, resulted in a number of 22 participants (w=0.96, Power =0.95, α =0.05). The effect size and power are determined from the results that they have found. For the second part of our experiment we performed a power analysis which resulted in a number of 39 participants (w=0.5, Power =0.80, α=0.05). Since this power analysis resulted in a greater number of needed participants we decided to recruit this amount of participants for our data collection.

Ultimately, we chose to base the effect size of the experiment as a whole on the part that investigates whether the CI reflects the prior. There is no previous research, nor are there any pilot

studies performed on this subject matter, hence no estimation of the effect size is available. Even though it is preferable to specify the a priori effect size on previous research, we partly have to rely on conventional standards. Additionally, Brinkman and Kennis (2017) found that when participants are confronted with their own implicit CI reflecting the assumed prior, it can differ to a great extent to how participants consciously perceive themselves. Which shows us that it could be very different from the posterior self-presentation. Furthermore, we expect that when the participants correctly recognize their own prior, it does not always resemble the objective visual image of the participant that the researchers perceive. So eventually the three CIs are assumed to differ enough to distinguish them by the face of it. Therefore, we expect a high effect size and effects observable for the naked eye. A value of w=0.5 is suggested as a 'high' effect size (Cohen, 1992), which will be chosen as the a priori effect size for our research. In total, 41 participants were recruited. 26 participants identify as women and 15 identified themselves as men. The recruited participants have a mean age of 24.35 (*SD*= 9.11).

Stimuli and materials

For the reverse correlation task, a neutral male base face from the Karolinska Institutet face database (Lundqvist, Flykt & Ohman, 1998) is used. The images were produced using a grey noise overlay over the base face as seen in figure 3, which resulted in a variety of 1000 different stimuli. The images were conducted from previous research from Dotsch (2016), which were built using RStudio within the RCICR-package in R.

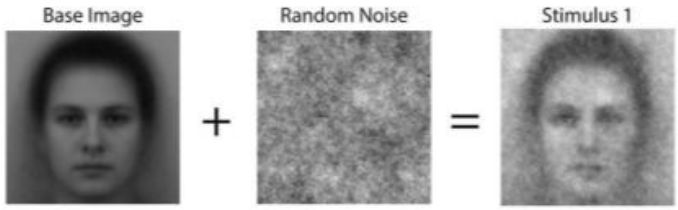


Figure 3. Construction of stimuli that is used in the reverse correlation task (Brinkman, Todorov & Dotsch, 2017). Every stimuli is computed by taking the base image and putting a random noise grey overlay on top of it.

The reverse correlation task and the recognition task are build within Gorilla (2018). The stimuli in each trial of the reverse correlation task comprised of twice the base face image, imposed with an opposite grey noise overlay. The same stimuli were used for the two reverse correlation tasks, with and without the mirror. The mirror, 30 cm by 30 cm, for the second reverse correlation task

is applied right above the computer screen, so that the participants were able to see their face completely.

The recognition task is likewise build in Gorilla. The participants receive three recognition tasks. The first one is the recognition task of the CI that concerns the prior, the second recognition regards the CI they made with the mirror and the third recognition task relates to the CI that we, as independent researchers, created from a photograph of the participant. These three recognition tasks are built so that they directly follow one and another. The participants receive a link through their mail, approximately one to two weeks after completing the reverse correlation task, in which they are instructed to try to recognize which CI out of six is their own. The five additional CIs are CIs made by other participants from the same gender and were randomly chosen.

Procedure

We conducted a within design research, in which a participant creates a CI exclusively from their mind, and subsequently an additional CI in which they simultaneously perceive visual input of their own face by means of a mirror. Furthermore, and objective researcher, one that does not know the participant, also creates a CI of the participant using a photograph of the participants face, which is taken beforehand. Before the experiment starts, we asked the participants for their permission to take their photo and to sign an informed consent that states that the photos are kept private and deleted after 8 weeks. The photo of the face of the participant was taken approximately 20 centimeters from the participant's face. The participant is asked to stand still in front of a white background, and look straight into the camera with a neutral face.

After the photo is taken, the participant is asked to take place in front of the computer. The task starts with an informed consent about the experiment, where the participants are informed about the experiment and their rights. Participants that study Psychology at Utrecht University can receive a compensation of 2 PPU. After agreeing to the informed consent, several demographic details such as age, gender, and educational level are asked. Next, participants receive a detailed description on how to perform the reverse correlation task, and will carry out a test trial. They are instructed to choose the image which best represents the actual self, on either a psychological and physical level. The participants receive five blocks of 100 trials with a short break in between the trials. The CI that follows from this part of the experiment will represent the implicit mental image.

The second part of the experiment consists of the same instructions as described above. However, a mirror is placed above the computer screen. Participants are asked to choose the image that best represents the image they see in the mirror, on both a psychological and physical level. When the participants leave, we perform a reverse correlation task for the participants based on the photo we took in the beginning of the experiment. To minimize bias, we made sure that the researcher performing the task, does not know the participant. The objective researcher will perform the reverse correlation task while continuously choosing the picture which best represents the photograph of the participant.

The last part of the experiment consists of the recognition task. This will be sent via email within two weeks after the reverse correlation task is completed, and can be done at home. The participants receive a link which directs them to task. In this recognition task, they are presented with six CIs. They receive the instruction to pick which CI they think is the CI that they created during the reverse correlation task without the mirror. When they pick incorrectly, they can pick again until the right CI is chosen. An illustration of this task is depicted in figure 4. Additionally, participants receive two more recognition tasks. The second recognition task will be of the CI that they made while looking in the mirror and the third recognition task will be of the CI that we made based on the photograph.

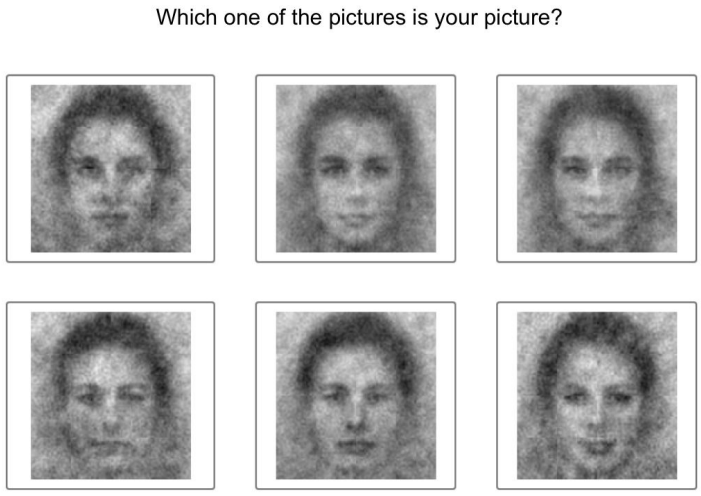


Figure 4. Example of the recognition task. One CI is the CI that is made by the participants themselves, were the five extra CIs are CIs that are made by other participants of the same gender. The participants receive the instruction to recognize the face that they constructed themselves. They can click on the pictures until they choose the right one.

When all the data is collected, the classification images that are created will be subjectively rated by our supervisor, Loek Brinkman, whom is asked to determine the right order of the classification images with reference to the photograph. If our hypothesis is right, he should be able to conclude that the mirror-CI is a combination of the CI without a mirror and the objective CI as created by us. The analysis of the results is described below.

Analyses

Firstly, all the raw data from the reverse correlation task is screened individually for every participant to see if any outliers are present. Outliers are participants who respond more than two times faster than the standard deviation. Secondly, if a left/right bias is present, participants are removed. A left/right bias is when a participant chooses the left or right picture more than 100 times in a row. The outliers are removed to prevent the data from being influenced by participants who don't complete the task accordingly because of boredom or disinterest. This exclusion criteria resulted in the removal of one participant, which leaves the data of 40 participants for the analysis.

Analysis 1; Recognition task

Subsequently, a Chi-square test for goodness of fit is performed to analyse if the participants are able to recognize the CIs that they created among six other CIs above chance level. This test is performed with IBM SPSS. An effect would eventually be considered present if participants are able to recognize their own CI in the first trial above a ⅙ chance level. Considering 40 participants, the expected distribution based on chance would be 16,7% correct guesses for each trial, which translates to the exact value of 6,67 participants that will correctly pick their CI in each trial. We compared this expected frequency distribution with the observed frequency we ultimately found and eventually tested the effect for significance. Additionally, the effect size is calculated via G*Power 3.1.

Exploratory analysis 1

In total, three CIs are created. For the recognition task we will use the CI made from memory, assumed to be the prior. This will be the CI that is used in therapeutic applications. Additionally, we created two more recognition tasks concerning the posterior and objective CI. A Chi-square

goodness of fit analyses will be performed within IBM SPSS over the two remaining recognition tasks and the expected frequency distribution will be compared with the observed frequency distribution. The recognition rates from the three different classification images will be compared with each other by checking for differences in the p-values and effect sizes.

Analysis 2; Sorting task

With a chi-square test for goodness of fit we tested whether the prior, posterior and objective visual input are recognizable, performed via IBM SPSS. Loek Brinkman, who is an expert in the field of predictive coding, was asked to label the three different classification images of each participant as either "prior", "posterior", or "objective visual input". In order to do so, he is given the three CIs plus a photograph of the participant in each trial. The researcher should be able to label the three images in the right order because we expect the posterior, or the CI created with the mirror, to better resemble the objective CI that reflects the objective visual input. This CI is expected to be a combination of the prior and the objective CI. In general, successfully labeling the three CIs should be above a chance level of ⅙.

Control analysis

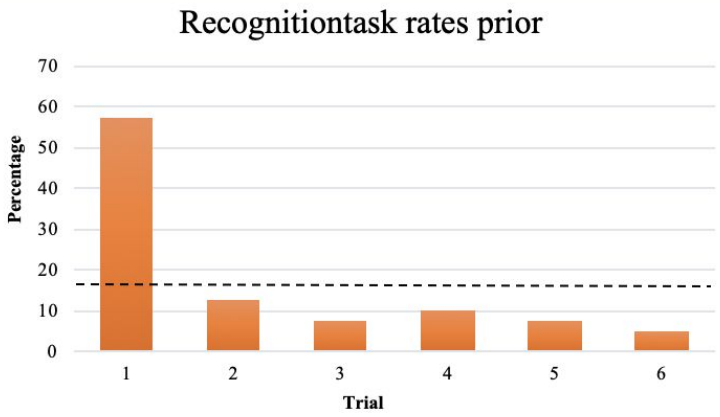
Because it is not yet clear if the CI we made from the photograph is not influenced by our subjective beliefs, and does eventually accurately represents the participant, we perform an additional control analysis. If the assumption that an objective classification image could be created from a photograph is true, Loek Brinkman should be able to label the photo objective CI as the 'objective visual input' above average, following the photograph. Since the task provides the researcher with three possible options, this should be above a ⅓ chance level. This too is performed through a chi-square goodness of fit in IBM SPSS.

Exploratory analysis 2

Since for therapeutic implications it is of great significance that subjects are able to recognize their own CI as themselves, we additionally only analyse the data from the sorting task of the participants that were able to recognize their own mental image via IBM SPSS. This analysis will again be a chi-square test for goodness of fit in which we exclude participants that did not recognize their self-image in the first try.

Analysis 1; Recognition task

To determine whether the participants recognized their self-image as conveyed within the CI from memory, we conducted a Chi-square goodness of fit test. The recognition rate for first trial was higher than expected, whereas the recognition rates for the remainder of the trials were lower than expected. The results show that participants recognize their self-image significantly from chance ($\chi 2$ (5, N = 40) = 48.8, p < .01., w = 1.10). Graph 1 shows the recognition rates for each trial. Within the first trial, 57.5% of the participants recognized their CI. 77.5% of the participants recognized their self-image within the first three trials.

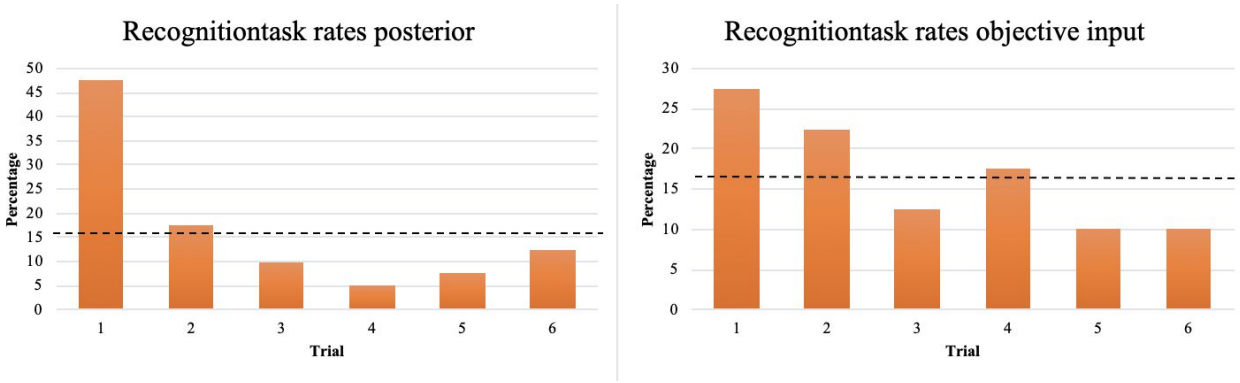


Graph 1. The recognition rates of the recognition task. The recognition rates shows the percentages of the participants that recognize their CI in a certain trial. The dotted line represents the percentages when the data will be equally distributed in a 1/6th chance, which is at 16,7%. Participants were able to recognize their own CI among 5 random CIs above chance only in the first trial, which is 57.5%.

Exploratory analysis 1

For the exploratory analysis we used the Chi-square goodness of fit test to conclude whether the recognition rates differ from chance level for the other two classification images. Graph 2 shows the recognition rates for each trial with regard to the different CIs. The results show that participants are able to recognize the CI they made while looking in the mirror significantly above

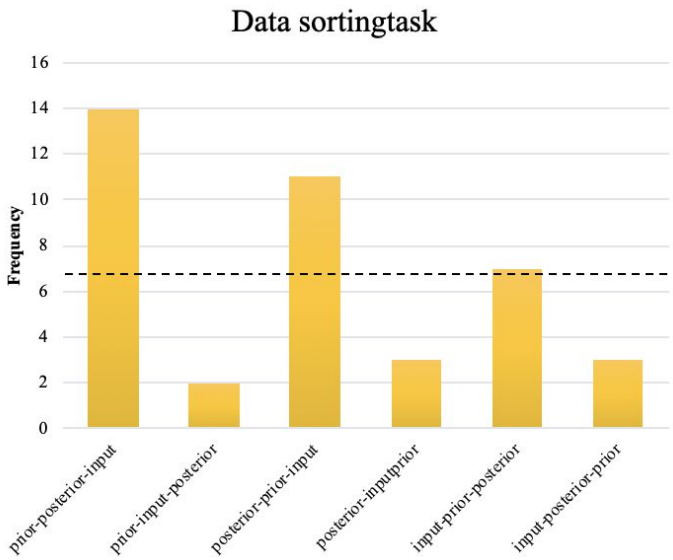
chance (χ2 (5, N = 40) = 29,6, p < .01., w = .86). Within the first trial, 47.5% of the participants recognized their CI, while 75% can recognize their CI within the first three trials. This is slightly lower than the recognition rates for the implicit self-image. The Chi-square goodness of fit test did not show significant results for the recognition rates regarding the objective CI made from the photograph (χ2 (5, N = 40) = 6,2, p = .25). Only 27.5% were able to recognize themselves in the first trial and 50% were able to recognize themselves within the first three trials.



Graph 2. The graphs show the recognition rates of the recognition task with the hypothesized posterior CI and objective CI. The recognition rates shows the percentages of the participants that recognize their CI in a certain trial. The dotted lines represents the percentages when the data will be equally distributed in a 1/6th chance, which is at 16,7%. Participants were able to recognize their posterior CI above the expected percentages in the first and second trials, which is 65%. Participants were able to recognize their objective CI around chance level and do not differ significantly.

Analysis 2; Sorting task

For the analysis of the sorting task we likewise conducted a Chi-square goodness of fit test. The test was used to determine whether our supervisor, Loek Brinkman, was able to arrange the three CIs made by the participant and an objective researcher, in the right order according to the predictive coding theory. The results show that the order in which Loek Brinkman labeled the three CIs deviates significantly from chance (χ2 (5, N = 40) = 18.20, p < .01., w = .67). As shown in graph 3, two out of six options are chosen significantly above chance, from which the first option is the right order; prior - posterior - visual input. This option is chosen 35% of the time. The third option is chosen in 27.5% of the cases and conveys the following order: posterior - prior - visual input. The fifth option is chosen 17.5% of the times, which is at chance level. Each of the remainders of the options is chosen in 7.5% of the cases or less.



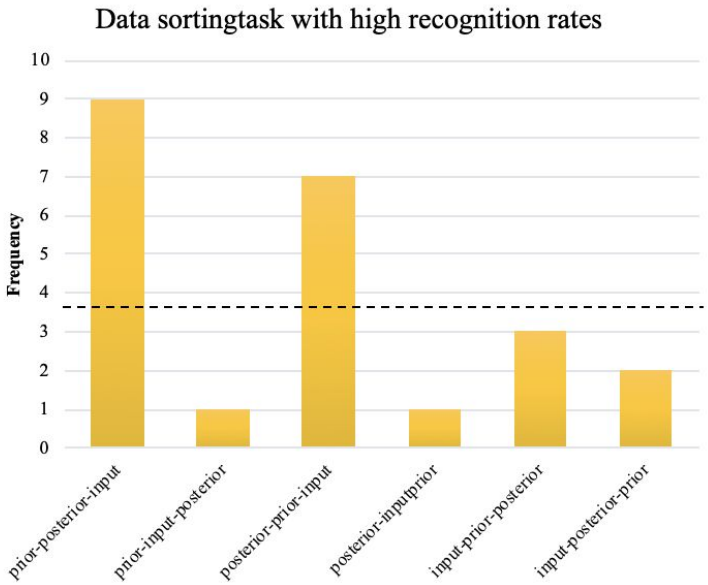
Graph 3. The sorting task in which Loek Brinkman had to label the three CIs in the correct order. The graph shows the observed frequency of the times the researcher labelled the CIs in that certain order. The dotted line represents the expected frequency on a 1/6th chance level which is 6.7 times. The correct order, prior-posterior-visual input, is chosen 14 times out of 40, which is above chance level.

Control analysis

To control for the subjectiveness of the objective CI, a chi-square goodness of fit test is performed. In 25 out of the 40 cases, Loek Brinkman was able to correctly identify the objective CI as the visual input. This differs significantly from chance level ($\chi 2$ (5, N = 40) = 15.65, p < .01., w = .63.). These results are visualized in graph 3, were Loek Brinkman correctly identified the objective CI by putting the CIs in the following order: prior-posterior-input and posterior-prior-input.

Exploratory analysis 2

For the exploratory analysis we used part of the data from the original sorting task. Since it is of therapeutic importance that subjects are able to recognize their own mental self-image, we analysed the data for the participants who were able to do this. As seen in graph 4, the Chi-square test for goodness of fit displays that again, the first and third option are more prominent than the other options. The distribution as a whole eventually provides us with a significant result ($\chi 2$ (5, N = 40) = 14.83, p = .01., w = .61.). The results do not seem to differ too much from the data in analysis 2, which includes all the participants. The correct order is chosen at a 38.89% rate and the third option is chosen 30.43% of the time.



Graph 4. The analysis of the sorting task, only including the participants that were able to recognize themselves in the recognition task in the first try. The graph shows the observed frequency of the times the researcher labelled the CIs in that order. The dotted line represents the expected frequency on a 1/6th chance level which is 3.8 times. In 9 out of 23 times, Loek Brinkman was able to correctly label the three CIs, which is above chance level.

In the first part of this experiment, we strived to replicate the study of Brinkman and Kennis (2017), in which they validated the reverse correlation task as a measure of the implicit self. When reverse correlation is a reliable measure to capture the mental image of a person, this technique can be implemented in therapeutic settings. Before the CIs that follow from the reverse correlation task can be used to confront patients about their actual self-image, research has to confirm that participants can identify themselves within the CI that they made. This is important because when participants cannot recognize themselves in the CI, they will be less prone to the confrontation with their self-image. We were able to replicate the results from Brinkman and Kennis with an effect size of w=1.10. This is a very high effect size, from which we can conclude that participants are indeed able to recognize themselves in the CI that they made from memory. Reverse correlation thus is an accurate method to create an image that is distinguishable by the participants themselves from five random CIs of the same gender.

For future research we must however mention it is important to control for ethnicity. We recruited two non-caucasian looking participants which might slightly influence the effect size. Participants with a different ethnicity other than caucasion can easily spot their CI among CIs from other ethnicities because the CI of a non-caucasian participant shows clear characteristics of the ethnicity. When participants with different ethnicities are included within the experiment, it becomes necessary that the five added random CIs match the ethnicity of the participant within the recognition task. Furthermore, a more representative CI can be made by using different types of base faces that also portray other ethnicities.

Additionally to the first part of the experiment, we performed an exploratory analysis over the recognition tasks which include the mirror CI and objective CI. From the statistical analysis of all three recognition tasks, we can conclude that participants are overall best at recognizing the CI that represents their implicit self-image or prior, slightly worse at recognizing the CI that represents the posterior, and do not recognize the CI that is created by an objective researcher. This provides us with very valuable information. The fact that participants are not able to recognize the objective CI could mean that the CI as created by the researcher does not capture the participant rightfully. Since the Loek Brinkman is able to match the objective CI to the

photograph within the sorting task (w=.63), the first intuition would be to conclude that this is not the case and that the CI does indeed capture the participant.

However, it could be the case that the participant was not able to recognize themselves in the objective CI, only because it is created by someone else. When an objective researcher makes the objective CI, they try to replicate the photo as closely as possible. However, participants called out that while making the self CI and mirror CI, they focussed on particular characteristics of themselves which they might think are important for their identity, like the eyes or mouth. The objective researcher that creates the objective CI is not aware of these focus points, which can result in an objective CI that does not capture these characteristics fully. This makes it easy to label the objective CI next to the photograph, as it is differently made than the self CI and mirror CI. In future research, it is important that the participants are asked afterwards on which characteristics they focus.

Another reason participants showed low recognition rates in the objective CI, could be explained in a predictive coding framework. People are not able to see the objective visual input because of their own prior, since they perceive a weighted average in the posterior. The high recognition rates for the implicit mental image, and the slightly lower recognition rates for the CI created with the mirror are in line with the theory of predictive coding. Participants should be less adequate in recognizing the posterior image since the objective visual input is inflicted upon the prior to form a posterior representation.

The low recognition rate of the objective CI could mean that participants are not be able to recognize themselves in photographs, which people in general seem to do quite well. The question arises whether they recognize their own face within photographs because they genuinely recognize themselves, or because they know that their face is present in the picture. The recognition of their own face could be simplified by external cues such as hairstyle, clothing or memory of environmental context. Following research could study whether people would be able to recognize their own image when these external factors are excluded and when they are not aware that they are captured within the photograph.

In the second part of our experiment we strived to empirically test if the CI that follows from the reverse correlation task represents the mental image, thus the prior, which was thus far only a theoretical assumption. At first glance, the results of the sorting task are significant and seem in line with the hypothesis. The fact that Loek Brinkman distinguished the right order and

identified the prior considerably above chance level implies that the reverse correlation task indeed captures the prior. This would also account for the mirror CI to represent the posterior and the photo CI to represent the objective visual input, especially since in 25 cases the Loek Brinkman was able to correctly place the photograph next to it. The fact that Loek Brinkman is able to relate the objective CI to the photograph above chance, follows logically from the idea that the visual input is not influenced by any existing priors.

However, the significance, high chi-square value and eventually high effect size could also be influenced by the fact that Loek Brinkman mentioned that it was easier to label the objective CI, since he knew that only this CI was made by another person than the participant that made the other two CIs, as mentioned before. He commented he could see a difference in the noisy pixel stimuli of the two CIs made by the participant and the one made by the objective researcher. Therefore, the objective CI would eventually be different and easier to label. This may weigh its effect in the eventual result. Therefore the chance that our supervisor correctly chose between which CI conveyed the prior or the posterior becomes higher, as you can see in the fact that he correctly chose the CI by memory to be the prior 14 times, but also incorrectly labeled the mirror CI to be the prior 11 times. Which without statistical testing does not differ too much from one and other and seems to be according to the ½ chance level.

Therefore, we suggest that in replicating this research, the objective CI should be discarded. Then the problems that are mentioned above, will not occur. To test whether the CI reflects the prior, only the assumed prior and mirror CI have to be placed in the right order next to the photograph, that is then considered the objective visual input. The mirror CI assumed to be the posterior should resemble the photo to a bigger extent than the implicit self-image, which could then be considered to convey the prior. However, the chance level would have to be significantly higher than ½, which means that more participants have to be recruited. However, it might ultimately also be the case that Loek Brinkman found it harder to distinguish the prior and posterior from one and another. This might be a result of the mere possibility that the representations of the prior and posterior resemble too much to tell them apart by the naked eye.

The fact that the objective CI lies so far apart could simply be due to the above mentioned flaws in the making and recognition of this CI. As for the resemblance of the prior and posterior, we have a student population expected to be fairly healthy in comparison to patients with a highly distorted self-image. We hypothesize that our supposedly healthy student population does not

have a highly distorted prior. Therefore, the prior and posterior are hardly distinguishable by the human eye, because they both resemble the sensory input. To solve this problem, we would lastly suggest a more objectively driven method to differentiate between the prior and posterior CIs. Such a method might be multidimensional scaling (Okada & Lee, 2016). This method would enable us to visualise the differences and similarities between the pixels in the prior CI and posterior CI data set and then provide a distance matrix. A computer algorithm could eventually be used to see if it is possible to objectively differentiate between the prior and posterior CI correctly above chance level.

Lastly, we performed an exploratory analysis of the sorting task with only including the participants that were able to recognize themselves within the first trial of the recognition task, since for therapeutic implications it is important that participants can identify themselves in their own CI. One reason for not recognizing this CI, can be that participants do not perform the reverse correlation task seriously. Factors like fatigue or boredom can result in CIs that do not fully represent their prior or posterior. When the CIs do not represent the mental image of a participant, there is a chance that the sorting task is influenced by this as well and Loek Brinkman will be less able to correctly label the three CIs. To prevent this from influencing the data, we analyse the data from the sorting task from only the participants that can identify themselves in their CI made by memory. These results are significantly as well, but have a slightly lower effect size (w=.61) than the data from the sorting task including all participants (w=.67). Concluding from this, the recognition rate of the CI does not have an influence on the ability to label the CIs as prior and posterior and visual input.

Most interesting however, might be the fact that our research method is able to empirically show that self-perception seems to work in line with predictive coding mechanisms and its hypothesized Bayesian method. Therefore, reverse correlation could be a newly considered valuable tool to investigate the predictive coding theory in (self)perception. We offer this suggestion, since our research seems to distinguish a prior, posterior and the objective visual input between the CIs made in the reverse correlation task. However due to the limitations, this method should be investigated further by implementing our improving suggestions to shed light in bayesian mechanisms in perception. Previous research has already offered a theoretical framework of such as a Bayesian account that seeks to integrate the process of predictive coding with representation and recognition of the self. According to Apps and Tsakiris (2014),

self-recognition is believed to also occur by forming probabilities of the prior predicted sensory input of the self, which are then matched to the actual input and eventually these predictions are altered to explain away the prediction error which eventually forms the actual self-representation, we for instance perceive in a mirror.

Furthermore, the fact that the recognition task works most effectively for the CI by memory, slightly less for the mirror CI, and considerably less for the objective visual input CI, suggests that self-recognition works likewise according to a bayesian account. Apps and Tsakiris (2014) already suggested that recognition of the self works through bayesian mechanisms that code the expected input that is "most likely to be me". Therefore, the self would not be considered a fixed concept, but rather a more flexible conceptualization based upon the probabilistic representations that predict what input is most likely to reflect the implicit self - the predicted prior - one sees in their mind's eye. The representation of the implicit self is then considered a probability distribution which explains why you can still recognize yourself when you for instance get a nose piercing, have formed a new scar or see yourself in limited pixel stimuli as is the case in the CI of the reverse correlation task.

In conclusion, we can state that participants are able to recognize themselves in the CI that they created themselves. It has to be taken into account that the five random CIs in the recognition task match the participants gender and ethnicity. When an objective researcher makes their CI, participants are not able to recognize themselves. This can be explained through the predictive coding theory, or simply because the objective researcher did not pay attention to the same characteristics as the participant. Through the sorting task, we were able to empirically test that the CI indeed represents the prior. However, there are a few weaknesses in our study design which could have led to significant results. It is important that in future research, the objective CI is disregarded and our other suggestions are implemented. By doing so, reverse correlation proves itself to be a valuable method to research in the context of predictive coding so that we eventually understand human perception.

Apps, M. A., & Tsakiris, M. (2014). The free-energy self: a predictive coding account of self-recognition. *Neuroscience & Biobehavioral Reviews*, *41*, 85-97.

Bentall, R. P., & Kaney, S. (1996). Abnormalities of self-representation and persecutory delusions: A test of a cognitive model of paranoia. *Psychological Medicine*, *26*(6), 1231-1237.

Brinkman, L., Todorov, A., & Dotsch, R. (2017). Visualising mental representations: A primer on noise-based reverse correlation in social psychology. *European Review of Social Psychology*, *28*(1), 333-361.

Brinkman, L. & Kennis, M. (2017). Herkenning van een visueel zelfbeeld. *Unpublished.*

Brinkman, L., Yilmaz, D., Verspui, M., & Brinkman, L. (2018). A Replication Study - Visualising the Implicit Self-Image. Retrieved from https://osf.io/6esrw/

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. Behavioral and brain sciences, *36*(3), 181-204.

Cohen, J. (1992). A power primer. *Psychological bulletin*, *112*(1), 155.

Dotsch, R., & Todorov, A. (2012). Reverse correlating social face perception. *Social Psychological and Personality Science*, *3*(5), 562-571.

Dotsch, R., Wigboldus, D. H., Langner, O., & van Knippenberg, A. (2008). Ethnic out-group faces are biased in the prejudiced mind. *Psychological science*, *19*(10), 978-980.

Esposito, R., Cieri, F., di Giannantonio, M., & Tartaro, A. (2018). The role of body image and self-perception in anorexia nervosa: the neuroimaging perspective. *Journal of neuropsychology*, *12*(1), 41-52.

Friston, K. (2010). The free-energy principle: a unified brain theory?. *Nature reviews neuroscience*, *11*(2), 127.

Gorilla. (2018). Retrieved from https://gorilla.sc//

Hedges, L. V., & Pigott, T. D. (2001). The power of statistical tests in meta-analysis. *Psychological methods*, *6*(3), 203.

Imhoff, R., Dotsch, R., Bianchi, M., Banse, R., & Wigboldus, D. H. (2011). Facing Europe: Visualizing spontaneous in-group projection. *Psychological science*, *22*(12), 1583-1590.

Karremans, J. C., Dotsch, R., & Corneille, O. (2011). Romantic relationship status biases memory of faces of attractive opposite-sex others: Evidence from a reverse-correlation paradigm. *Cognition*, *121*(3), 422-426.

Lick, D. J., Carpinella, C. M., Preciado, M. A., Spunt, R. P., & Johnson, K. L. (2013). Reverse-correlating mental representations of sex-typed bodies: the effect of number of trials on image quality. *Frontiers in psychology*, *4*, 476.

Lundqvist, D., Flykt, A., & Öhman, A. (1998). The Karolinska Directed Emotional Faces – KDEF, CD ROM from Department of Clinical Neuroscience, Psychology section, Karolinska Institutet, ISBN 91-630-7164-9.pl_1), S242-S242.

Okada, K., & Lee, M. D. (2016). A Bayesian approach to modeling group and individual differences in multidimensional scaling. *Journal of Mathematical Psychology*, *70*, 35-44

Schmack, K., de Castro, A. G. C., Rothkirch, M., Sekutowicz, M., Rössler, H., Haynes, J. D., ... & Sterzer, P. (2013). Delusions and the role of beliefs in perceptual inference. *Journal of Neuroscience*, *33*(34), 13701-13712.

Shorten, C. A. (2017). Similarity Between Actual and Possible Selves and Its Relationship to Self-esteem.

Young, A. I., Ratner, K. G., & Fazio, R. H. (2014). Political attitudes bias the mental representation of a presidential candidate's face. *Psychological Science*, *25*(2), 503-510.