Social Behavior and the Brain: An Introduction to Social Neuroscience

Dr Ark Verma
Department of Cognitive Sciences
Indian Institute of Technology Kanpur

Week - 02
Lecture – 09

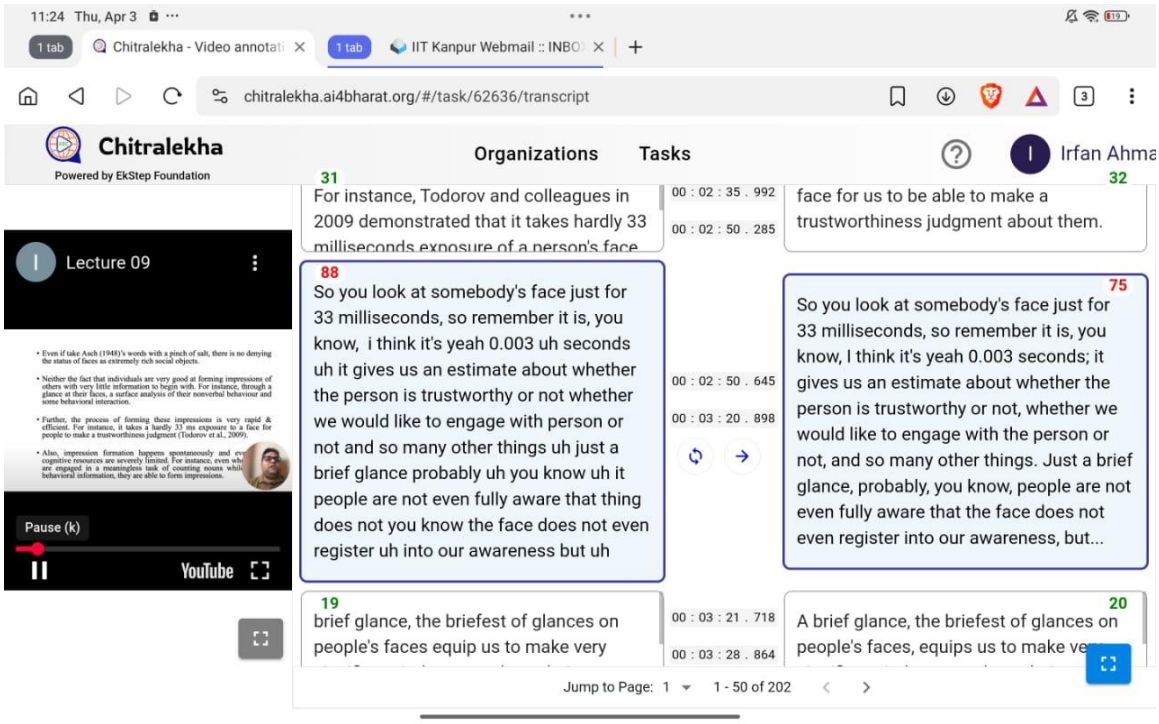
Hello and welcome to the course Social Behaviour in the Brain. I am Dr. Ark Verma, an Associate Professor in the Department of Cognitive Science at IIT Kanpur. This is week 2 and we have been talking about retrieving and representing personal knowledge in the brain. In this lecture, I am going to be talking about faces and their significance as social stimuli. Now, Solomon Asch in 1948 opined that a simple look at an individual's face are sometimes sufficient for us to form an impression about them.

A slight eye contact, an exchange of a few words provided with them, it provides us with a first impression of people. Obviously with more interactions these impressions will become richer and closer to truth that faces are the special objects that carry a wealth of social information about individuals and the human brain is capable and it does basically take all of that information in a very short time very rapidly and uses it for other kinds of approximations in social behavior. Now, even if we take Ash's words with a pinch of salt, there is no denying the status of faces as extremely rich social objects. They carry a lot of information.

They carry information about people's intentions. They carry information about their mental states. How are they feeling at this point? A lot of times people, you know, claim or boast to be able to judge other people's personalities based on their faces. And they sort of.

The mode of interaction, the manner of interaction is decided based on the first exchange of a glance or a facial interaction with individuals. So again, neither the fact that individuals are very good at forming impressions of others with very little information to begin with, For instance, through a glance at their faces, a surface analysis of their non-verbal behavior and some behavioral interaction actually sort of, you know, we cannot deny that it carries that kind of weight. Now, the process of forming these impressions is actually very rapid and it is very efficient. For instance, Todorov and colleagues in 2009 demonstrated that it takes hardly 33 milliseconds exposure of a person's face for us to be able to make a trustworthiness judgment about them. So you look at somebody's face just for 33 milliseconds, so remember it is, you know, i think it's yeah 0.003 uh seconds uh it gives us an estimate about whether the person is trustworthy or not whether we would like to engage with person or not and so many other things uh just a brief glance probably

uh you know uh it people are not even fully aware that thing does not you know the face does not even register uh into our awareness but uh brief glance, the briefest of glances on people's faces equip us to make very significant judgments about their personality. Also, impression formation is not something that we voluntarily decide and we plan and we strategize to do. It happens rapidly and even in cases where cognitive resources are very limited. So for instance, even if we are engaged in a secondary task, a meaningless task of counting nouns while gathering behavioral information, we are able to form impressions based on faces. Moreover, people have been shown to be capable of using the dynamic changes in the face such as through expression of emotions to understand the immediate meaning of a given context and even invariant facial features to be able to identify others.

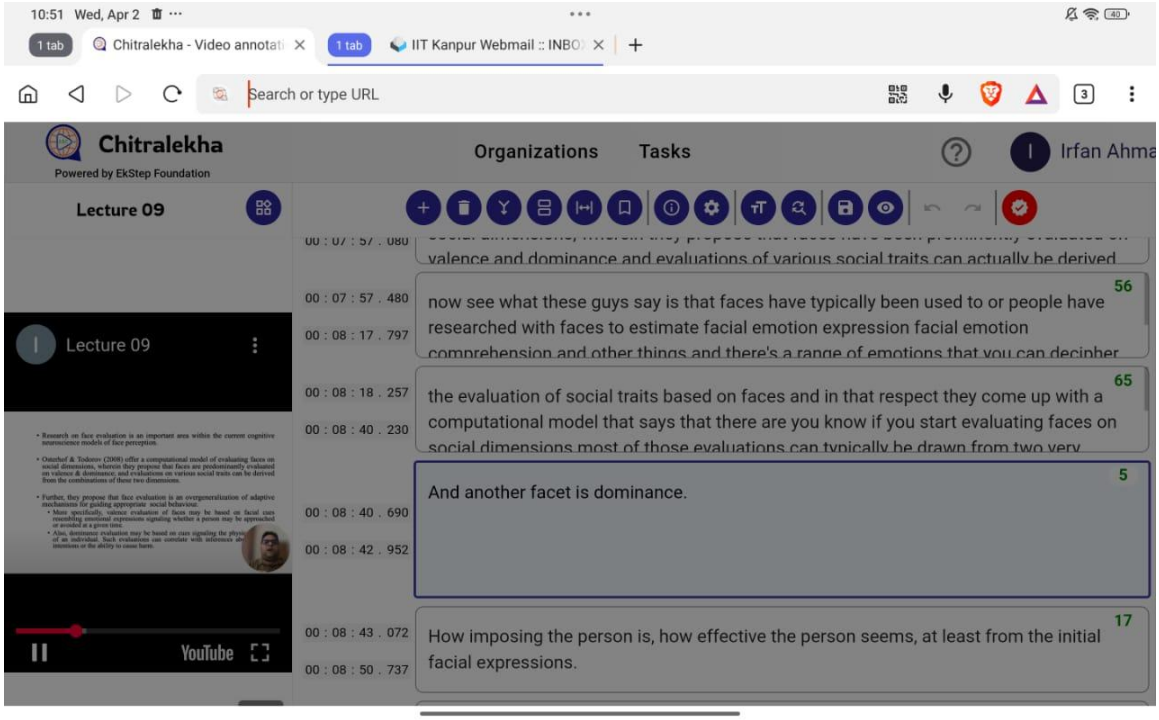


Sometimes you can basically identify others based on their specific properties, maybe the way their faces look, but also competitions of what kind of emotions somebody typically carries on their face will help us identify that individual. Also we should remember and this is again something that we'll keep revisiting in this lecture is that faces act as a rich reservoir of personal inferences even though sometimes those inferences can be erroneous. You might all have learned or read about come across the dictum that says oh love at first sight. What is love at first sight? It's basically a degree of interaction that one feels when you look at somebody somebody's faces and maybe the positive attributes about the face or maybe you know the aesthetics the beauty the symmetry any number of these things encourage you to have a positive impression of that individual based on their face and you basically and can extrapolate it to an infatuation or feeling

like you know you're already in love even though you've not seen this person earlier you don't know it the person from before But again you can see faces sometimes act as a window to individual's personalities and a lot of times people are swayed by these decisions. Now, interestingly, again, while I'm saying all of this, you must be cautious of the fact that obviously first impressions are not super accurate.

You know, they may miss a bunch of things. But however inaccurate these personal inferences can be to start with, you know, face perception. they can actually help predict social outcomes ranging from electoral success, you know, people might vote for candidates who they see having a smiling expression on their face, a pleasant personality, a good one to be around with, to even sentencing decisions. So, for example, in judiciary, some studies have shown that the judges were inclined to, you know, to be relatively softer to individuals who came across as pleasant amicable having a pleasant smiling expression on their faces and so on so for instance and just to extend this for instance inferences of competencies from faces are able to predict electoral success and in and basically you know things like dominance are able to predict even uh eventually eventual military rank attainment so you know in one of these studies they looked at the faces of army generals and people made a judgment of say for example whether they are going to go on and you know achieve a higher highest rank in the military and a lot of people made these decisions based on individual spaces and those decisions actually were able to predict that these individuals eventually reached higher ranks because of the expression of dominance on their face. So, this is why faces are extremely rich stimuli.

Also, research on face evaluation is a very important area with the current cognitive neuroscience, you know, the current cognitive neuroscience models of perception. For example, Ostarov and Todorov offer a computational model of evaluating faces on social dimensions, wherein they propose that faces have been prominently evaluated on valence and dominance and evaluations of various social traits can actually be derived from the combination of these two dimensions. now see what these guys say is that faces have typically been used to or people have researched with faces to estimate facial emotion expression facial emotion comprehension and other things and there's a range of emotions that you can decipher from the faces but something that they point out that may have been missing is the evaluation of social traits based on faces and in that respect they come up with a computational model that says that there are you know if you start evaluating faces on social dimensions most of those evaluations can typically be drawn from two very important you know facets and those two facets are valence of the face you know ranging from positive to negative And another facet is dominance. How imposing the person is, how effective the person seems, at least from the initial facial expressions. Also, they propose that face evaluation is an overgeneralization of an adaptive mechanism which guides appropriate social behavior.



So overgeneralization is an interesting concept here. Basically what they are saying is, that people basically base or their decisions are derived from the first impressions, the first happy or sad or angry expressions on people's faces and these can be overgeneralized to interpret people as positive, negative, aggressive, submissive and so on. okay so they basically propose in more specifically that valence evaluations of faces may be based on facial cues resembling emotional expressions that signal whether a person may be approachable or that person should be avoided at a given time on the other hand dominance evaluation may be based on cues signaling the physical strength of an individual see we talked about that study where army generals photographs were rated on the fact whether they can actually uh eventually go on to get a higher social rank. So if the face signals, you know, enormous physical strength, mental strength maybe, although that is not included in the study, that eventually hints at us, it provides us with a hint of the fact that, oh, these people, you know, will be dominant or not so dominant and so on. So let us move on to this cognitive neuroscience you know perspective on face perception.

We will look at the model that Osseroff and Todorov have provided in much more detail and we will try and understand the ramifications of this model for the study of faces. Now, as I was saying, most research in cognitive neuroscience of face perception are mainly focused on either face recognition or recognition of emotional expressions. For instance, for that matter, ephemeralized studies have shown that whereas areas in the fusiform gyrus are more responsive to facial identity information, areas in the superior temporal sulcus are more responsive to facial emotion expression information. Now

obviously Todorov and Oserov sort of looked at this data, they looked at a lot of this research and they started building on this cognitive neuroscience model of face processing, results from single cell recordings, fMRI data, model of the neural processes that may underline facial perception and they developed this model to capture the disparity between processing of facial identity and emotional expressions. Now according to this model, the model that they propose, the major distinction in face processing is between the invariant facial features and the dynamic changes such as eye gaze and expressions which keep on changing momentarily.

For example, the invariant facial features may be, you know, the way my nose is or my mouth is and the way I, you know, my face looks when I am you know showing a neutral expression but depending upon say for example how I am interacting with you if this were a class you probably there would be some questions somebody would be distracted so based on that interaction my facial my eye gaze will move here and there my facial emotions will change dynamically so this basically is the difference between invariant facial features and dynamic facial features As is said, whereas invariant facial features are more important for facial recognition, for personal recognition because you can sort of use those to distinguish between two people's faces, dynamic facial changes are very important for estimating the mental state of others. So looking at the face of others, looking at the dynamic expressions that keep coming and going, you can determine whether a person is happy or sad, is angry, disappointed, disgusted even at a given point in time. And this model has actually been shown to account for these observed dissociation that we were talking about between the facial identity and emotional expressions. For instance, prosopagnosics, despite their ability to recognize faces, show near normal perception of emotional expressions. So you can see, even from neuropsychological results, even from neuropsychological data, there is enough evidence of the fact that computation of facial identity and computation of emotional expressions on the face are actually relatively different things.

Now however the mechanism of how person inferences such as things like trustworthiness and competence you know that dominance or competence or trustworthiness come about okay let's let's look at those those mechanisms are not super clear to us and these inferences although are based on you know invariant facial features expressions of emotions affect trait judgments as well, and it has been possible to observe these dissociations between processing of facial identity and impression formation. Let's take an example. In a study with four developmental prosopagnostics having severe impairments in both memory and perception of facial identity, the authors report that their judgments of facial trustworthiness across three different types of sets of faces were actually within normal range of the control judgment. And the performance of 2 out of 4 process prognostics were actually very typical. It was normal or above normal and so on.

So this finding actually suggests that indeed the mechanisms for identifying faces versus the mechanism for understanding facial emotions are actually dissociable and different mechanisms. Now these authors propose that while there has been some research on facial identity perception and recognition of facial expressions, very little research has been dedicated to the focused evaluation of faces across social dimensions. You know things like whether a person is trustworthy or not, is cooperative or not, is dominant or not, those kind of social dimensions, social interaction dimensions. And something that has been sort of researched to a relatively lower level is this idea of trustworthiness. For example, Adolphs and colleagues demonstrated that patients with bilateral amygdala damage should impaired discrimination between trustworthy and untrustworthy faces.

So you can see that these kind of trait estimations are also contingent on the limbic system structures that are involved in computation of emotions. Also, fMRI studies with normal individuals have also confirmed the involvement of amygdala in the evaluation of facial trustworthiness. So people are you know people are actually making these judgments based on you know facial estimation of trustworthiness and other social traits and people have been able to others have been able to identify the neural mechanisms underlying or supporting these estimates Now this model proposed by Todorov and colleagues is referred to as a dimensional model of evaluation of emotionally neutral faces. Let's move on with that. While people engage in a variety of trait judgments from faces, these judgments actually turn out to be highly correlated with each other.

For instance, using a set of standardized faces, authors obtained a correlation of 0.83 between judgments of trustworthiness and emotional stability 0.75 with judgments of attractiveness trustworthiness and attractiveness and minus 0.76 with judgments of trustworthiness and aggressiveness and 0.63 you know of correlation with judgments of intelligence so you can see if you are able to delineate let's say trustworthiness some of these other positive traits estimation will actually go hand in hand with that and is highly correlated so you can see already positive traits forming another cluster and negative traits may be forming another cluster Given the high correlations among judgments of different traits, there seems to be a difficulty in identifying first a neural trait specific to a trait dimension and facial configurations that vary along within this dimension only.

For instance, if the modeler's goal were to model the responses to phases as a function of multiple trade judgments, these correlations among judgments would introduce serious collinearity problems because you would not know that whether it is the same measure sort of reflecting or we are actually being able to identify different variables here. Proposal of these models actually seem to undertake an approach of reducing the judgments on multiple trade dimensions to a few orthogonal dimensions. So based multiple trade judgments to few orthogonal dimensions. They say because this is so you know correlated and they cluster together let us figure out two major dimensions along which you can sort of cluster these different trade judgments okay so what did they do

they first identified the trade judgments on which faces were spontaneously evaluated and then they collected these judgments on you know different trade dimensions that we were just saying Finally, what they did was they submitted these judgments to a principal component analysis to identify the underlying dimensions of phase evaluation. So what they said, oh, what are the traits that we judge phases on X, Y, Z, A, B, C and so on.

But what are the underlying dimensions between these judgments? So maybe there are just D and E. So basically, you see what clusters with D, what clusters with E and that basically that information can come through the principal component analysis. So the first principal component actually accounted for 63.3% of the variance of the mean trade judgments. all positive judgments to be elaborate all positive judgments so attractiveness responsibility punctuality actually had positive loadings and all negative trade judgments such as aggressiveness unpleasant behavior and so on had negative loadings so it seems that you know this all of these things together can be judged across the dimension of valence so valence basically uh you know you can think of valence as a continuum where on one end there are all positive kind of judgments on the and on the other end there are all negative judgments .

The second principle component accounted for 18.3 percent of the variance so judgments of dominance aggressiveness and confidence had the highest loading on this component you know on dominance suggesting that it can be interpreted alongside the dominance evaluation so what are the two dimensions they come up with They come up with valence where from all positive to negative judgments are aligned. And then there is this dimension of dominance where aggressiveness, vehicle strength, those kind of things come up. Now such a two-dimensional structure of face evaluation seems consistent with the well-established dimensional models of social perception. So this is not something new, it is just basically you are trying to bring together a lot of analysis, trying to put it together and trying to understand and discern patterns that would basically tell us the underlying mechanisms of facial emotion perception or facial trait judgment for that matter.

So starting with a large set of traits describing interpersonal relationships, Wiggins and colleagues had also shown that these traits can be represented within a two dimensional space defined by affiliation and dominance. So they actually are also, you know, several years earlier, I think almost 20 years earlier, Wiggins and colleagues also tried to do this. but their two dimensions were affiliation and dominance and these dimensions if you see are actually very similar to valence and dominance that Ostarov and Todorov actually point out the principal component analysis also gave some different insights so they demonstrated that the valence and dominance dimensions can actually be you know approximated by single trade judgments as well For example, the judgments of trustworthiness had the highest loading 0.94 on the first principal component, remember

that is valence and it was practically uncorrelated with the second principal component. So it is not correlated with dominance but it is correlated with valence.

Whereas judgments of dominance had the highest rate loading on the second principal component and the lowest loading on the first principal component. So you can see that the data sort of trade judgment and of emotion expression judgment is clustering very nicely through this principal component analysis that these authors seem to have carried out. now this is not something limited to that analysis this was the case even when principal components were obtained from another analysis excluding these two judgments to avoid biasing the PCA solution to be more specific what they did was that what they found was that trustworthiness judgments were highly correlated with the first principal component that was on valence but not with the second principal component which was dominance okay And dominance judgments were highly correlated with the second principal component but not with the first principal component. So it basically tells us that the data are sort of very clearly segregated. Additional analysis showed that the two-dimensional solution is robust with respect to the set of traits used to estimate the principal components and the phases of stimuli.

So you can see very easily that according to this dimensional model of social evaluation of faces, two components actually come out very nicely which is the component of valence and dominance. And then there is this idea of trustworthiness which sort of lies along the valence dimension and the idea of physical dominance that lies along the dominance dimension. So let us see, we will zoom in a little bit more into their computational model. okay so moving on from this principal component analysis the authors move towards the modeling of phase trustworthiness and phase dominance okay what do they do as the judgments of trustworthiness and dominance could actually be used as approximations of the underlying dimensions valence and dominance you know Phase evaluation, you know, the authors actually started to build a computational model for representing how phases would vary on trustworthiness and dominance Okay, because we saw that these things, you know, these two dimensions could be used to predict the underlying dimensions Then they said okay. Let us look at how these two dimensions work out when you are doing phase evaluation So authors employed a data-driven statistical model of face representation in which faces were represented as points in a multi-dimensional space.

I'll show you the figure in a bit. And the input to this model was a database of faces which was laser scanned in three dimensions. Wherein the shape of the 3D image was represented by vertex positions of a polygon model of fixed mesh topology. I will show you that in a bit. And finally what they did was they again used principal component analysis to represent using that the representation of each face was reduced to a single point with a number of independent components.

You can look at this figure here. So you can see here the faces are presented with you know along this dominance and trustworthiness dimension as a point in this vector space. So using the model of the face representations the authors generated emotionally neutral faces. So basically they wanted the model to learn this classification or categorization so they first developed emotionally neutral faces. And they intentionally used only white faces so as to avoid any confounds based on race, ethnicity and so on. Now the participants initially were asked to judge these phases on trustworthiness and dominance and then the authors computed the mean trustworthiness and dominance judgments and you know for each phase and then they use this data to find vectors in the 50 dimensional space whose direction was optimal in changing trustworthiness and dominance you know respectively.

So you look at that entire continuum. Now these vectors were actually based on the best linear fit of the judgments as an interpretation as a function of the 50 shape components, all of these phases. And finally to obtain an orthogonal solution, you know, these categories, the authors rotated the dominance vector to make it orthogonal to the trustworthiness thing. So here they are now being able to independently observe the values for dominance and trustworthiness. Now they wanted to evaluate the computational model. So the authors randomly started generating faces.

Next for each face, the authors created several versions that varied along the respective dimensions. Remember, so they take the face of one individual and they generate several versions of it, starting from the least trustworthy to the most trustworthy, positive valence to negative valence. And then they asked human participants to rate these faces across the set dimensions. So valence and trustworthiness. Now these studies showed when these studies were conducted, they showed that the models of trustworthiness and dominance were able to successfully manipulate trustworthiness and dominance.

How did they do that? Trustworthiness and dominance judgments of faces generated by these models actually tracked trustworthiness and dominance as predicted by the model respectively. So they were able to use this vector space to generate exactly correct faces which were depicting the real traits. For example, the phases generated by models varying along different points in the continuum of trustworthiness actually were rated as such. So, if a model is generating a less trustworthy phase, the individual, human individual is rating that as a less trustworthy phase. If a model is generating a high trustworthy phase, then the participants are able to rate that as a high trustworthy phase.

So, you can see here that the computational model with these dimensions is actually successfully manipulating trustworthiness or dominance across a particular vector scale.

Now, interestingly, whereas dominance judgments of faces generated by the dominance dimension were related in a linear fashion to the face dominance, trustworthiness

judgments of a face generated by the trustworthiness dimension were related in a quadratic fashion. Again, this is not something I would bother you with, but just understand that there was a systematic relationship between these traits and the actual social trait judgment based on these faces. Now, in a bit more detail, people were found to be more sensitive to the changes at the negative end than at the positive end of the trustworthiness dimension. And this should not be surprising because how trustworthy is an individual? On the higher side, let's say on the latter end of the continuum where people are 90% trustworthy or 92% trustworthy or 85% trustworthy does not really matter that much.

However, if the people are less trustworthy and even less trustworthy and so on, that basically will create some kind of survival questions, engagement questions and so on, whether you want to engage with this kind of a person, whether you want to interact with such kind of a person. So people were very sensitive to be and they were able to pick up the variations. on the negative side of this continuum very well even better than they were able to pick out you know differences on the positive side so they carried out validation studies and those studies basically exemplify some of the advantages of using these kinds of formal models you know of how faces can vary across social dimensions So, first these models could generate and what are these advantages let us go through them one by one. First is that these models could actually generate an unlimited of number of faces that vary on these dimensions of interest. So, for example, if you want to go between happy and sad let us say valence or you want to go between dishonest and honest, trustworthiness, you can actually generate hundreds of faces along that continuum and you can get them validated by human participants and you have a lot of data to play with and understand how these mechanisms of identifying trustworthiness or valence are working within the human brain.

The next advantage is that the variation that these models can produce can basically allow for very precise manipulation and the range of differences can be little and large and basically you can manipulate or calibrate these things so as to allow you to test your very specific hypothesis and understand what is changing with the human brain with respect to these things. For instance, a lot of previous studies had failed to find that trait judgements could be made even after subliminal exposure to faces. However, the stimuli may not have been sufficiently different from the trait judgements of interest. So for example, see when you create facial databases and there are several facial databases that are there, so rat-bound faces database, There is Caltech faces database and you will see a lot of these databases have faces which show emotions across a continuum. Most of these databases show six basic emotions and they show surprise and disgust.

There are not a lot of databases which do different degrees of happiness. So what happens is that when you can artificially create these pictures through a model, these phase

pictures through a model which specifies and understands the mechanism then you can actually you know very precisely do these manipulations along the dimension of your interest be it valence be it dominance and then you can use that large set of stimuli to understand how things go because when you ask somebody to show let's say if you are to taking my photograph and you were to ask me oh can you show me 100% happiness I'll make a face and then you tell me oh can you show 90% happiness then I'll make another face then you ask me to show 70 or 80% happiness while I'll probably be doing it more honestly you'll be probably clicking photos with you know honesty it is not guaranteed that myself as an individual or many other human subjects that we interact with and collect data with we'll be able to distinguish these different layers or these different levels of emotional expressions. A computer model can do it much better that way. So authors use the faces generated by this trustworthiness dimension to test for subliminal effects and presented these faces just for 20 milliseconds and then immediately masked them by the neutral version of faces. So they showed a trustworthy face and masked it with a neutral face.

The trustworthy face was shown just for 20 milliseconds. What did they find? They actually found that the trustworthiness judgments of the neutral faces were more negative when these faces were preceded by untrustworthy than by the trustworthy faces. So it seems that these primes are being able to work even at such small time intervals such as 20 millisecond difference. And these findings together suggest that people are actually able to extract information and judgments even when the faces are presented below their level of awareness. Again, something I said in the beginning of this lecture. Now, the third advantage that comes through using these computational models is that these models can be used as a discovery tool to identify the variations in PHEQs that produce or map with very specific judgments.

Although these models are typically holistic in the sense that they are constrained by any set of facial features, they can be used to discover the important features a posteriori. So once you've sort of generated a bunch of these faces and you've got your judgments, you can actually go back and look at which facial feature was changing systematically and how does the change in this facial feature contribute to the judgments of trustworthiness or untrustworthiness and you can say oh you can identify oh this feature is something very critical or this other feature is something very critical. So these are some of the very interesting offerings of computational models that are very difficult to get through if you are using real databases of faces because the precise manipulation, the a posteriori sort of understanding of these features, bunch of these things are very difficult to achieve through these dimensions. Now, experiments have confirmed that these two dimensions, trustworthiness and valence, are actually sensitive to different types of facial information.

Let's just look at these mechanisms. For instance, the authors first randomly generated

faces and created extreme versions of the faces on the trustworthiness and dominance dimensions. And then in a study where participants were asked to categorize these faces, as expressing you know one of the six basic emotions they found that these people were able to classify the exaggerated faces in the negative direction on the trustworthiness you know judgment as angry and similarly exaggerated faces in the positive direction as happy so you can see that these dimensions or variations along these dimensions can be actually used to predict you know the mental state or the facial expression emotion expression from these faces. Now another thing, again we are talking about validation studies for these models. In additional 9 studies, participants were asked to rate the faces on continuous scales such as on a scale of 1 to 5, rate these faces as being angry or happy, where 1 is angry and 5 is happy.

And then you can use other things. Oh, is this baby-faced expression or mature-faced person or is this face very masculine or very feminine? You can actually ask people to do these dimensions. Okay, so the authors manipulated these kind of information available for the judgments and in three of the studies participants rated the intact faces, in three studies they rated the faces with their external features were masked and in three they rated the faces with their internal features masked. So they sort of are manipulating the features and trying to understand which of these features are critical for you know judgment of let's say anger, happiness, baby face, mature face or femininity and masculinity. The studies actually showed that whereas the trustworthiness dimension was more sensitive to features resembling happy and angry expressions, the dominance dimension was more sensitive to features signaling physical strength.

Again, something that we saw much earlier as well. Now sort of you know just bringing this all together most of the diagnostic information for the trustworthiness dimension was present in the internal features of the face. So trustworthiness seems to correlate more or draw more from the internal features of the face whereas most diagnostic information about the dominance dimension was basically drawing from the external or the overall shape of the face. So and in principle this 2D model that you know these guys have proposed was able to represent social judgment from faces. It was able to simulate social judgment from faces. This another dimension they develop a little bit later is this dimension of threat.

So they basically found that threat judgements which are particularly important from a survival point of view are also highly correlated with both trustworthiness and dominance judgements. So for example threatening faces are found to be both untrustworthy and dominant looking and authors sort of you know they basically built a separate dimension in the space defined by trustworthiness and dominance dimensions by giving equal weights to these two dimensions and found that these were practically identical to a threat dimension based on threat judgment. So while the take home message is actually that most of the social trade judgments of faces are along the dimensions of valence and

dominance and an example is trustworthiness, but there are also other dimensions possible, one of which is the dimension of threat. So I hope you appreciate the character of faces as extremely rich social stimuli. We will continue talking about this in the next lecture, which will be the last lecture of this week. Thank you.