

## Lecture 9 – 08/04

An fMRI Investigation of Attributing Negative Social Treatment to Racial Discrimination – Masten, Telzer & Eisenberger (2011)

### INTRODUCTION

They start with some pretty clear things about social relevance, which is good because in a lot of studies social relevance is hard to grasp:

60% of Black Americans perceive racial discrimination on a daily basis

Repeated exposure to discrimination...may contribute to well-established health disparities between Black and White individuals (at least in the USA)

If you go back to 30 years ago, people talked about discrimination and racism but when the implicit attitudes have been discovered (about 20 years ago) we started to consider it differently.

Currently when people talk about racism in culture, there is something called microaggression, which is a small subtle aggressive tone or statement that does not necessarily have to be overt but they have a significant effect on people. So this little insults that accumulate throughout the days, months and years have a major effect on people rather than the less common and larger racist event where people overtly say explicit racist comments.

This is the background : people in the USA – but probably it happens in most of the cultures for minorities – (and in this case we are talking about Black Americans) experience this on a daily basis.

### LITERATURE

Research on the victim's experience has focused on self-reports, and has not considered the moderating role of attributional "Style".

There are two things to pay attention to in this statement:

What are the neural substrates?

Self-reports have limitations and this is one of the reasons why people moved away from them and rather started to use other scales such as the IAT, which is measuring non conscious biases.

Often people self-reports are biased by memories, self-presentation, social desirability effect, and they might deny their biases and prejudices even to themselves. This might have to do also with experience: they may not want to see themselves as victims, and appear strong and so they may deny experiences.

In this research, they want to get beyond self reports and look to the neural substrates: instead of asking people how they feel, they want to look at what brain areas are activated during these experiences.

What is the moderating effect of attributing the negative experience to racial discrimination?

This can go two ways:

INTERNAL (I'm Black):

One possibility is that they can either attribute it internally to themselves as something that cannot be changed. The fact that they blame it on themselves can have the effect of being victimized.

## EXTERNAL (they're prejudice)

In external attribution, they would say that the person is racist, a bad person and they want to escape from them. In this case there is nothing to do with the individual, but it has to do with the person that expressed the racist attitudes and behaviors.

This gets into a long history of attributional statuses regarding depression. For example, in this internal vs external there is a sort of depressive genic attributional style, meaning that it leads to higher levels of depression: people can attribute their failures to internal and stable characteristics of themselves (e.g. someone fails in exam, and attributes the failure to himself -> it is internal and stable since it is his intelligence and it cannot be changed) or to external characteristics (e.g. someone fails in an exam and blames the professor). Anyone can make one type of attribution or the other, but some people have the tendency to do the first one.

This is what are they trying to understand in this study: does this internal attributional style -which has been associated with depression in regard to other aspects of life- also play a role in experiencing racist aggressions or events?

These are the two main things that distinguish this study from the past researches that were focusing on the substrates of being a victim and the possible moderating role of making different attributions.

Question: so this means that if I am experiencing racist microaggression, then what is it in my thoughts that is making me experience this?

Answer: In a very basic level: the subjects are going to be exposed to some racist or potentially racist event. Most of the people would probably interpret it as a racist but it is not a clear overt racist event. Then they are trying to study what effects it has on a person's level of distress and what effect the person's attributions have on the level of distress. So we have racist experience leads to distress and then we have the idea of attributions that could change that possible effect.

But when we are going to get to the methodology it will become clearer.

## NEURAL SUBSTRATES

They focus on past research:

Dorsal Anterior Cingulate Cortex (dACC) and the Anterior Insula -> have been related to "painful" or distressing aspects including social rejection, which is the type of experience they are going to have.

These two areas have to deal with the affective response.

Lateral and Medial Prefrontal Cortex (PFC) and Rostral ACC (rACC) -> regulation of negative affective responses to social or painful threats.

These two areas have to deal with regulation of affective response.

## PREDICTION

With regards of the effect of the social exclusion (which is specifically what they are focusing at), they predicted that the distress associated with negative social treatment would be associated with:

Greater activity in regions associated with social distress (dACC, anterior insula)

Reduced activity in regions involved in emotional regulation (PFC, rACC)

This is a little bit surprising: they are saying that these two types of brain areas are going to be inversely related to each other (one goes up and the other goes down). As we experience more affective response, we experience less regulatory effect and vice versa.

The professor suggested though, that we engage more regulatory thoughts to down-regulate the effect. So the regulatory response is going to depend on the active effect of response and to engage regulation I first need to experience affective response. If I am not affectively activated, I am not angry or offended, there is no reason for me to engage in cognitive response, in regulatory response. So it should be easier to say that they are positively rather than negatively correlated.

When they say they are correlated, they are actually not talking about correlation but rather they are saying that with regards to distress, one should go up and the other down. So for them, distress is defined by greater activity in the affective regions which should mean that we go down in regulation, but for the professor, the more we became affectively offended, then we should start engaging in more regulatory activation (we are not going to need regulation until we are offended and that's what's going to drive regulation).

In the study they don't really give an explanation of why they expect this effect, without considering a different model. This is a limitation because every time we propose a model, we should always consider an alternative model or hypothesis.

About the Moderation of Attribution of Discrimination, they say it could go either way (and the professor agrees):

If : it's personally threatening

Then : it will increase distress

And : we are going to have More dACC and Anterior Insula and Less PFC and rACC

That is the same if the attribution are personally threatening or raise the threat: if the attribution is internal and stable, they are going to be more personally threatening (it's my fault and it's a stable thing, I am always going to be a victim and there is nothing I can do about it) and this is going to increase distress and it's going to lead to more activation in the affective areas and less in the regulatory ones.

If : the attributions represent effective coping, and in this case probably external and probably less stable (situation in which I am saying that the person is a bad)

Then : there is going to be a reduced distress

And : there is going to be Less dACC and Anterior Insula and More PFC and rACC

Question: what do they mean with "reduced distress"?

Answer: I use the terms "increased and reduced distress" a little bit loosely without a good definition of it. In their study they are not looking at increased or decreased distress but just at the levels of distress. The professor though changed the words into "high distress" and "low distress". It is a subtle difference, but it

is a significant difference theoretically because we have two different variables: offence (racist treatment, social exclusion) which will lead to some experience of distress and then there is the moderating variable that can lower that. So it is sort of a two step process .

In this case one can ask is: does the attribution decrease relatively to the immediate offence or increase relatively to the immediate offence? They actually don't study this at all and this is one of the limitations.

## METHODS

### Participants:

Eighteen Black individuals (9 women) -> almost enough participants to analyze data by gender. They don't do it but maybe they could have since it's the same number of female and male

Age: 18 to 28 -> reasonable

Participants were recruited through fliers distributed around the UCLA campus, announcements made in undergraduate classes, posting on student list serves, and on-line advertisements in the community. Race was never mentioned in any of the recruitment materials.

### Procedure

The subjects were scanned while they were playing "Cyberball" (computerized ball-tossing game)

They met 2 white confederates (1 male, 1 female) before playing

All personnel in lab were white

They never mention race but they were just presented to the other two white players and were said that they were all going to go to different rooms with different machine and that they were going to play the game

They played cyberball during the scan

### How:

both, the black participants and the two white confederates completed a consent form together

They were introduced to each other

They were told that the others would also be in scanners

During the fMRI scan, participants played two rounds of Cyberball, one in which they were included equally for the whole game and one in which they were excluded after initially being included for 10 throws.

So they are watching the screen and they are seeing two other players in the background and they see their hands. So if someone throws the ball to me, I can throw it to one person or another person, and when they get it they can throw it back to me or to another person.

Since it is all controlled by a computer, the amount of times one receives the ball can be exactly manipulated:

In the first round, the amount of time is equally distributed.

In the other rounda, they receive it for the same amount of times for the first ten rows and then they get excluded and no one throws them the ball

This is exactly how it looks like:

First both the two cartoons figures and the participant throw the ball to each other and then the participant is excluded. They are not even looking at each other, and at the beginning they are equally throwing the ball to each other. It is just after 10 rows in the second round that the participant is excluded. So the fact that the other two participants were white was never mentioned.

Post-scan measures:

They used them to measure distress after the scans:

observer-Rater Distress : they did an interview with the participant in which they asked about being excluded and how they felt about it

Self-Reported Social Distress. The answers have been recorded in videotapes and they have been coded by 21 coders. There was a very high agreement: the correlation was of .91, which is very high and so pretty reliable

Discriminatory Attributions (e.g. why they thought they have been excluded from the game)

Comments from the professor:

The order of the rounds was not counterbalanced: the inclusion round was always first and the exclusion round was always second. This because once you have been excluded, you cannot undo that effect for a second round.

Studies about the Cyberball are hard to do on children because after a few rows they are excluded they start crying, so they had to find another paradigm for social exclusion. In this study there are adults, but the effect is similar in the results

## RESULTS

Correlation of the post-scan measures:

Self-Reported Distress – Observer-Rated Distress = .50

It is a reasonable result. Often when we want to measure the same construct, we want to find a correlation of .60 or above because we are measuring the same thing. A correlation of .50 means that there is a 25 % of variance in common between the two measures and therefore a 75 % of variance that is not in common. It is ok, but not great.

Discriminatory attribution – Self-Reported Distress = .51

Discriminatory attribution – Observer-Rated Distress = .32

It is not significant when you have 18 subjects. When you have a little number of subjects, you have to have a little less confidence in any correlation.

Both of the correlations between Discriminatory attribution and the two distress measures are going in the same direction.

So from this, thinking at their initial hypothesis with regards to attributions and distress, we can conclude that attributions are modulated by distress, but we need to replicate that because the sample is very small. 0.32 is not a strong correlation, but if you have a good sample (e.g 40 participants) it is going to be

significant. We can therefore live this result out, but it is important not to ignore it, and instead to replicate. There is a difference in demonstrating the null hypothesis and not having statistical power to really test it well.

This result is consistent with the idea that attributing social exclusion to racism and discrimination seems to be associated with personal threatening because it is increasing stress

fMRI

Areas in response to exclusion:

In the table there are the brain areas activated during exclusion vs inclusion. So we are not looking at distress and neither to attribution, it is only brain areas activated during exclusion as compared to inclusion.

The areas that are circled are the ones that have a sort of hypothesis; they didn't have any hypothesis about occipital cortex, cerebellum, premotor cortex... these are areas that were actually associated with the manipulation but were not part of the hypothesis. So they put them in the table but they do not talk about them.

In blue we have the affective response area, and the t has a positive effect: they are more active during exclusion as compared to inclusion

In red we have the areas that should be more of the affective regulation and they are also positively associated with exclusion.

So related to exclusion, we are getting an increase both in affective and regulatory areas. This is consistent with the professor's idea, and so that we should not expect inverse association: to get regulation we need affective response and so we get an affective response which then increases the regulatory response.

Correlation with Self-Reported social distress:

In the previous table there were only the areas active in exclusion relative to inclusion. Now we are looking at correlation (so it is a different kind of statistics) between distress and brain activation. What we see is:

Subgenual Anterior Cingulate Cortex, which is not hypothesised by them that shows a pattern of correlation with distress

The dorsal medial prefrontal cortex which is negatively correlated with distress, and this correlation makes sense: affective regulation lowering the symptoms of self-reported distress. It is not completely consistent with the theory but a little bit

Correlation with Observer Rated social distress:

There is a lot more signal:

Dorsal Anterior Cingulate and Anterior Insula are both positively correlated with distress and that is exactly what they would predict (what we should expect);

some prefrontal cortex areas are also correlated

negative association between prefrontal cortex areas and distress. If we look at the two horizontally circled areas, we can see that a lot of data is consistent with their expectations: there is a positive correlation between observer-rated distress and neural activation in affective regions; there is a negative correlation between observer rated stress and activations in areas that are associated with regulatory. So the more prefrontal cortex we see, the less observer related distress we see and this makes sense.

NOTE: The DMPFC is written both in the positive and negative correlation...why?

Maybe there is an activation broadly in the same area but there is not much overlap, but if you look at the Brodman areas, it is 8 and 8.9 and the coordinates are -6, 54, 48 and -12, 42, 42. It is very hard to discriminate between these areas, so how are we getting a significant positive and a significant negative signal from areas that are essentially overlapping?

So for the Self-reported Distress there is not much important to say. In Observer-related distress the only thing that emerges is the dorsal and ventral medial prefrontal cortex which has a positive correlation with observer-related and a negative correlation in the same area with observer-related distress.

Out of that, some results are pretty consistent.

fMRI images

It would have been curious to know if they superimposed on these images the DMPFC that showed a negative and a positive correlation.

Usually these scans are not very useful but in this case it would have, in order to understand this positive and negative correlation of two areas that are almost on top of each other, one positively and one negatively correlated to each other.

The graph below could be interesting: sometimes you do not see a correlation especially with a small sample like 18 subjects, that are driven by one outlier. So in this case it is nice to see the dots on the graph so you can see that it has not been driven just by one subject.

Correlation with discriminatory attribution:

They found a positive correlation with rACC (regulatory region) and a negative correlation with the dACC (affective response).

So what they are concluding here is that attributing exclusion to discrimination leads to lower levels of affective response neurologically and to higher levels of regulatory neurological response.

We can contrast this with the correlation table we saw earlier: we have tons of correlations between attribution and distress. Here we are only seeing a negative correlation between attributions and neurological indicators of affect and distress (Dorsal anterior cingulate activation would be a neurological activation of distress).

So the results are a bit contradicting: attributions lead to lower level of neurological distress (activation of affective areas) but when we look at the correlation between attribution and distress, there is positive correlation

#### SUMMARY

Exclusion: associated with

Increased social pain-related neural activity

Increased emotion regulatory neural activity

Social Distress: associated with

Increased social pain-related neural activity

Decreased emotion regulatory neural activity

Really broad conclusions

Attributing exclusion to discrimination:

reduced social pain-related neural activity

increased emotion regulatory neural activity

This attribution is associated with higher levels of distress that were measured by self-report and maybe even observer-related report even though the correlation is weaker.

So when we look at neural activity we can see that attributing exclusion to discrimination was associated with reduced social pain-related neural activity (neural indicator of distress) but when we look at discriminatory attributions, it is actually positively correlated with both measures of the stress that we had even though this was not significant.

#### TWO MAIN PROBLEMS OF THIS STUDY

The first is a methodological problem that they should have taken into account:

“Each of the two rounds of Cyberball was completed during a functional scan lasting for 2 min 48 sec”

What is the time course of these effects?

They do not look at the time course at all.

To understand this, we can look at Cadinu et al. (2005) study, that looked at the time course in relation to the stereotype threat manipulation.

Stereotype threat = where people's performance – if they belong to a minority group (in this case it was women) – can be manipulated by making negative stereotypes about their group. So if you bring in women and you make negative stereotypes about them, you can influence their performance.

In this specific study, they had two groups: a control group that was told that there was no difference between men and women in math performance and a stereotype threat condition in which the women were told that it has been scientifically demonstrated that women are not as good as men.

These were all students that were going to a scientific high school in Italy, which means that for them math would be an important subject

Then they were asked to solve 7 math problems, and in each of them they were asked what they were thinking.

After the experiment they were debriefed and were told about the real purpose of the study.

In the graph there is the difference between the experimental and the control group: a positive number means that the experimental group did better than the control group, and a negative number ( below the line) means that the experimental group is doing worse than the control.

During the first trial, women that were told that man are better in math, did a bit better than the control group; this disappeared after the second question and then the difference started to increase: they got worse in the third question, in the fourth, in the fifth, a bit better in the sixth and in the last one they were about 40 % worse.

So they went from 10 % better to 40 % worse in the manipulation, but it didn't all happen at the same time: it gradually changed over time, just across 7 math problems. The idea is that there has been an accumulation of the idea that they have been insulted, being told that women are worse than man, they did one good problem, but then they would start to think that maybe they did not do so well. They start to do badly and the worse they do, the more they think about it and the lower the percentage goes.

In the study that we just saw on social exclusion, they did not account for this, so we can imagine that in the Cyberball task, if you take the first 20 sec of being excluded, you are not going to get any difference. But if it is a 2 minutes scan of being excluded, they could separate it in the first 30 sec, the second 30 sec, the third and the fourth and analyze them separately with the idea being that the effects of discrimination don't all happen in the first trial, but it's going to take the person a while to realize that they have being excluded, and then ruminate on it, and after a minute they are going to be offended and in a minute and a half they are really going to be pissed off about it. In this study they just put it all together rather than changes across time.

So the idea is: even if the time is not very long you can think about those as longitudinal data and the way they have been analyzed is that the authors collapsed all of it together.

The second has to deal with how difficult is to interpret these data and part of it goes back to the question regarding the increased or decreased response, which the professor previously answered to saying that there seems to be a two step procedure processes: the offence and the affective response and then the moderation by the attributions that can change this response. The study does not address that.

## SOPHISTICATION MODELS

In this study this was the sophistication model:

They predicted the rejection or social exclusion, with resulting changes in BOLD signal in different brain areas

They expected that distress would be correlated with attributions

They expected that the BOLD signal would be correlated with distress

They expected that the BOLD signal would be correlated with attributions

These are four different sets of analysis but they do not integrate them in any way. And this is their analysis: both their analytical approach and the theoretical structures are at this level, and this is not very sophisticated either analytically or theoretically.

The professor instead would suggest something like that:

In this structure you have something that is more theoretically meaningful: the green lines are what they analyze, the blue are the ones the professor thought should be included.

There are also two more things:

one has to do with the second problem: distress is moderated by attribution, but it actually happens before attributions and so we can divide it in BOLD Distress and Observed Distress. This way we have Rejection that leads to Distress, Distress leads to Attributions and Attributions modify the effect of Rejection on Distress. And so we need distress to happen more than just once as a final outcome: it is a final outcome but it also is what moderates the attributions, and the Attribution influences the change in distress.

The other is the signs (+ and -) which are the actual moderation. They talk about moderation but they do not examine it. This type of analysis could be done in one single analysis called structural equation modeling.

So they could do something much more sophisticated, not only at the theoretical level but also at an analytical level, and in doing this they could also include the time factor by looking at changes across time in the neural signal.