Understanding Others (Brain Mechanisms of Theory of Mind, Empathy, Social Pain and Altruistic Punishment)

Sana Aminnaji

August 8, 2023

Abstract

This article provides a general overview of the basic concepts defined in the field of social neuroscience and the linking works and experiments with neuroeconomics. After a brief investigation on the most popular definitions in this field, then, we summarize the major findings about the neural basis of mentalizing and empathizing and discuss their implications for economics and discuss a number of economical and psychological experiments. Finally, there will be a comprehensive overview of a new experiment inspired by the previous ones in order to determines the presence of the social concepts in personal economical decisions.

1 Introduction

human kind is often engaged in social interactions, in real-life there are several social situations where the thoughts and feelings about self and its relationship with others are elicited. In this regard, understand other people's minds (Theory of Mind), feelings (empathy) and actions (action observation and 'mirror neuron systems') and also other different feelings appeared facing others, have traditionally been the focus of research in social neuroscience.[6]

In the first part of this essay we will define all the basic concepts mentioned previously, then the related works and experiments are going to be introduced containing all the results and conclusions in the end by combining a number of the mentioned experiments a new approach will be presented and the expected results will be discussed.

2 Literature

2.1 Theory of Mind(ToM)/ Mentalizing/ Cognitive Perspective Taking

This term refers to a person's ability to make attributions about mental states such as intentions, desires or beliefs to others (and oneself) and to understand that others have beliefs, desires and intentions that are different from one's own. [6]

Some experts in the field of neuroeconomics presented a different horizon of this concept and related experiments, in which subjects are examined while playing strategic economic games with another person outside the scanning room. Brain activation elicited when playing against intentional versus non-intentional actors is compared. [1]

2.2 Mirror neurons

There are a specific type of neurons first found in monkey's brain. These neurons respond when monkeys see someone else performing a specific action, as well as when the monkey itself performs that particular action. It has been suggested that these mirror neurons represent the neural basis for imitation. Thus, when we imitate someone, we first observe the action and then try to reproduce it. Other findings have revealed that the neural response when observing others' actions is somatotopically organized, stronger for familiar and well-trained actions, and modulated by contextual information.

2.3 Empathy

Broadly speaking, empathy occurs when observing or even simply imagining another person's affective state thereby triggering an isomorphic affective response in the observer. The person experiencing empathy is aware that the source of his or her emotional response is the other person's affective state.

Observation or imagination of another person in a particular emotional state automatically activates a representation of that state in the observer with its associated autonomic and somatic responses. [7]

There are four conditions that can be considered as the crucial conditions of feeling empathy. There is empathy if: (a) one is in an affective state; (b) this state is isomorphic to another person's affective state. (c) this state is elicited by the observation or imagination of another person's affective state. (d) one knows that the other person is the source of one's own affective state. [2]

2.4 Compassion/ Sympathy

Compassion can be defined as the emotion one experiences when feeling concern for another's suffering and desiring to enhance that individual's welfare. [5]. In most definitions of compassion two crucial components are involved: (a) the affective feeling of caring for a suffering person and (b) the motivation to relieve the other person's suffering. [6]

In contrast to empathy, having compassion or sympathy when exposed to the suffering of another person can involve positive feelings of warmth and love and a strong motivation to help rather than negative feelings of distress.

2.5 Social Pain

Research over the past century, from social psychology to behavioural neuroscience, has demonstrated the importance of social bonds for mammalian well-being and survival.

Experiences of social pain — which is defined as the unpleasant experience that is associated with actual or potential damage to one's sense of social connection or social value (owing to social rejection, exclusion, negative social evaluation or loss) — may be processed by some of the same neural circuitry that processes physical pain. The pain signal interrupts ongoing behaviour; promotes quick responses aimed at terminating, reducing or escaping the source of threat; and serves as a punishment based reinforcer to teach organisms to avoid threatening stimuli in the future. [3]

2.6 Altruistic Punishment

Recent models of social preferences define utility functions that incorporate a motive to sanction violations of fairness and cooperation norms. These models predict actual behavior better than do models based on self-interested preferences, lending support to the idea that people are motivated to punish norm violations.Additionally, recent models of the evolution of human cooperation indicate that altruistic punishment has deep evolutionary roots. This suggests that proximate mechanisms evolved that induce humans to bear the cost of punishing others. Because altruistic punishment is not an automatic response, such as the digestion of food, but rather is an action based on deliberation and intent, humans have to be motivated to punish. The typical proximate mechanism for inducing motivated action is that people derive satisfaction from the action. [4]

3 Previous Work

In this section we are going to explain three different experiments in this field, mentioning the conclusions and results each separately.

3.1 Compassion Training[5]

Methods: At pre-training measurement, a total of 94 women were scanned by means of fMRI while viewing 1 of 3 parallel video sets from the newly developed SoVT. Thirty of 33 participants completed the memory training, 28 of 31 participants completed the compassion training, and 30 participants were included as an additional validation sample at pre-training measurement. In the main experiment(both pre and post-training), each fMRI measurement presented participants with 12 HE(High Emotion) and

12 LE(Low Emotion) videos (duration 10–18 s). After each video, participants rated their subjective experience of empathy, positive affect, and negative affect. Every mini-block of 3 HE or LE videos was followed by a fixation cross (displayed for 10 s). The compassion training group attended a 1-day course of loving kindness meditation, followed by guided practices (15–30 min) of strengthening feelings of warmth and care through the visualization of a close loved person. similary, the memory control group attended a 1-day course (6 h) of experience in teaching the method of loci, a technique to memorize items in an ordered sequence. As with the compassion group, the memory group also participated in evening classes or trained at home.

Results: In summary, compassion training (but not memory training) led to significant increases in ratings of positive affect for LE and HE videos and to an increase in empathy ratings for LE videos. Compassion training compared with memory training: (a) increased positive affect, even in response to others' suffering, and (b) was associated with stronger activations in a specific neural network including the mOFC, the pallidum, the putamen, and the VTA/SN—brain regions previously implicated in positive valuation, as well as love and affiliation. The present study actually shows that positive emotions were not only increased in response to everyday life situations, but even in response to witnessing the distress of others. The observed increase in reported positive affect—even when exposed to the suffering of others—suggests that persons trained in compassion can encounter social situations in general and distressing situations in particular with positive, other-oriented affect.

3.2 Social Exclusion[3]

Methods: In this study, participants believed they were playing a virtual game of catch, called 'Cyberball', with two other individuals over the Internet. In reality, the other players were computer-controlled and the game was preset so that participants were first included in the game and then excluded when the two players stopped throwing them the ball. The participant, who is located in the MRI system, depicted by the hand at the bottom of each screenshot, is included in the ball-tossing game with the two other players, depicted in the upper left and right hand corners of the screenshots.

Results: In response to social exclusion versus inclusion, participants showed increased activation in the dorsal anterior cingulate cortex(dACC) and anterior insula(AI). Moreover, greater activity in the dACC was associated with greater feelings of social distress (for example, "I felt rejected") in response to social exclusion. Activity in the dACC and AI was greater during social exclusion than during social inclusion.

3.3 Altruistic Punishment[4]

Methods: Two human players, A and B, interact anonymously with each other. Both players know that they face a human player, and each of them is endowed with 10 money units (MUs). They can increase their income substantially if player A trusts B, and B acts in a trustworthy manner. More specifically, A makes the first decision. He can send his endowment of 10 MUs to B (case 1) or he can keep his endowment (case 2). If A trusts B and sends his endowment (case 1), the experimenter quadruples the amount sent so that B receives 40 MUs. At that moment, B has 50 MUs in total—his endowment plus the 40 units just receivedand A has nothing. Then B has the choice of sending back nothing or half of the 50 MUs. Thus, if B acts trustworthily and sends back half, both players earn 25 MUs, but if B keeps all the money, he earns 50 MUs and A, who trusted B, earns nothing. In case 2, that is, if A does not trust B, both players keep their endowment of 10 MUs. In the cases which B did not act trustworthy, A will be located in a PET scanner and decide whether punish B or not. The punishments are described in four different manners: (intentional and costly (IC))there is a costly decision for A. Every punishment point assigned to B costs one MU for A and reduces B's payoff by two MUs. (intentional and free (IF))the punishment is not costly for A. Every punishment point assigned to B costs nothing for A, whereas B's payoff is reduced by two MUs. (intentional and symbolic (IS))punishment has only a symbolic meaning. Every punishment point assigned to B costs neither A nor B anything. (nonintentional and costly (NC))in which a random device determines B's decision, removing the responsibility for it from player B. Punishment is again costly for both A and B; A loses one MU and B loses two MUs per punishment point assigned to B.

Results: player A exhibits a strong desire to punish B in all three intentional conditions, but this desire is nearly absent in the NC condition. Moreover, player A imposes much higher payoff reductions on B in those conditions in which B intentionally abuses his trust, whereas almost no punishment is

imposed on B in the NC condition. Twelve of 14 subjects punished B if he kept all the money in the IC condition, and all 14 subjects punished B in the IF condition. This contrasts with the NC condition, in which only 3 of 14 subjects reduced B's payoff, and those who did so punished only a little. The activation of the caudate in those conditions in which subjects expressed a strong desire to punish and could indeed punish is particularly interesting in light of this region's prominent role in the processing of rewards. Also a increased blood flow in the thalamus in those conditions in which subjects expressed a strong desire to punish and could punish (IC and IF) relative to the symbolic punishment condition was observed. No thalamus activation was found when IC and IF were compared with the NC condition, in which the desire to punish was absent. Activations in the thalamus have been reported in human neuroimaging studies investigating processing of monetary reward.

4 Method

In this section we describe a new experiment inspired from the explained researches in the previous part. At first the experiment's approach will be defined comprehensively, then we will note some expected results due to the findings in mentioned experiments.

Methods: This experiment is a combination of the Social Exclusion and Altruistic Punishment experiments, we aim to find out the importance of social pain in the economical environments and answer to the question of how it can effects our economical decisions. In order to investigate the social environment we have to generalize the trust game to three participants. Subject A believes that is playing with two other humans, while player B and C's decisions are taken computerized. The first move is done by player A and in each round the player whose turn is will decide pass half of the money to which other player. in the pre-train phase the computers play randomly, thus, at the end of this phase everyone has almost equal amount of wealth. In the next phase we will add the option of no move in each turn and also place participant in the PET system in order to record the activation of different brain regions during the task. In this part the computers try to exclude player A by ignoring him to pass the money to. The question is does the same brain regions responsible of coding social pain(dACC and AI) start to function? If the answer is positive what is the decision made by the participant in this condition. does he try to tempt the others by continue playing or try to escape the sense of rejection by making no move. There can be another analysis about punishment, in this case we can give the opportunity of punishing a player instead of making no move and investigate whether the subject try to tempt the others by trusting them or start to punish them.

Expectations: we expect that the subject feel the social pain similar to the case happened in cyberball game, there can be an approximate estimation about the correlation between the amplitude of pain and the player's strategy. In the punishment phase we predict that the sense of revenge elicits after feeling the social pain by activation of related regions. The subject will give up tempting others and start punishing them after a specific threshold correlated with the subjects characteristics and the intensity of blood flow in the mentioned areas.

5 Conclusion

this essay can be considered as a step in the recent attempts in "neuroeconomics" and the "cognitive neuroscience of social behavior" to understand the social brain and the associated moral emotions. In the first part of this study we tried to give a comprehensive presentation about the most common concepts in these Fields. The second part of this paper was assigned to a brief description of some related practical experiments and the result of each one was mentioned separately. Finally, we introduce a new experiment with the purpose of answering the question of what is the role of pain produced by social exclusion in our economical decisions.

References

[1] Gallagher H.L. Jack A.I. Roepstorff A. Frith C.D. "Imaging the intentional stance in a competitive game". In: *NeuroImage* (2002).

[2] Tania Singer Frederique de Vignemont. "The empathic brain: how, when and why?" In: *TRENDS in Cognitive Sciences* (2006).

[3] Naomi I.Eisenberger. "The pain of social disconnection: examining the shared neural underpinnings of physical and social pain". In: (2012).

[4] Dominique J.-F. de Quervain Urs Fischbacher Valerie Treyer Melanie Schellhammer Ulrich Schnyder Alfred Buck Ernst Fehr. "The Neural Basis of Altruistic Punishment". In: *Science* (2004).

[5] Olga Klimecki Susanne Leiberg Claus Lamm Tania Singer. "Functional Neural Plasticity and Associated Changes in Positive Affect After Compassion Training". In: (2012).

[6] Tania Singer. "The past, present and future of social neuroscience: A European perspective". In: *NeuroImage* (2012).

[7] Ernst Fehr Tania Singer. "The Neuroeconomics of Mind Reading and Empathy". In: *Neuroscientific Foundations of Economic Decision-making* (2005).