King's College London

PG Cert/PG Dip/MSc Examination

7PADSPRII Research Skills: Synoptic Project

Final word count: 4000
(word limit: 4000 words)

Please note that a 5% penalty will be applied when essays exceed the word limit. There is no additional 10% allowance in this submission. Markers will stop reading once they reach the word limit. This will have an impact on the grade awarded, so please respect the word limit.

Plagiarism Statement

This exam script was written by me using my own words. It does not use text from work I, or any other student/researcher, has already submitted for formative or summative assessment in another module, another course, publication or source anywhere in the world.

The university Academic Honesty and Integrity Policy can be found here:
http://www.kcl.ac.uk/governancezone/Assessment/Academic-Honesty-Integrity.aspx

Student ID Number 19071714 Date 19 April 2022

Plan

The role of the insula in processing fair and unfair distributions in experimental games in the light of the reward-risk prediction model

Humans are social animals (Tomasello, 2014). Interactions with others are what make our civilisations and progress possible. In this context, being able to face the uncertainty of trusting a stranger to engage in transactions is crucial and fairness is one of its essential components (Henrich et al., 2010). A good understanding of how fairness works in the human brain is therefore an important topic for various disciplines including psychology, economics and neuroscience and will have broad societal impacts, both for healthy and clinical population (Kishida et al., 2010; Montague et al., 2012).

Experimental tasks based on social dilemmas such as in the Ultimatum Game (UG) (Güth, et al., 1982) were used in laboratories to look at behaviours when facing unfair distributions. The emergence of the interdisciplinary field of neuroeconomics (Camerer, 2005) made possible to further examine these behaviours from a neurobiological perspective using neuroimaging techniques and draw models from experimental economics and psychology to understand the underlying processes involved (Fehr & Kraijbich, 2014).

In the case of the UG, neuroimaging results found correlation between the activation of the insula in response to unfair offers (Sanfey et al., 2003). Nevertheless, given the different roles played by the insula and especially in signaling uncertainty and risk

(Preuschoff et al., 2008), it has been questioned whether this region could be necessarily responding to fairness itself in the context of the UG (Fehr & Krajbich, 2014).

This critical review aims to outline a brief narrative of the evolution of research to identify the role of the insula in the relationship between fairness and risk prediction and risk prediction error in the context of experimental games as well as highlighting gaps and proposing some directions for future research.

2.1.1. Valid consent

2.1.2. Deception

2.2. Guidelines applied to neuroimaging studies

2.2.1. Contraindications, information and screenings

2.2.2. Confidentiality

2.2.3. Incidental finding and giving advice

Synoptic Project 2

Further analyses will be done to compare and contrast the measurements performed in the neuroimaging studies to highlight overlaps and gaps.

The methodology used in the experimental designed will be examined and limitations such as reverse inference, ecological validity, publication bias, underpowered studies will be discussed.

Part 1 – Methods and procedures

1.1. fMRI data collection – consequences and limitations

1.1.1. Sample sizes in fMRI studies

1.1.2. Combining measurements methods

The role of the anterior insula in processing fair and unfair distributions in ultimatum games in the light of the reward-risk prediction model

–

A review of methods and analysis

Introduction

The role played by the anterior insula (AI) in fairness-related experimental games like the Ultimatum Game is still unclear (Fehr & Krajbich, 2014). As discussed in the first part of the Synoptic Project (SP1), fMRI studies found an increased activity of the anterior insula (AI) in healthy controls in presence of unfair distributions when playing the UG (Sanfey et al., 2003). Given that the activation of the insula has been found to be correlated with emotional awareness (Craig, 2009), an early interpretation of this finding suggested that the increased activity of the AI is linked to the possible strong emotions elicited by the unfair offers. Another interpretation proposed that the activation of the AI could correspond to norm violation detection (Xiang et al., 2013) which could be using the same error detection mechanism found in the non-social context of the reward/risk prediction model (Preuschoff et al., 2008).

However, a few studies based on different experimental designs and different populations found inconsistent results (Kirk et al., 2011; Yoder & Decety, 2020). This could suggest that the AI might not necessarily respond to fairness *per se* as mentioned in SP1.

The research question is therefore to clarify the role of the insula found in fairness studies, in light with the reward/risk prediction paradigm, in the context of experimental game designs on healthy controls. The specific aim of this part two of the Synoptic Project is to appraise the methods and analysis of the studies mentioned in SP1, to further compare and contrast them, present their limitations in terms of methods and analyses as well as highlight the gaps, and suggest directions for future research design.

Part 1 – Methods and procedures

In this part, the methods used to study how the brain is responding to fairness will be discussed, especially the measurements of the brain activity with neuroimaging techniques (1.1.) while participants are involved in socio/economic tasks (1.2.). The measurements and data collection will be presented as well as some limitations that need to be addressed for future study designs.

1.1. FMRI data collection – consequences and limitations

As seen in Table 1., the large majority of the studies rely on fMRI neuroimaging techniques for data collection. This non-invasive imaging method is often used in neuroeconomics studies as it allows recordings of the entire brain when participants are asked to perform behavioural tasks. Its high spatial resolution and good temporal resolution make this technique the most versatile and common neuroimaging technique used as observed in this literature review (see Table 1). Nevertheless, as with any other technique, fMRI comes with its limitations. This section will review one of the consequences of choosing fMRI on the sample size (1.1.1.) and the strengths and weaknesses of this technique as well as one approach to mediate the effects of its intrinsic constraints (1.1.2.).

Study	Design	Participants	Task	Measurements	Analysis	Results
Sanfey et al., 2003	One group of healthy controls	*n* = 19 (11 females and 8 males) Origins not mentioned	Classic UG	fMRI	General linear model (GLM)	t-statistic show greater activation of bilateral AI following unfair as compared with fair offers (P < 0.001). Participants with stronger AI activation to unfair offers rejected a higher proportion for these offers (right insula: correlation coefficient r = -0.45, P = 0.025, one-tailed; left-insula: r = -0.39, P = 0.05, one-tailed). Across participants, activation in the right AI was significantly greater in response to unfair offers that were rejected (P = 0.028, one-tailed).
Preuschoff et al., 2008	One group of healthy controls	*n* = 19 (9 females and 10 males) Origins not mentioned	Card game	fMRI	GLM	Two types of activations found: - Activation in AI correlates with risk prediction (random effects, df=18; p<0.0005) - Activation in AI correlates with risk prediction errors (random effects, df=18; p<0.0005) and increasing risk prediction errors was correlated in an increased activation (right insula, p < 0.001; r^2 = 0.65). The activation that correlates with the risk prediction after seeing the first card is delayed (about 5s), in an area in insula that is slightly superior and anterior.
Tabibnia, 2008	One group of healthy controls	*n* = 29 (18 females and 11 males)	UG	fMRI	GLM	Increased AI activity (compared to resting baseline) when unfair trials were rejected (x = 36, y = 18, z = -8), t(11) = 4.39, Prep > .99, d = 2.95.

		University undergraduates				
Kirk et al., 2011	Two groups: healthy controls and experienced meditators	Controls = 40 Meditators = 26 Origins not mentioned	Classic UG	fMRI	Univariate GLM	In meditators, greater activation for unfair offers compared to fair offers in the bilateral posterior insula, left mid-anterior insula. In controls, greater activity for unfair offers compared to fair offers in bilateral anterior insula. An 5-mm spherical ROI in the bilateral AI found that the right and left AI showed a significantly higher activity for unfair versus fair offers (left paired t = 3.4, p , 10-4. Right paired t = 2.6, p < 0.008) while no difference was found in meditators (left: paired t = 1.3, p < 0.2. Right: paired t = 0.7, p < 0.4). Left AI showed a significant difference between unfair offers in controls compared to meditators (two sample t = 1.9, p < 0.05) but no difference in the right AI between unfair offers in controls compared to meditators (two sample t = 0.8, p < 0.4). Stronger activation for unfair offers shows lower acceptance rate in controls (Left: R = -0.41, p = 0.004, one-tailed. Right: R = -0.45, p = 0.002, one-tailed) while activty in AI did not correlate with acceptance rates for unfair offers in meditators (Left: R = -0.23, p = 0.12, one-tailed. Right: R = -0.31, p = 0.06, one-tailed).
Xiang et al., 2013	One group of healthy controls	*n* = 127 (71 females and 56 males)	Classic UG	fMRI	Univariate GLM	Activity in the AI was correlated with variance prediction error.

		Origins not mentioned				The ROI analysis showed that the BOLD responses of the AI displayed a U-shape activation pattern (found in Preuschoff et al., 2008) to the norm prediction error.
Corradi-Dell'Acqua et al., 2013	One group of healthy controls	*n* = 23 (9 females and 14 males) Origins not mentioned	Modified version of the UG	fMRI	Univariate GLM	Left AI found to be active both in myself and third-party condition
Yoder & Decety, 2020	One group of healthy controls	Study 1 – *n* = 27 (15 females and 13 males) Study 2 – *n* = 36 (23 females + 13 males) Origins not mentioned	Modified version of the UG	FMRI + EEG	Univariate GLM & Multivariate pattern analysis	Whole-brain analysis found AI activity in Self-Fairness conditions in response to unfair offers but NOT in Other-Fairness conditions.

Table 1. Comparison of experimental designs

1.1.1. Sample size

Choosing a study design that involves the use of fMRI has intrinsic implications, especially on the sample size. As shown in Table 1, in almost all the studies using fMRI, the sample size is small (less than 100 participants), with a mean of 40 and a median of 28 participants for these studies of interest. This limited sample size would lead to a low statistical power and an increased likelihood of false positives. This will have a direct impact on the reliability and the replicability of the studies (Turner et al., 2018).

The limitation of the number of participants is to be explained by the high expense of the fMRI scanners. Traditionally, group fMRI studies are indeed composed of between 10 to 30 subjects (Button et al., 2013). Previous guidelines estimated a minimum sample of 24 subjects to obtain reliable results with a rate of true positives of more than 80% (Desmond & Glover, 2002). A power calculation would determine the adequate sample size (Mumford, 2012), although this step was rarely mentioned in fMRI studies in general (Szucs & Ioannidis, 2020) or in the studies of interest. This has been said to lead to frequent underpowered studies (Button et al., 2013). Nevertheless, the limited sample size can be counterbalanced by the large number of trials conducted to offset the noise of fMRI datasets (Dale and Buckner, 1997). Noise in fMRI studies is linked to BOLD responses that are not related to brain activity. This can include noise related to participants' motion (in the scanner), physiology (e.g. breathing that can change the blood flow) (Krüger and Glover, 2001) as well as the scanner itself (e.g. magnetic field instabilities). Given these large sources of noise, one approach is to repeat the trials to allow signal averaging (Dale and Buckner, 1997). This repetition of identical trials might therefore help to mediate the effects of the small sample size and obtain a better estimate of the hemodynamic response function (Schram et al., 2019).

In conclusion, although limited, the sample size of the studies cited in the literature review complies with the guidelines for fMRI studies.

1.1.2. fMRI strengths and weaknesses and the advantage of combining neuroimaging techniques

Thanks to its flexibility, fMRI neuroimaging techniques have extended the understanding of behavioural findings in decision-making and social interactions in general (Rilling & Sanfey, 2011) and fairness in particular (Sanfey et al., 2003). As seen in Table 1., fMRI has been widely used to study brain activity in the context of the UG. Its main advantage lies in its noninvasiveness as well as its high spatial resolution. Nevertheless, like any technique, several inherent constraints in the use of fMRI could be mentioned (Logothetis, 2008), including its low temporal resolution, limited to the speed of the hemodynamic response (Schram & Ule, 2019), its vulnerability to noise as mentioned above and false positives (Eklund et al., 2016) as well as the issue of the interpretation of the results (see 2.2. below for further discussion).

To overcome these limitations, a promising approach has emerged by combining fMRI with other imaging techniques. Associating techniques with complementary strengths would mediate the weaknesses of each method and lead to additional insights. Thus, one way was to combine the fMRI low temporal resolution with the EEG high temporal resolution (Philiastides & Sajda, 2007). This approach has, for instance, been chosen by the most recent study cited in Table 1. (Yoder et al., 2020).

1.2. Study design and experimental tasks

This section will address the question of using the UG as a task to study fairness (1.2.1) and the question of the population chosen and the cross-cultural study design (1.2.2.).

1.2.1. Measuring fairness with the UG

The UG has extensively been used to study social preference and fairness (Gabay et al., 2014) since the first experiment by Güth and colleagues (1982). The majority of the studies selected in this project relied on the classic design of the UG as the behavioural task (Table 1.). Like other economic games, the UG has the advantage to allow direct behavioural observations in a standardised settings which makes it less prone to faking compared to self-report assessments in behavioural designs (Baumert et al., 2014) and permit the measurements of task-related brain activity (Sanfey et al., 2003; Gabay et al., 2014).

A question has been raised as to whether the UG could be a relevant measure of the notion of fairness. Thus, in these studies, the concept of fairness is equivalent to fairness *in the context of the U*G. It has been said that in addition to the limited amount of research investigating the psychometrics properties of the UG, there was no significant positive correlations found with dispositions related to fairness concerns (Baumert et al., 2014). The authors suggested therefore that rejection decisions in the UG might not reflect concerns for fairness *per se* but the belief that a person may have about what they legitimately deserve (Baumert et al., 2014). The question of being able to measure fairness *per se* is indeed a central point. Experimental designs in a laboratory are faced with the question of the ecological validity and have to find a balance between the necessity of a standardised experimental setting and the representation of an accurate real-life behaviour in a naturalistic environment. Some authors pointed out that the UG would only represent fairness as a synonym for an equal distribution of money and wonder to what extend the various models developed based on the UG might include the concept of fairness outside the game itself (Debove et al., 2016). Thus, the authors highlighted that in daily life, fairness might refer to more than just equal distribution but also divisions that matched individual contributions (Adams, 1963).

Furthermore, as seen in SP1, results from fMRI studies involving the UG might not necessarily correspond to fairness (Fehr & Kajbich, 2014). It has been suggested that the increased activity of the insula for unfair offers noted by Sanfey and colleagues (2003) might represent a negative prediction error for the participants own's payoff as subsequently found by Preuschoff and colleagues (2008). The use of the classical design of the UG was not able to differentiate between the two concepts. As discussed in SP1, looking at games where fairness and the player's own payoff are decorrelated might help to clarify this question (Fehr & Kajbich, 2014). Thus, as seen in Table 1., a few studies have used variations of the classic UG to attempt to decorrelate participants' payoff with fairness, where participants could play for themselves in one condition and on behalf of other players in another (Corradi Dell'Aqua et al., 2013) or where Self-fairness could compete with Other-fairness (Yoder & Decety, 2020). However, there is no study from this review where fairness and risk prediction and risk prediction error have directly been measured in the same experiment. Incorporating a game based on the model used in the study by Preuschoff and colleagues (2008) might be a way to directly compare and distinguish between these two models.

1.2.2. Population and the value of cross-cultural experiments

The need to take into account different cultures has been a point of attention in behavioural studies to avoid biases in studies that included predominantly participants from Western, Educated, Industrialised, Rich and Democratic (WEIRD) countries (Henrich et al., 2006). This is all the more crucial with culturally dependent domains like fairness which is subject to high variability among populations (Henrich et al., 2010). Thus, behavioural studies found that although the principle of fairness is present in many cultures, some variations have been observed (Blake et al., 2015; Cochard et al., 2021).

The participants' origins were rarely mentioned in the studies of interest (Table 1.). One study did mention participants' being university undergraduates (Tabibnia, 2008). Future neuroimaging studies would benefit from including a more variety of participants both in terms of countries and cultures.

Moreover, the majority of the studies involved one group of healthy controls, except for one study comparing controls to experienced meditators (Table 1., Kirk et al., 2011). Future designs that allow comparison between participants with brain dysfunction/manipulation or different cognitive abilities would also provide valuable insights (Fehr & Krajbich, 2014).

Part 2 – Analysis and reporting

Neuroeconomics results relying on neuroimaging data are based on specific data analysis techniques (2.1.) and can present specific challenges when it comes to their interpretation (2.3.) such as the issue of reverse inference (2.2.).

2.1. Univariate general linear model versus supervised multivariate pattern analysis

2.1.1. The classical approach of hypothesis-driven analysis

In neuroeconomics studies, the classical approach for data analysis relies on a variant of regression analysis called general linear model (GLM; Friston et al., 1994). This data analysis approach is based on a hypothesis testing-driven model where the researchers try to find whether A correlates with B in C (with A being the behaviour of interest, B the data type (BOLD hemodynamic response in the case of fMRI studies) and C the brain region).

As seen in Table 1., almost all the studies mentioned in the literature review follow this classical approach of GLM for data analysis. Most of them used mass univariate linear regression to estimate the relationship between the variable of interest and the physiological data recorded at multiple locations.

All the traditional statistical tools (t-tests, ANOVAs, etc...) can then be used in this GLM at each source of the neural signal (i.e. voxels in fMRI) as noted in Table 1.

2.1.2. The exploratory approach of the multi-voxel pattern analysis (MVPA)

Another more recent approach for data analysis is based on a supervised machine-learning technique (Poldrack and Farah, 2015). This method uses supervised learning algorithms to classify data according to predetermined classes the program has been trained to detect. This technique presents the advantage of leveraging algorithms to identify recorded signals associated with cognitive processes of interest to avoid making assumptions about the underlying mechanisms. This might be particularly insightful when two tasks might be engaging the same brain region.

One study took advantage of the two techniques by combining univariate analysis with the model-based approach to first characterise the hemodynamic signals and then distinguish the signals (Yoder & Decety, 2020). In this study, the MVPA used was based on the technique of linear support vector machine (SVM) which allows to do both regression and supervised classification. Interestingly, this study highlighted that the AI was not found to be activated in the Other-fairness condition (Table 1.) which might suggest that AI might not be involved in processing fairness *per se* but could be linked with other processes. Further studies using tasks that can distinguish between fairness and risk prediction models could take advantage of the strengths of the MVPA approach.

2.2. The issue of reverse inference

Interpretation of fMRI results, especially in the classical hypothesis-driven data analysis approach, faces the common challenge of *reverse inference* (RI) (Poldrack, 2006). The RI refers to the way of inferring mental processes and behaviours from the activation of certain parts of the brain. As opposed to *forward inference* where an activation of certain brain regions is observed from participants' behaviours, in RI, cognitive processes are, *a contrario*, inferred from neuronal activity. In this latter case, activations of the brain region are interpreted as evidence for the involvement

of a cognitive process. This inference is often based on previous studies which reported that the region of interest was previously found to be correlated with these processes. Such interpretation presents a logical fallacy as the knowledge that a certain mental process *X* involving a certain brain region *Y* is not a sufficient evidence to conclude that the participant is necessarily engaging in this mental process *X.* Indeed, this region Y could be associated with other different functions as well.
It can be an issue with regions that are found to be commonly recruited like the insula (Duncan & Owen, 2000; Nelson et al. 2010; Yarkoni et al. 2011).
The AI has been reported to be involved in various cognitive functions such as awareness of pain, body movements, emotions, prediction of risk, uncertainty and anticipation (Craig 2009, Singer et al. 2009). When Sanfey and colleagues (2003) reported that the AI was also found to be activated when participants of an UG are presented with unfair offers, the inference that unfairness in the context of the UG is found to be correlated with the activation of the AI was a valid interpretation. What could be discussed however is the subsequent interpretation suggesting that as the AI has been found to be previously activated in presence of strong emotions, pain and distress, the activation would reflect the participants' negative emotions. This interpretation has been repeated by many studies looking at the UG without being necessarily directly tested, although it can be mentioned that some subsequent studies using skin conductance response (SCR) were able to correlate the results with emotional response (Van't Wout et al., 2006). Studies combining SCR and fMRI would be a way to directly test the accuracy of the model suggested. Nevertheless, as mentioned above, the study design would need to ensure to decorrelate participant payoff with fairness to avoid confounding results.

2.3. Timeline and signal comparison

Without being able to show experimental evidence in the study of interest that there is an activation of this brain region following a particular mental process, there might be as many possible interpretations of the role of this region as there are previous studies finding correlation with it. This is especially the case when a closer look at the results might suggest some overlaps. This section aims to propose a novel perspective of the results of the studies led by Sanfey (2003) and Preuschoff (2008) by comparing the timeline and the signals found in the studies.
A recall of Sanfey and colleagues' experimental timeline show that for each trial (that lasted in total 36s), after a fixation cross of 12 seconds, the participant of the UG (playing the role of the Responder) is presented with 1/ a picture and a name of its partner which lasted 6 seconds, 2/ then the offer made by the Proposer which lasted 6 seconds, 3/ the alternative of accepting or rejecting this offer for 6 seconds, time during which he/she has to decide, and 4/ the outcome of the game.

In Preuschoff and colleagues' study, the experimental paradigm based on a card game was designed to study the participant's risk prediction. In that case, two cards were drawn consecutively, with the participant asked to predict whether the second card would be higher than the first; the participant's predictions occurring twice, one before the first card is drawn and one again before the second card is shown. The risk prediction and the risk prediction error would appear respectively at these two timings. Interestingly, the interval between the display of the first card and the second card is around 7 seconds.

When comparing the experimental design and especially the timeline along with the signal activation, it can be suggested an overlap of the sequence. This might contribute to suggest that risk prediction and risk prediction error might be found in Sanfey's experimental design.

However, this proposition has not been tested yet. Further studies could design an experiment where both fairness and risk prediction could be measured in the same trial.

Conclusion

As presented in this critical review (Table 1.), fMRI studies are a powerful imaging method which has allowed a better understanding of the neural processes of social behaviours such as rejections of unfair offers in the UG. Its non-invasiveness and high spatial resolution make it one of the most versatile imaging techniques (Shram & Ule, 2019). However, like any technique, fMRI is subject to some intrinsic constraints including noise and low temporal resolution. The small sample size of the studies also made them less replicable (Turner et al., 2018). An explicit reference to the power calculation as well as including a more diverse population would improve the effectiveness and generalisation of the results for this research question (Szucs & Ioannidis, 2020).

As discussed, the evolution of fMRI analysis from a hypothesis-driven analysis to an exploratory approach has permitted to find new insights. Nevertheless, interpretation of fMRI data demands certain caution, especially in terms of reverse inference (Poldrack, 2006). It is crucial to avoid backward reasoning and draw conclusions from the observation of brain activity instead of previous results, especially with regions that are multipurpose domain-general like the AI.

From this critical review, it is still unclear whether the results presented could be linked to the concept of fairness *per se*. The use of the UG as the task limits the results to the context of equal distribution in economic games in an experimental setting (Debove et al., 2016). Furthermore, the comparison of the signals with subsequent studies shows possible confounding elements in the activation timeline. Thus, future studies where fairness and reward/risk prediction models could be measured in the same trial and participants' own interest decorrelated with fairness, would lead to a better understanding of the involvement of the AI in one of the essential components of trust and human cooperation (Fehr & Schmidt, 1999).

Reference list

Adams, J. S. (1963). Towards an understanding of inequity. *The journal of abnormal and social psychology*, *67*(5), 422

Baumert, A., Schlösser, T., & Schmitt, M. (2014). Economic games. *European Journal of Psychological Assessment*.

Blake, P. R., McAuliffe, K., Corbit, J., Callaghan, T. C., Barry, O., Bowie, A., ... & Warneken, F. (2015). The ontogeny of fairness in seven societies. *Nature*, *528*(7581), 258-261.

Cochard, F., Le Gallo, J., Georgantzis, N., & Tisserand, J. C. (2021). Social preferences across different populations: Meta-analyses on the ultimatum game and dictator game. *Journal of Behavioral and Experimental Economics*, *90*, 101613.

Corradi-Dell'Acqua, C., Civai, C., Rumiati, R. I., & Fink, G. R. (2013). Disentangling self-and fairness-related neural mechanisms involved in the ultimatum game: an fMRI study. *Social cognitive and affective neuroscience*, *8*(4), 424-431.

Craig, A. D., & Craig, A. D. (2009). How do you feel--now? The anterior insula and human awareness. *Nature reviews neuroscience*, *10*(1).

Dale, A. M., & Buckner, R. L. (1997). Selective averaging of rapidly presented individual trials using fMRI. *Human brain mapping*, *5*(5), 329-340.

Debove, S., Baumard, N., & André, J. B. (2016). Models of the evolution of fairness in the ultimatum game: a review and classification. *Evolution and Human Behavior*, *37*(3), 245-254.

Desmond, J. E., & Glover, G. H. (2002). Estimating sample size in functional MRI (fMRI) neuroimaging studies: statistical power analyses. *Journal of neuroscience methods*, *118*(2), 115-128.

Duncan, J., & Owen, A. M. (2000). Common regions of the human frontal lobe recruited by diverse cognitive demands. *Trends in neurosciences*, *23*(10), 475-483.

Fehr, E., & Krajbich, I. (2014). Social preferences and the brain. In *Neuroeconomics* (pp. 193-218). Academic Press.

Fehr, E. & Schmidt, K. M. A theory of fairness, competition, and cooperation. Q. J. Econ. 114, 817–868 (1999)

Friston, K. J., Holmes, A. P., Worsley, K. J., Poline, J. P., Frith, C. D., & Frackowiak, R. S. (1994). Statistical parametric maps in functional imaging: a general linear approach. *Human brain mapping*, *2*(4), 189-210.

Gabay, A. S., Radua, J., Kempton, M. J., & Mehta, M. A. (2014). The Ultimatum Game and the brain: A meta-analysis of neuroimaging studies. *Neuroscience & Biobehavioral Reviews*, *47*, 549-558.

Haxby, J. V., Connolly, A. C., & Guntupalli, J. S. (2014). Decoding neural representational spaces using multivariate pattern analysis. *Annual review of neuroscience*, *37*, 435-456.

Henrich, J., McElreath, R., Barr, A., Ensminger, J., Barrett, C., Bolyanatz, A., ... & Ziker, J. (2006). Costly punishment across human societies. *Science*, *312*(5781), 1767-1770.

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). Most people are not WEIRD. *Nature*, *466*(7302), 29-29.

Kirk, U., Downar, J., & Montague, P. R. (2011). Interoception drives increased rational decision-making in meditators playing the ultimatum game. *Frontiers in neuroscience*, *5*, 49.

Krüger, G., & Glover, G. H. (2001). Physiological noise in oxygenation-sensitive magnetic resonance imaging. *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine*, *46*(4), 631-637.

Logothetis, N. K. (2008). What we can do and what we cannot do with fMRI. *Nature*, *453*(7197), 869-878.

Mumford, J. A. (2012). A power calculation guide for fMRI studies. *Social cognitive and affective neuroscience*, *7*(6), 738-742.

Nelson, S. M., Dosenbach, N. U., Cohen, A. L., Wheeler, M. E., Schlaggar, B. L., & Petersen, S. E. (2010). Role of the anterior insula in task-level control and focal attention. *Brain structure and function*, *214*(5), 669-680.

Philiastides, M. G., & Sajda, P. (2007). EEG-informed fMRI reveals spatiotemporal characteristics of perceptual decision making. *Journal of Neuroscience*, *27*(48), 13082-13091.

Poldrack, R. A. (2006). Can cognitive processes be inferred from neuroimaging data?. *Trends in cognitive sciences*, *10*(2), 59-63.

Poldrack, R. A., & Farah, M. J. (2015). Progress and challenges in probing the human brain. *Nature*, *526*(7573), 371-379.

Poline, J. B., & Brett, M. (2012). The general linear model and fMRI: does love last forever?. *Neuroimage*, *62*(2), 871-880.

Preuschoff, K., Quartz, S. R., & Bossaerts, P. (2008). Human insula activation reflects risk prediction errors as well as risk. *Journal of Neuroscience*, *28*(11), 2745-2752.

Rilling, J. K., & Sanfey, A. G. (2011). The neuroscience of social decision-making. *Annual review of psychology*, *62*, 23-48.

Sanfey, A. G., Rilling, J. K., Aronson, J. A., Nystrom, L. E., & Cohen, J. D. (2003). The neural basis of economic decision-making in the ultimatum game. *Science*, *300*(5626), 1755-1758.

Singer, T., Critchley, H. D., & Preuschoff, K. (2009). A common role of insula in feelings, empathy and uncertainty. *Trends in cognitive sciences*, *13*(8), 334-340.

Schram, A., & Ule, A. (Eds.). (2019). Handbook of research methods and applications in experimental economics.

Szucs, D., & Ioannidis, J. P. (2020). Sample size evolution in neuroimaging research: An evaluation of highly-cited studies (1990–2012) and of latest practices (2017–2018) in high-impact journals. *NeuroImage*, *221*, 117164.

Turner, B. O., Paul, E. J., Miller, M. B., & Barbey, A. K. (2018). Small sample sizes reduce the replicability of task-based fMRI studies. *Communications Biology*, *1*(1), 1-10.

Van't Wout, M., Kahn, R. S., Sanfey, A. G., & Aleman, A. (2006). Affective state and decision-making in the ultimatum game. *Experimental brain research*, *169*(4), 564-568.

Woo, C. W., Krishnan, A., & Wager, T. D. (2014). Cluster-extent based thresholding in fMRI analyses: pitfalls and recommendations. *Neuroimage*, *91*, 412-419.

Xiang, T., Lohrenz, T., & Montague, P. R. (2013). Computational substrates of norms and their violations during social exchange. *Journal of Neuroscience*, *33*(3), 1099-1108.

Yarkoni, T., Poldrack, R. A., Nichols, T. E., Van Essen, D. C., & Wager, T. D. (2011). Large-scale automated synthesis of human functional neuroimaging data. *Nature methods*, *8*(8), 665-670.

Yoder, K. J., & Decety, J. (2020). Me first: Neural representations of fairness during three-party interactions. *Neuropsychologia*, *147*, 107576.