Universiteit Leiden
The Netherlands

Opleiding Informatica & Economie

Predicting the outcome of

behavioral economic games by analyzing robot images.

Elske van Zwol

Supervisor:
Dr.ir. R.E. de Kleijn
Second reader:
Dr.ir. D.J. Broekens

BACHELOR THESIS

Leiden Institute of Advanced Computer Science (LIACS)
www.liacs.leidenuniv.nl 12/07/2024

Abstract

Behavioral economic games can help us understand human behavior better. A dictator game is a type of behavioral economic games that is used to measure altruism and sharing behavior. A previous study looked at dictator game behavior with images of robots, to study human-robot interactions (6). They found that you could predict the dictator game offers by looking at three components: likeability, anthropomorphism, and utility. This study further builds on this previous research by using image feature extraction methods to analyse the digital images of the robots used in the previous study. In this study, we then determine if there is a relationship between these features, the dictator offers and questions of the previous study. We found that local contrast as digitally extracted from the image, has a significant correlation with the dictator game offer. We also found that all extracted features have one or more significant correlations with questions from the previous study. We can thus say that based on the outcomes and within the scope, extracted features from digital images could be useful to make predictions around dictator game behavior. These findings may help reduce cost to evaluate robot design along various features.

Contents

1 Introduction

Behavioral economic games help us model and understand human behavior. Within these behavioral economic games, dictator games are well known. They are particularly useful to help us understand altruism and sharing behavior (7). During a dictator game participants are given an amount of money which they can either keep all to themselves or share with their opponent.

This thesis is built upon previous research by de Kleijn et al. (6). The researchers did a study on dictator games with commercially available robots. In the study participants were shown images of 18 different commercially available robots and had to answer 12 questions about each robot (table 1), these questions were used to understand how participants felt about each robot. After answering the questions the participants had to give a final dictator game offer per robot. The researchers did this to determine what drives the outcome of the dictator game.
They found, using a Principal Component Analysis, that the outcome of these dictator games depended on three components: an Anthropomorphism index (AI), a Likeability index (LI), and an Utility index (UI). The first component includes physical and cognitive similarity to humans. The second component represents general likeability and the third component represents a measure of utility. The twelve questions were allocated to these three components.
This study aims to build on this research by using computer science methodologies to analyse and predict the outcome of the dictator game using digital photographs of robots. The research question is formulated below.

Can we predict from images alone, how much money people allocate to robots?

In order to use digital images, features need to be extracted. Image feature extraction is a common topic in computer science. These features can help you understand an images better. In this research the goal is to see if the outcome of dictator games can be predicted by looking at features from digital images of robots. This research will focus on the following features; image complexity, brightness, global contrast and local contrast. The reason for choosing these features is twofold. On the one hand these are characteristics of any image that have been proven to impact aesthetic judgement of photographs and on the other hand existing coding is available. This is further addressed in the methods section. In addition, all of these features can be captured in a single digit, which is helpful for statistical analysis.

Being able to understand how different features drive AI, LI and UI, has numerous practical and economical applications. In robot production it can be used to optimize certain features in the industrial design process, rather than having to prototype multiple robot models. This can be done digitally which saves both time, money and reduces the need to use human test panels. Furthermore, similar design advantages should work in any industry where the ultimate product design needs to have a certain appeal to humans. Examples could be the car industry or various home appliances. Combining behavioral economic theory with computer science technologies in this way can contribute to better outcomes. The scope of this research is within a specific sample set of robots and thus any conclusions of this research will apply on this group. However, the methodology used in this research, should be relatively easily applied to other data sets or test subjects.

1.1 Thesis overview

This section contains the introduction; Section 2 discusses related work and some definitions; Section 3 describes the methods used in this study; Section 4 states the results found in this research; Section 5 concludes the research.

This research was done as a bachelor thesis at LIACS and FSW, with the help of Roy de Kleijn and Joost Broekens. I would like to thank both for all their help and support during the completion of this bachelor thesis.

2 Related Work

2.1 Human-Robot Interactions

Compared to human-human dictator games, little research has been done on human-robot dictator games. However research on Human-Robot Interactions (HRI) and Human-Agent negotiation is available. The research on HRI is mainly focused on the amount of trust humans have in robots versus other humans. The following studies do not contain dictator games, but give us insights into HRI.
Jessup et al. (11) looked at the differences in affect when participants were paired with a human or a robot while playing a modified version of the investor game. They found that there were no differences in affect between partner types when the partner performed a trustful behavior, however the findings did suggest that people are more sensitive to distrust behaviors that are displayed by a robot versus those by a human.
Razin et al. (20) looked at bridging the gap between HRI and game theory. They stated that both fields have their own theories of trust and started to bring the two together. According to their research HRI can learn from human-human interactions. They can do this by either studying game theory or by looking at the relation between the underlying constructs of trust in HRI research and game-theoretic trust.
In the research by Maggioni et al. (15) students were given random opponents, either a human or a humanoid robot, and had to play the prisoners dilemma. The research showed how communication could have strong influences on the decision making.

The following two studies looked at Human-Agent negotiation.
Lee et al. (14) wanted to see if people's behavior changed during the dictator game, ultimatum game, and negotiation against artificial agents while varying agents' minds on two dimensions of the mind perception theory: agency (cognitive aptitude) and patiency (affective aptitude) via descriptions and dialogs. During their dictator game study they found that agents with emotional capacity garnered more allocations.
Jonker et al. (12) state that AI can be used to assist humans during negotiation. They state that while humans are better at understanding the context and fluctuations in human-human interactions, they can be influenced by emotions and have difficulty handling the complexity of negotiation spaces. As a next step of the research they wanted to develop a negotiation support system.

As mentioned before de Kleijn et al. (6) did research on human-robot dictator games. They found that the outcome of the dictator game could be predicted by the three components, Anthropomorphism index (AI), a Likeability index (LI), and an Utility index (UI). The questions are allocated to these three components, they are shown in table 1.

Variable	Question	Component
care_family	Would you let this robot take care of your family?	Utility
cook	Would you let this robot cook for you?	Utility
creepy	How creepy is this robot?	Likeability
feel_emo	Can this robot feel emotions?	Anthropomorphism
friendly	How friendly is this robot?	Anthropomorphism
like	How much do you like this robot?	Likeability
phys_sim	How physically similar is this robot to a human?	Anthropomorphism
plan_indep	Can this robot plan its own actions independently?	Anthropomorphism
think_hum	Does this robot think like a human?	Anthropomorphism
touch	Would you like to touch this robot?	Likeability
vacuum	Would you let this robot vacuum your house?	Utility
want_to_have	Would you want to have this robot?	Likeability
dictator_offer	Amount of money offered in the dictator game	

Table 1: The questions used in the research by de Kleijn et al. (6)

2.2 Image properties

Chikhman et al. (3) found that for outline images the best measure of complexity is the number of turns. The term "turns" is used as an overarching term for "points", "angles" or "sides" of image outlines (2). However, the number of turns is not always the best measure of complexity for any outline image, several polygons can have the same number of turns but different complexities. Images could also be considered simple if their shape is close to a circle. Other research stated that complexity is proportional to the squared perimeter of the image area (2; 19). This would indeed confirm that circles are the simplest images. The number of turns on its own is not the optimal measure, it is likely that one should also take into account perimeter, the degree of turn and its direction (3). Entropy quantifies disorder and is closely related to the concepts of algorithmic (Kolmogorov) complexity (29). In this research the complexity of robots in images will be determined by looking at the Shannon entropy. The Shannon entropy looks at the orientation of the edges of an image. The edges are compared to each other and the product of their intensity is counted in a histogram. If all the edge orientations are about equally prominent in an image, the entropy is close to maximal. The entropy maximum is $-log_2(\frac{1}{24}) \approx 4.585$. The more prevalent particular orientations are compared to others, the less uniform is the histogram; as a results the entropy is lower (21). Complex images will have a higher amount of edge orientations than simple images. Thus Shannon's entropy score will get higher the more complex an image is.

When you are looking at how colors affect human decision making you can for example look at hue, saturation and lightness or brightness (25). According to Schloss and Palmer (25) when humans look at context-free colors their preferences are primarily dominated by hue, while when they look at the color of objects, the preferences are more strongly affected by lightness/brightness and saturation levels. Lakens et al. (13) stated that picture brightness influences evaluations. They found that brighter versions of of neutral pictures were evaluated more positively than darker versions of the same pictures.

In this research, contrast is defined in two ways. The first definition describes the global contrast of the image, which is commonly understood as the ratio between the brightest and the darkest spots in the image (16). The second definition is the local contrast of an image. Local contrast looks at neighboring pixels and the combination of their brightness levels (1). Sadr and Krowicki (23) manipulated images of human faces by reducing the level of contrast of the total image. They found that decreased contrast produced an increase in perceived attractiveness. They observed that manipulated loss of visual information results a significant increase in the attractiveness of the subject. Tinio et al. (28) looked at the influence on aesthetic judgements from contrast, sharpness and grain. They found that aesthetic judgments of images that are degraded in contrast are evaluated least positively. Their research suggested that of the three elements of image quality, contrast is the most influential to aesthetic judgments of photographs. Gershoni and Kobayashi (9) did a study where participants rated manipulated versions of images and the original. They manipulated the image by increasing the contrast in increments. They found that the overall preference was highest for the original versions of the photographs, and preference decreased linearly with each increment in contrast.

3 Method

3.1 Data Collection

For this study two types of data were used. The first one is the data from the research from de Kleijn et al. (6) The sample consisted of participants from both India and the Netherlands (N=443; n(India)=402, n(Netherlands)=43), all students from various universities in India and Leiden University in the Netherlands. Each participant was shown images of 18 different robots collected from the robots guide website (22). An overview of these robots is shown in figure 1 For each robot the respondents were asked 12 questions about characteristics of the robots followed by a dictator game. The second type of data collected, was extracted from the images themselves. To get more data points per robot, four images per robot were used instead of the one in the study of de Kleijn et al. (6). These were mainly collected from the robots guide website (22). When the robots guide website did not provide enough images, we did a google image search for the robot in question to find additional ones. To ensure that the background wouldn't influence the results, these were removed in Adobe Photoshop, producing an alpha channel background for each image. This allows us to focus on the robots, while ignoring the background. All the images were resized in Python to have 480000 pixels while maintaining the aspect ratio. Feature extraction from the images was mainly done with existing packages and modules in Python. We looked at the following features: brightness, global contrast, local contrast and Shannon entropy.

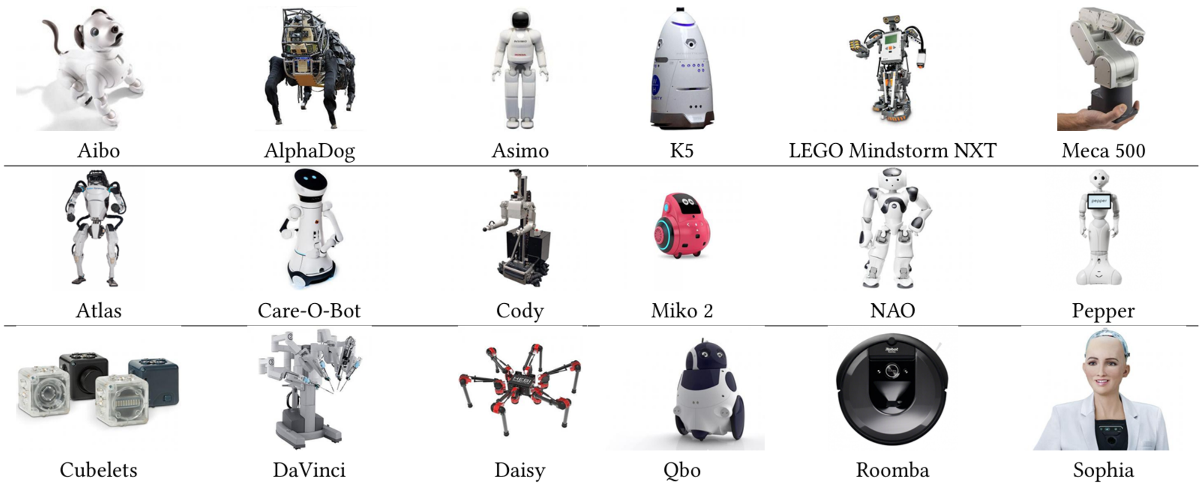


Figure 1: Robots used in the study

3.1.1 Brightness

To determine the brightness of an image the average value for each of the red, green, and blue color channels in the image was determined and then transformed to the perceived brightness (26). To calculate the perceived brightness, the HSP color module is used (8). The model uses three constants (.299, .587, and .114) to represent the different degrees to which each of the primary (RGB) colors affects human perception of the overall brightness of a color (8). We looked at the RGB values, disregarding the alpha channel of the background. This allowed us to focus solely on the robot in the image, excluding the background.

3.1.2 Global contrast

To determine the global contrast the code by Andreas Merentitis (17) was used. The function by Merentitis calculates the contrast of an image based on the range of its pixel values. It converts the image from RGB to grayscale and then looks at the overall range of the pixel values. During the conversion, the alpha channels were ignored, and thus excluded from the analysis.

3.1.3 Local contrast

To determine the local contrast the Gray Level Co-occurrence Matrix (GLCM) is used. The GLCM characterizes texture via spatial pixel intensity relationships, it considers the relationship between neighboring pixels. It analyses how often gray levels occur together within an image (1). Various statistical measures can be derived from the GLCM. In this research only the contrast was looked at. The contrast measures the local variations in the image. High contrast values indicate large differences between neighboring pixel intensities (1). Before using the GLCM to get the local contrast score, the images were converted to grayscale using the same method as for Global contrast, thus ignoring all alpha channels.

3.1.4 Shannon entropy

The Shannon entropy is used to gain more insights on the complexity of the robots. A high entropy score tells us that there are edges in a high number of different angles. A lower entropy score tells us that there is a low number of different angles (21). To determine the Shannon entropy we used the code from the research of Redies et al. (21). We used the bug fixed version from 2021. The code returns the average Shannon entropy in different pixel ranges, the first order shannon entropy and the edge density. The code converts the images to grayscale and uses edge detection to find the edges.

3.2 Analysis

3.2.1 Pearson correlation coefficient

After all the results were collected the data was analysed. This research aims to see if outcomes of dictator games can be predicted by looking at digitally extracted features from robot images. The outcomes of the dictator games themselves are known through the research by de Kleijn et al. (6). Firstly we would like to determine if there is a relationship between features and dictator offers. Secondly we analyse if features correlate to the 12 questions and their corresponding index components. To see if there is a correlation between the data from the research by de Kleijn et al and the feature data collected in this research, the Pearson correlation coefficient was used. The Pearson correlation coefficient measures the linear relationship between two datasets. It is a score that varies between -1 and 1 with 0 implying no correlation. The scipy package in Python provides a module that calculates the Pearson correlation coefficient and its corresponding p-value which shows statistical significance.

To get the best understanding of our data, we calculated the Pearson correlation coefficient in two separate ways. As we have four separate pictures of each individual robot, we calculate the Pearson correlation coefficient in two separate ways. Firstly, we use the average feature values of the four images of each robot. These average values are then used to calculate the correlations versus dictator offer and question answers. The advantage of this method is that we use all the information using a straight forward calculation. Going forward this method will be defined as the "average method".

Secondly, we use the fact that we have four images each, to create four separate datasets, each with only one image per robot. The result is four datasets with 18 images of the 18 robots. We then calculate all correlations per dataset and lastly average the outcomes. This method has the advantage of using the multiple images to create multiple datasets, leading to more robust outcomes. Going forward this method will be defined as the "group method".

3.2.2 Interaction effect

Because we have data from both India and the Netherlands we want to look at the interaction effect to see if the data differs. We will do this for the dictator offer and its significant correlations. The statsmodel.api package in Python provides a module that calculates this interaction effect (5). Using the p-value we can determine if there is a difference between the countries or not. An insignificant p-value ($p>0.05$) will tell us that the data doesn't show a significant difference between in this case the data from India and the Netherlands. A significant p-value ($p<0.05$) will tell us that the data does show a significant difference.

3.2.3 Stepwise regression

To find our strongest predictors we will use a stepwise regression. Stepwise regression is a technique, which utilises an automatic procedure to determine a choice for the predictor variables. By using forward selection, the model starts with no potential variables and adds one at a time to determine the optimal model (18). R is used to automatically run the stepwise regression (24).

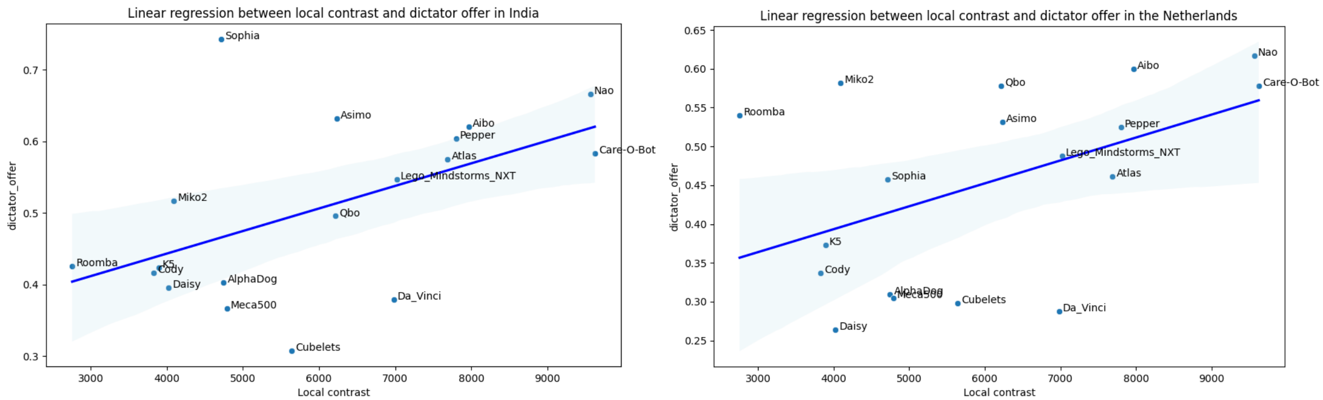
3.2.4 Internal consistency

Internal consistency is used to indicate whether items on a test, that are intended to measure the same construct, produce consistent scores (27).
To measure the internal consistency Cronbach's alpha was used. Cronbach's alpha is a way of assessing reliability by comparing the amount of shared variance, or covariance, among the items. Cronbach's alpha is the average of all possible split-half reliabilities. Often it is helpful to examine what the Cronbach's alpha becomes after a particular item is deleted. If Cronbach's alpha goes up considerably upon deletion of an item, the item may not belong in the measure (4).
To determine the Cornbach's alpha the psych package in R was used (10).

4 Results

Feature	Pearson	P-value	Country	Method
Brightness	0.407	0.094	India	Average
Global contrast	0.431	0.074	India	Average
Local contrast	0.525	0.025	India	Average
avg-shannon20-80	-0.125	0.623	India	Average
avg-shannon80-160	0.002	0.994	India	Average
avg-shannon160-240	0.000	0.999	India	Average
Brightness	0.426	0.078	Netherlands	Average
Global contrast	0.440	0.067	Netherlands	Average
Local contrast	0.478	0.045	Netherlands	Average
avg-shannon20-80	-0.404	0.096	Netherlands	Average
avg-shannon80-160	-0.244	0.328	Netherlands	Average
avg-shannon160-240	-0.312	0.208	Netherlands	Average

Table 2: Pearson correlation coefficient for the dictator offer using the average method



(a) India (b) the Netherlands

Figure 2: Relationship between local contrast and dictator offers. Line of best fit with 95% confidence interval

4.1 Pearson correlation coefficient

4.1.1 Average method

Of the four features that we digitally extracted from the robot images, only local contrast showed a statistically significant correlation ($p<0.05$) with the dictator offer. Local contrast shows a significant positive correlation with data from both India ($p=0.025$) and the Netherlands ($p=0.045$) (table **??**). The other features don't show any direct statistically significant correlation with the dictator offer in both countries.
From this result we can state that a higher local contrast in a digital image results in a higher dictator game offer.
Figure 2 shows the linear regressions between the local contrast and the dictator offer for both India (figure 2a) and the Netherlands (figure 2b). Both figures show a positive relationship between local contrast and the dictator offer.

Besides the correlation of features and the dictator offers, this study also compared the features with how each question (table 1) is answered. While only one feature correlated with the the dictator offer, all four features did have significant correlations with part of the questions.

As we have 4 features and 12 questions, as well as two countries, we calculate 48 correlations per country using the average method. All correlations can be found in appendix C. The correlations show some interesting results that are described per index component below.
Based on correlations between the 4 features and 12 questions, we can also look at the relationships between features and the three indexes from de Kleijn et al. (6).

When looking at the questions related to the Anthropomorphism index (AI), we find that both in India and the Netherlands (table 3 and table 4) contrast has significant positive correlations. Interestingly this is true for both global and local contrast. Our data suggests a relationship between higher contrast and a more human perception of the digital images.

Feature	Question	Pearson	P-value
Brightness	creepy	-0.529	0.024
Brightness	touch	0.498	0.036
Brightness	care_family	0.481	0.043
Global contrast	care_family	0.536	0.022
Global contrast	friendly	0.521	0.027
Global contrast	touch	0.517	0.028
Global contrast	creepy	-0.485	0.042
Local contrast	care_family	0.540	0.021
Local contrast	like	0.534	0.023
Local contrast	cook	0.491	0.038
Local contrast	think_hum	0.480	0.044
Local contrast	phys_sim	0.469	0.050
avg-shannon20-80	creepy	0.587	0.010
avg-shannon160-240	creepy	0.473	0.047

Table 3: Pearson correlation coefficient scores from India calculated using the average method.

Within the likeability Index (LI), two questions stand out. The first question is 'How creepy is this robot?', this question is positively correlated to image complexity. The higher the Shannon entropy score, the creepier the robot is perceived in both countries. In India brightness and global contrast have a negative correlation with creepiness.
The second observation is that in India answers to the question 'Would you like to touch this robot?' are positively correlated to brightness and global contrast. While in the Netherlands the answers to the same question are negatively correlated to image complexity and positively correlated to global contrast.
Lastly India also shows a positive correlation between local contrast and how the question 'How much do you like this robot?' is answered.

Within the Utility Index (UI), in both India and the Netherlands we find a significant correlation between the features brightness, global/ local contrast and the answers to the questions 'Would you let this robot take care of your family?' and 'Would you let this robot cook for you?'. The only other significant correlation is in the Netherlands where more complexity points to a lower willingness to let the robot vacuum your house.

Overall we can see that in India local contrast has a significant correlation with questions from every index component. Brightness has correlations with questions from LI and UI. Global contrast has correlations with AI and LI, while image complexity only has a significant correlation with one LI question.
In the Netherlands global contrast has a significant correlation with questions from every index component. Local contrast has correlations with questions from AI and UI. Image complexity is significantly correlated with both LI and UI questions. Brightness is only significantly correlated to UI questions.

Feature	Question	Pearson	P-value
Brightness	care_family	0.602	0.008
Brightness	cook	0.573	0.013
Global contrast	care_family	0.761	0.000
Global contrast	cook	0.662	0.003
Global contrast	friendly	0.619	0.006
Global contrast	feel_emo	0.560	0.016
Global contrast	touch	0.505	0.032
Local contrast	care_family	0.635	0.005
Local contrast	cook	0.554	0.017
Local contrast	feel_emo	0.508	0.031
Local contrast	friendly	0.501	0.034
avg-shannon20-80	vacuum	-0.637	0.005
avg-shannon20-80	creepy	0.501	0.034
avg-shannon20-80	touch	-0.481	0.043
avg-shannon80-160	vacuum	-0.535	0.022
avg-shannon160-240	vacuum	-0.642	0.004
avg-shannon160-240	creepy	0.483	0.042

Table 4: Pearson correlation coefficient scores from the Netherlands calculated using the average method

4.1.1.1 Conclusion

Among the features, only local contrast showed a statistically significant positive correlation with the dictator offer in both countries ($p<0.05$), indicating that higher local contrast in images leads to higher offers in the dictator game.

4.1.2 Group method

Looking at table 5 we can see that no features show any significant correlations with the dictator offer. Local contrast with the dictator offer answers from India are close but fall short to be statistically significant with a p-value of 0.051.

We can see that this group method, has fewer significant correlations than the average method. However the correlations that do show up, are also present within the average method correlation results.

The AI only shows one significant correlation. It is only significantly correlated with the answers to the question 'How friendly is this robot?' in the Netherlands.

The LI only shows 2 significant correlations, both in the India dataset and with the question 'How creepy is this robot?'. It has a negative correlation with brightness and a positive correlation with image complexity.

Feature	Pearson	P-value	Country	Method
Brightness	0.379	0.126	India	Group
Global contrast	0.347	0.180	India	Group
Local contrast	0.473	0.051	India	Group
avg-shannon20-80	-0.111	0.669	India	Group
avg-shannon80-160	0.000	0.949	India	Group
avg-shannon160-240	0.003	0.880	India	Group
Brightness	0.397	0.106	Netherlands	Group
Global contrast	0.351	0.204	Netherlands	Group
Local contrast	0.428	0.091	Netherlands	Group
avg-shannon20-80	-0.341	0.174	Netherlands	Group
avg-shannon80-160	-0.209	0.409	Netherlands	Group
avg-shannon160-240	-0.272	0.286	Netherlands	Group

Table 5: Pearson correlation coefficient for the dictator offer using the group method

Feature	**Question**	**Pearson**	**P-value**
Brightness	creepy	-0.499	0.039
Local contrast	care_family	0.487	0.044
avg-shannon20-80	creepy	0.513	0.036

Table 6: Pearson correlation coefficient scores from India calculated using the group method

UI is the only index that has significant correlations within both countries. The answers to the questions 'Would you let this robot take care of your family?' and 'Would you let this robot cook for you?' are correlated to brightness, global contrast and local contrast in the Netherlands. The first question is also significantly correlated with local contrast in India.
The Netherlands shows more UI correlations, it shows a negative correlation between the answers to the question 'Would you let this robot vacuum your house?' and image complexity.

4.1.2.1 Conclusion

This method adds additional insights by seemingly being more selective, with fewer correlations. We see that UI has correlations with both countries, while AI and LI only have one question correlate in one country.

4.2 Interaction effect

We also wanted to see if we could find a significant difference between the data from the Netherlands and India when it came to the relationship between local contrast and dictator offer. An interaction effect between local contrast and country when predicting dictator offers was looked at. We used the data from the average method, as this was the only method that had any significant correlations with dictator offer.
Local contrast and country didn't show a significant interaction effect when predicting dictator offers, b=-1.979-e06, t(32)=-0.106, p=0.916.

Feature	Question	Pearson	P-value
Brightness	care_family	0.564	0.015
Brightness	cook	0.537	0.023
Global contrast	care_family	0.596	0.010
Global contrast	cook	0.520	0.039
Global contrast	friendly	0.495	0.046
Local contrast	care_family	0.568	0.019
Local contrast	cook	0.494	0.047
avg-shannon20-80	vacuum	-0.556	0.017
avg-shannon160-240	vacuum	-0.567	0.020

Table 7: Pearson correlation coefficient scores from the Netherlands calculated using the group method

The regression model is statistically significant overall, as indicated by the Prob (F-statistic) of 0.0116.

4.2.1 Conclusion

Local contrast and country didn't show a significant interaction effect when predicting dictator offers. The regression model is statically significant. However, the country variable and its interaction with local contrast are not significant, suggesting that the effect of local contrast is not different for the Netherlands compared to India.

4.3 Stepwise regression

Using the forward stepwise regression to find best predictors for the dictator offer, we found that the final model included the predictors local contrast, avg-shannon20-80, avg-shannon80-160, brightness, and avg-shannon160-240. Global contrast was left out.
the Models R-squared is 0.35, indicating that about 35% of the variance in dictator offer is explained by the predictors included in the model. The F-statistic, which shows the overall significance of the model, is 14.82 with a p-value of 1.25e-11. Indicating that the overall model is statistically significant, and at least one of the predictors has a significant effect on dictator offer.

4.3.1 Conclusion

The regression results suggest that local contrast, avg-shannon20-80, avg-shannon80-160, and to a lesser extent, Brightness, are significant predictors of dictator offer. These variables collectively explain a moderate amount of variance in dictator offer, as indicated by the R-squared.

4.4 Internal consistency

The analysis produced a standardized Cronbach's alpha of 0.71. A value higher then 0.7 means an acceptable internal consistency. However, the raw alpha of -0.013 suggests problematic raw internal consistency. The initial analysis showed a standardised Cornbach's alpha of 0.66. by adding the "check.keys=TRUE" option in the alpha formula the outcome improved to 0.71. The items brightness and global contrast were automatically reversed due to initial negative correlations with the first principal component. Dropping local contrast increased the raw alpha to a positive value, while removing global contrast resulted in the highest standardized alpha.

4.4.1 Conclusion

Overall the outcome is not very clear and no clear conclusions can be drawn. The output of the model can be found in the appendix B.

5 Conclusions

To answer the research question "Can we predict from images alone, how much money people allocate to robots?", the positive correlations indicates that local contrast has some predicting value for dictator game behavior. There is a positive significant correlation between the dictator offer and local contrast in both countries using the average method.
The other three researched features did not have any direct significant correlations with the dictator game offer. They did however have significant correlations with one or more of the questions from the research by de Kleijn et al. (6). Using their research you would be able to further predict the dictator game outcome, because they found that you could predict the dictator game outcome by looking at the answers to the questions. This means that when looking at the combined results, you could state that there is an indirect relationship between all features and dictator game behavior. For example in the research by de Kleijn et al. (6) they stated that likeability is the most important determinant of dictator game behavior. We saw that image complexity and global contrast in both countries and all features in India have significant correlations with the likeability questions.

When comparing the results of the average method and group method, we see overlap in statistical significance. All of the significant correlations found using the group method can also be found in the average method. The average method however, showed notably more significant correlations. Especially with the AI and LI related questions. The group method had hardly any correlations with those indexes. When looking at the outcome of both methods, there is no clear conclusion on why and how they differ.

As we have datasets for two countries, we also looked at the difference between these datasets. The interaction effect however showed no significant difference in how local contrast and country affects the dictator offer.

In addition we also used a forward stepwise regression. This showed us that the predictors local contrast, avg-shannon20-80, avg-shannon80-160, Brightness, and avg-shannon160-24 best predicted the dictator offer. This is mostly in line with findings from the Pearson correlation data. One notable thing is that global contrast was dropped from the model. An explanation for this could be that local contrast was enough for the model to reach its optimum, as these are both contrast measures.

Previous research supports these findings as it already stated that the features used in this study have an influence on how humans judge images. For example that Lakens et al. (13) stated that picture brightness influences evaluations. Tinio et al. (28) found that from the features they looked at, contrast was the most influential on the aesthetic judgements of photographs.

We do have to keep one thing in mind. The dataset of robots is relatively small. As this research built on the research by de Kleijn et al. (6) we were limited to the same dataset. This could have influenced the outcome. To be able to make more solid conclusions, the dataset should be expanded, which would mean conducting another round of interviews with new robots.

To conclude, we can say that based on the outcomes of this study and within the scope, extracted features from digital images could be useful to make predictions around dictator game behavior. It would be interesting to further research this topic, for instance looking at more features, a bigger dataset, a different subject matter or to further evaluate the indirect feature correlations.

References

[1] G. Ajmera. Feature extraction of images using glcm (gray level cooccurrence matrix), 2024. URL: https://medium.com/@girishajmera/feature-extraction-of-images-using-glcm-gray-level-cooccurrence-matrix-e4bda8729498.

[2] F. Attneave. Physical determinants of the judged complexity of shapes. *Journal of Experimental Psychology*, 53(4):221–227, 1957.

[3] V. Chikhman, V. Bondarko, M. Danilova, A. Goluzina, and Y. Shelepin. Complexity of images: Experimental and computational estimates compared. *Perception*, 41:631–47, 10 2012. doi:10.1068/p6987.

[4] L. Collins. Research design and methods. In J. E. Birren, editor, *Encyclopedia of Gerontology (Second Edition)*, pages 433–442. Elsevier, New York, second edition edition, 2007. URL: https://www.sciencedirect.com/science/article/pii/B0123708702001621, doi:10.1016/B0-12-370870-2/00162-1.

[5] H. Daryani. Interaction effects in machine learning, 2021. URL: https://medium.com/analytics-vidhya/interaction-effects-in-machine-learning-c8374a9e62dc.

[6] R. de Kleijn, A. Dev, and S. Mukherjee. Behavioral-economic games with commercially available robots. *Companion of the 2024 ACM/IEEE International Conference on Human-Robot Interaction*, 2024.

[7] A. Edele, I. Dziobek, and M. Keller. Explaining altruistic sharing in the dictator game: The role of affective empathy, cognitive empathy, and justice sensitivity. *Learning and Individual Differences*, 24:96–102, 2013.

[8] D. R. Finley. Hsp color model — alternative to hsv (hsb) and hsl, 2006. URL: https://alienryderflex.com/hsp.html.

[9] S. Gershoni and H. Kobayashi. How we look at photographs as indicated by contrast discrimination performance versus contrast preference. *Journal of Imaging Science and Technology*, 50(4):320–326, 2006.

[10] S. Jackson. Five ways to calculate internal consistency, 2016. URL: https://www.r-bloggers.com/2016/08/five-ways-to-calculate-internal-consistency/.

[11] S. Jessup, A. Gibson, A. Capiola, G. Alarcon, and M. Borders. Investigating the effect of trust manipulations on affect over time in human-human versus human-robot interactions. *Proceedings of the 53rd Hawaii International Conference on System Sciences*, pages 553–562, 2020.

[12] C. M. Jonker, K. V. Hindriks, P. Wiggersa, and J. Broekens. Negotiating agents. *AI magazine*, pages 79–91, 2012.

[13] D. Lakens, D. A. Fockenberg, K. P. H. Lemmens, J. Ham, and C. J. H. Midden. Brightness differences influence the evaluation of affective pictures. *Cognition and Emotion*, 27(7):1225–1246, 2013. doi:10.1080/02699931.2013.781501.

[14] M. Lee, G. Lucas, and J. Gratch. Comparing mind perception in strategic exchanges: human-agent negotiation, dictator and ultimatum games. *Journal on Multimodal User Interfaces*, 15(2):201–214, 2021. doi:10.1007/s12193-020-00356-6.

[15] M. Maggioni and D. Rossignoli. If it looks like a human and speaks like a human ... communication and cooperation in strategic human–robot interactions. *Journal of Behavioral and Experimental Economics*, 104, 2023.

[16] K. Matkovic, L. Neumann, A. Neumann, T. Psik, and W. Purgathofer. Global contrast factor - a new approach to image contrast. *Computational Aesthetics in Graphics, Visualization and Imaging*, pages 159–167, 2005. doi:10.2312/COMPAESTH/COMPAESTH05/159-167.

[17] A. Merentitis. detect-contrast-blurriness, 2020. URL: https://github.com/AndreasMerentitis/detect-contrast-blurriness.

[18] A. Miller, J. Panneerselvam, and L. Liu. A review of regression and classification techniques for analysis of common and rare variants and gene-environmental factors. *Neurocomputing*, 489:466–485, 2022. URL: https://www.sciencedirect.com/science/article/pii/S092523122101907X, doi:10.1016/j.neucom.2021.08.150.

[19] D. G. Pelli, C. W. Burns, B. Farell, and D. C. Moore-Page. Feature detection and letter identification. *Vision Research*, 46(28):4646–4674, 2006. URL: https://www.sciencedirect.com/science/article/pii/S004269890600232X, doi:10.1016/j.visres.2006.04.023.

[20] Y. Razin and K. Feigh. Committing to interdependence: Implications from game theory for human–robot trust. *Journal of Behavioral Robotics*, 12:481—-502, 2021.

[21] C. Redies, A. Brachmann, and J. Wagemans. High entropy of edge orientations characterizes visual artworks from diverse cultural backgrounds. *Vision Research*, 133:130–144, 2017. URL: https://www.sciencedirect.com/science/article/pii/S0042698917300354, doi:10.1016/j.visres.2017.02.004.

[22] Robotsguide. Robots, 2024. URL: https://robotsguide.com/robots.

[23] J. Sadr and L. Krowicki. Face perception loves a challenge: Less information sparks more attraction. *Vision Research*, 157:61–83, 2019. Face perception: Experience, models and neural mechanisms. URL: https://www.sciencedirect.com/science/article/pii/S0042698919300215, doi:10.1016/j.visres.2019.01.009.

[24] S. P. Sanderson. A complete guide to stepwise regression in r, 2023. URL: https://www.r-bloggers.com/2023/12/a-complete-guide-to-stepwise-regression-in-r/.

[25] K. B. Schloss and S. E. Palmer. *Color Preference*, pages 1–7. Springer Berlin Heidelberg, Berlin, Heidelberg, 2020. doi:10.1007/978-3-642-27851-8_70-14.

[26] stackoverflow. What are some methods to analyze image brightness using python?, 2017. URL: https://stackoverflow.com/questions/3490727/what-are-some-methods-to-analyze-image-brightness-using-python.

[27] W. Tang, Y. Cui, and O. Babenko. Internal consistency: Do we really know what it is and how to assess it? *Journal of Psychology and Behavioral Science*, 2:205–220, 01 2014.

[28] P. Tinio, H. Leder, and M. Strasser. Image quality and the aesthetic judgment of photographs: Contrast, sharpness, and grain teased apart and put together. *Psychology of Aesthetics, Creativity, and the Arts*, 5(2):165–176, 2011. doi:10.1037/a0019542.

[29] D. H. Zanette. Quantifying the complexity of black-and-white images. *PLoS ONE*, 13(11), 2018. doi:10.1371/journal.pone.0207879.

A Stepwise regression

```
Call:
lm(formula = dictator_offer ~ Local_contrast + avg_shannon20_80 +
    avg_shannon80_160 + Brightness + avg_shannon160_240, data = full_data)

Residuals:
     Min       1Q   Median       3Q      Max
-0.21781 -0.06768  0.00000  0.07287  0.32727

Coefficients:
                     Estimate Std. Error t value Pr(>|t|)
(Intercept)         1.061e-01  3.862e-01   0.275 0.783883
Local_contrast      1.869e-05  5.861e-06   3.189 0.001766 **
avg_shannon20_80   -4.642e-01  1.314e-01  -3.532 0.000560 ***
avg_shannon80_160   7.022e-01  1.912e-01   3.672 0.000343 ***
Brightness          1.076e-03  5.077e-04   2.119 0.035839 *
avg_shannon160_240 -2.082e-01  1.098e-01  -1.896 0.060067 .
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.1013 on 138 degrees of freedom
Multiple R-squared:  0.3494,	Adjusted R-squared:  0.3258
F-statistic: 14.82 on 5 and 138 DF,  p-value: 1.25e-11
```

Figure 3: Summary of the stepwise regression.

B Cornbach's alpha

C All Pearson Correlation Coefficient results

```
Reliability analysis
Call: alpha(x = feature_data, check.keys = TRUE)

  raw_alpha std.alpha G6(smc) average_r S/N    ase mean  sd median_r
    -0.013      0.71    0.91      0.29 2.5 0.0034 4837 369     0.33

    95% confidence boundaries
         lower alpha upper
Feldt    -0.42 -0.01  0.31
Duhachek -0.02 -0.01 -0.01

 Reliability if an item is dropped:
                   raw_alpha std.alpha G6(smc) average_r S/N alpha se var.r med.r
Brightness-          0.00013      0.66    0.88      0.28 2.0  6.9e-05  0.26  0.22
Global_contrast-    -0.01344      0.80    0.92      0.45 4.1  3.5e-03  0.17  0.44
Local_contrast       0.02442      0.81    0.92      0.47 4.4  4.6e-03  0.14  0.50
avg_shannon20_80    -0.01351      0.52    0.83      0.18 1.1  3.5e-03  0.23  0.29
avg_shannon80_160   -0.01352      0.51    0.82      0.17 1.0  3.5e-03  0.21  0.26
avg_shannon160_240  -0.01354      0.55    0.85      0.20 1.2  3.5e-03  0.22  0.22

 Item statistics
                    n raw.r std.r r.cor r.drop    mean      sd
Brightness-        72 -0.41  0.66 0.614  -0.42 11478.3 2.8e+01
Global_contrast-   72 -0.64  0.23 0.136  -0.64 11553.8 3.8e-02
Local_contrast     72  1.00  0.18 0.088  -0.42  5976.0 2.2e+03
avg_shannon20_80   72  0.20  0.93 0.948   0.20     4.2 2.0e-01
avg_shannon80_160  72  0.27  0.95 0.978   0.27     4.3 1.6e-01
avg_shannon160_240 72  0.33  0.88 0.890   0.33     4.2 1.9e-01
```

Figure 4: Output of the Cornbach's alpha.

Feature	**Question**	**Pearson**	**P-value**
Brightness	creepy	-0.529	0.024
Brightness	touch	0.498	0.036
Brightness	care_family	0.481	0.043
Brightness	friendly	0.466	0.051
Brightness	feel_emo	0.415	0.087
Brightness	want_to_have	0.442	0.066
Brightness	cook	0.417	0.086
Brightness	dictator_offer	0.407	0.094
Brightness	like	0.385	0.115
Brightness	vacuum	0.381	0.119
Brightness	think_hum	0.364	0.138
Brightness	plan_indep	0.314	0.204
Brightness	phys_sim	0.294	0.237
Global contrast	care_family	0.536	0.022
Global contrast	friendly	0.521	0.027
Global contrast	touch	0.517	0.028
Global contrast	creepy	-0.485	0.042
Global contrast	think_hum	0.417	0.085
Global contrast	want_to_have	0.455	0.058
Global contrast	feel_emo	0.464	0.052
Global contrast	cook	0.461	0.054
Global contrast	like	0.458	0.056
Global contrast	dictator_offer	0.431	0.074
Global contrast	phys_sim	0.381	0.118
Global contrast	plan_indep	0.332	0.178
Global contrast	vacuum	0.264	0.290
Local contrast	care_family	0.540	0.021
Local contrast	like	0.534	0.023
Local contrast	dictator_offer	0.525	0.025
Local contrast	cook	0.491	0.038
Local contrast	think_hum	0.480	0.044
Local contrast	phys_sim	0.469	0.050
Local contrast	feel_emo	0.467	0.051
Local contrast	friendly	0.465	0.052
Local contrast	want_to_have	0.461	0.054
Local contrast	plan_indep	0.456	0.057
Local contrast	touch	0.430	0.075
Local contrast	creepy	-0.336	0.173
Local contrast	vacuum	0.259	0.299

Table 8: All Pearson correlation coefficient scores from India using the average method Part 1

Feature	**Question**	**Pearson**	**P-value**
avg-shannon20-80	creepy	0.587	0.010
avg-shannon20-80	touch	-0.395	0.105
avg-shannon20-80	vacuum	-0.364	0.137
avg-shannon20-80	want_to_have	-0.324	0.190
avg-shannon20-80	friendly	-0.299	0.228
avg-shannon20-80	like	-0.204	0.416
avg-shannon20-80	care_family	-0.185	0.463
avg-shannon20-80	feel_emo	-0.139	0.584
avg-shannon20-80	dictator_offer	-0.125	0.623
avg-shannon20-80	cook	-0.108	0.671
avg-shannon20-80	think_hum	-0.046	0.856
avg-shannon20-80	phys_sim	0.016	0.951
avg-shannon20-80	plan_indep	0.005	0.985
avg-shannon80-160	creepy	0.460	0.055
avg-shannon80-160	touch	-0.249	0.319
avg-shannon80-160	vacuum	-0.234	0.349
avg-shannon80-160	want_to_have	-0.175	0.489
avg-shannon80-160	friendly	-0.151	0.549
avg-shannon80-160	phys_sim	0.132	0.602
avg-shannon80-160	plan_indep	0.110	0.663
avg-shannon80-160	think_hum	0.074	0.771
avg-shannon80-160	care_family	-0.060	0.813
avg-shannon80-160	like	-0.041	0.870
avg-shannon80-160	feel_emo	-0.018	0.944
avg-shannon80-160	cook	0.018	0.944
avg-shannon80-160	dictator_offer	0.002	0.994
avg-shannon160-240	creepy	0.473	0.047
avg-shannon160-240	vacuum	-0.288	0.246
avg-shannon160-240	touch	-0.246	0.326
avg-shannon160-240	want_to_have	-0.189	0.453
avg-shannon160-240	friendly	-0.148	0.557
avg-shannon160-240	phys_sim	0.128	0.614
avg-shannon160-240	plan_indep	0.113	0.656
avg-shannon160-240	think_hum	0.076	0.766
avg-shannon160-240	like	-0.059	0.815
avg-shannon160-240	care_family	-0.033	0.895
avg-shannon160-240	cook	0.028	0.913
avg-shannon160-240	feel_emo	-0.007	0.979
avg-shannon160-240	dictator_offer	0.000	0.999

Table 9: All Pearson correlation coefficient scores from India using the average method Part 2

Feature	Question	Pearson	P-value
Brightness	creepy	-0.499	0.039
Brightness	touch	0.467	0.051
Brightness	care_family	0.449	0.066
Brightness	friendly	0.435	0.072
Brightness	want_to_have	0.413	0.091
Brightness	cook	0.391	0.110
Brightness	feel_emo	0.387	0.114
Brightness	dictator_offer	0.379	0.126
Brightness	like	0.359	0.144
Brightness	vacuum	0.355	0.154
Brightness	think_hum	0.340	0.168
Brightness	plan_indep	0.292	0.245
Brightness	phys_sim	0.275	0.271
Global contrast	care_family	0.426	0.089
Global contrast	friendly	0.415	0.099
Global contrast	touch	0.411	0.125
Global contrast	feel_emo	0.371	0.146
Global contrast	cook	0.361	0.167
Global contrast	like	0.371	0.177
Global contrast	want_to_have	0.366	0.177
Global contrast	dictator_offer	0.347	0.180
Global contrast	think_hum	0.329	0.200
Global contrast	creepy	-0.372	0.240
Global contrast	phys_sim	0.297	0.246
Global contrast	plan_indep	0.270	0.309
Global contrast	vacuum	0.215	0.441
Local contrast	care_family	0.487	0.044
Local contrast	like	0.478	0.051
Local contrast	dictator_offer	0.473	0.051
Local contrast	cook	0.441	0.071
Local contrast	think_hum	0.433	0.081
Local contrast	feel_emo	0.420	0.090
Local contrast	phys_sim	0.424	0.091
Local contrast	friendly	0.417	0.091
Local contrast	want_to_have	0.412	0.097
Local contrast	plan_indep	0.412	0.098
Local contrast	touch	0.383	0.128
Local contrast	creepy	-0.298	0.247
Local contrast	vacuum	0.230	0.372

Table 10: All Pearson correlation coefficient scores from India from the group method Part 1

Feature	Question	Pearson	P-value
avg-shannon20-80	creepy	0.513	0.036
avg-shannon20-80	touch	-0.339	0.187
avg-shannon20-80	vacuum	-0.324	0.194
avg-shannon20-80	want_to_have	-0.281	0.273
avg-shannon20-80	friendly	-0.245	0.353
avg-shannon20-80	like	-0.171	0.514
avg-shannon20-80	care_family	-0.160	0.532
avg-shannon20-80	feel_emo	-0.110	0.626
avg-shannon20-80	dictator_offer	-0.111	0.669
avg-shannon20-80	cook	-0.103	0.687
avg-shannon20-80	plan_indep	0.002	0.823
avg-shannon20-80	think_hum	-0.044	0.833
avg-shannon20-80	phys_sim	0.000	0.855
avg-shannon80-160	creepy	0.404	0.113
avg-shannon80-160	touch	-0.214	0.401
avg-shannon80-160	vacuum	-0.207	0.417
avg-shannon80-160	want_to_have	-0.151	0.553
avg-shannon80-160	friendly	-0.123	0.637
avg-shannon80-160	phys_sim	0.106	0.686
avg-shannon80-160	plan_indep	0.094	0.712
avg-shannon80-160	think_hum	0.061	0.803
avg-shannon80-160	cook	0.012	0.833
avg-shannon80-160	care_family	-0.051	0.836
avg-shannon80-160	feel_emo	-0.011	0.861
avg-shannon80-160	like	-0.033	0.868
avg-shannon80-160	dictator_offer	0.000	0.949
avg-shannon160-240	creepy	0.418	0.092
avg-shannon160-240	vacuum	-0.253	0.319
avg-shannon160-240	touch	-0.213	0.414
avg-shannon160-240	want_to_have	-0.164	0.525
avg-shannon160-240	friendly	-0.125	0.646
avg-shannon160-240	phys_sim	0.113	0.660
avg-shannon160-240	plan_indep	0.102	0.691
avg-shannon160-240	think_hum	0.069	0.761
avg-shannon160-240	cook	0.027	0.799
avg-shannon160-240	feel_emo	-0.001	0.801
avg-shannon160-240	like	-0.049	0.821
avg-shannon160-240	care_family	-0.025	0.843
avg-shannon160-240	dictator_offer	0.003	0.880

Table 11: All Pearson correlation coefficient scores from India from the group method Part 2

Feature	Question	Pearson	P-value
Brightness	care_family	0.602	0.008
Brightness	cook	0.573	0.013
Brightness	friendly	0.456	0.057
Brightness	dictator_offer	0.426	0.078
Brightness	touch	0.421	0.082
Brightness	feel_emo	0.365	0.136
Brightness	vacuum	0.362	0.139
Brightness	think_hum	0.356	0.147
Brightness	creepy	-0.314	0.204
Brightness	like	0.311	0.208
Brightness	phys_sim	0.283	0.255
Brightness	want_to_have	0.224	0.372
Brightness	plan_indep	0.090	0.723
Global contrast	care_family	0.761	0.000
Global contrast	cook	0.662	0.003
Global contrast	friendly	0.619	0.006
Global contrast	feel_emo	0.560	0.016
Global contrast	touch	0.505	0.032
Global contrast	dictator_offer	0.441	0.067
Global contrast	like	0.438	0.069
Global contrast	think_hum	0.390	0.109
Global contrast	phys_sim	0.373	0.127
Global contrast	creepy	-0.322	0.193
Global contrast	want_to_have	0.218	0.385
Global contrast	vacuum	0.110	0.664
Global contrast	plan_indep	0.079	0.757
Local contrast	care_family	0.635	0.005
Local contrast	cook	0.554	0.017
Local contrast	feel_emo	0.508	0.031
Local contrast	friendly	0.501	0.034
Local contrast	dictator_offer	0.478	0.045
Local contrast	think_hum	0.439	0.068
Local contrast	phys_sim	0.437	0.070
Local contrast	like	0.391	0.108
Local contrast	touch	0.303	0.221
Local contrast	plan_indep	0.290	0.243
Local contrast	want_to_have	0.216	0.390
Local contrast	creepy	-0.146	0.563
Local contrast	vacuum	0.022	0.930

Table 12: All Pearson correlation coefficient scores from the Netherlands using the average method Part 1

Feature	Question	Pearson	P-value
avg-shannon20-80	vacuum	-0.637	0.005
avg-shannon20-80	creepy	0.501	0.034
avg-shannon20-80	touch	-0.481	0.043
avg-shannon20-80	want_to_have	-0.459	0.055
avg-shannon20-80	dictator_offer	-0.404	0.096
avg-shannon20-80	like	-0.390	0.109
avg-shannon20-80	friendly	-0.319	0.197
avg-shannon20-80	care_family	-0.287	0.248
avg-shannon20-80	cook	-0.249	0.319
avg-shannon20-80	plan_indep	0.161	0.523
avg-shannon20-80	feel_emo	-0.032	0.900
avg-shannon20-80	think_hum	-0.013	0.959
avg-shannon20-80	phys_sim	0.009	0.973
avg-shannon80-160	vacuum	-0.535	0.022
avg-shannon80-160	creepy	0.449	0.061
avg-shannon80-160	touch	-0.371	0.129
avg-shannon80-160	want_to_have	-0.321	0.194
avg-shannon80-160	dictator_offer	-0.244	0.328
avg-shannon80-160	like	-0.240	0.338
avg-shannon80-160	plan_indep	0.226	0.367
avg-shannon80-160	friendly	-0.164	0.516
avg-shannon80-160	care_family	-0.163	0.517
avg-shannon80-160	cook	-0.127	0.615
avg-shannon80-160	phys_sim	0.121	0.631
avg-shannon80-160	think_hum	0.081	0.749
avg-shannon80-160	feel_emo	0.077	0.762
avg-shannon160-240	vacuum	-0.642	0.004
avg-shannon160-240	creepy	0.483	0.042
avg-shannon160-240	want_to_have	-0.446	0.064
avg-shannon160-240	touch	-0.411	0.090
avg-shannon160-240	dictator_offer	-0.312	0.208
avg-shannon160-240	like	-0.310	0.210
avg-shannon160-240	plan_indep	0.195	0.439
avg-shannon160-240	friendly	-0.168	0.504
avg-shannon160-240	care_family	-0.123	0.626
avg-shannon160-240	phys_sim	0.123	0.627
avg-shannon160-240	feel_emo	0.103	0.685
avg-shannon160-240	think_hum	0.095	0.709
avg-shannon160-240	cook	-0.083	0.743

Table 13: All Pearson correlation coefficient scores from the Netherlands using the average method Part 2

Feature	Question	Pearson	P-value
Brightness	care_family	0.564	0.015
Brightness	cook	0.537	0.023
Brightness	friendly	0.425	0.082
Brightness	dictator_offer	0.398	0.106
Brightness	touch	0.395	0.107
Brightness	feel_emo	0.341	0.169
Brightness	vacuum	0.339	0.172
Brightness	think_hum	0.334	0.176
Brightness	creepy	-0.295	0.236
Brightness	like	0.290	0.249
Brightness	phys_sim	0.265	0.288
Brightness	want_to_have	0.208	0.409
Brightness	plan_indep	0.082	0.750
Global contrast	care_family	0.596	0.010
Global contrast	cook	0.520	0.039
Global contrast	friendly	0.495	0.046
Global contrast	feel_emo	0.448	0.072
Global contrast	touch	0.400	0.166
Global contrast	dictator_offer	0.351	0.204
Global contrast	like	0.352	0.235
Global contrast	think_hum	0.308	0.235
Global contrast	phys_sim	0.289	0.264
Global contrast	creepy	-0.248	0.354
Global contrast	vacuum	0.091	0.515
Global contrast	want_to_have	0.178	0.526
Global contrast	plan_indep	0.075	0.694
Local contrast	care_family	0.568	0.019
Local contrast	cook	0.494	0.047
Local contrast	feel_emo	0.456	0.068
Local contrast	friendly	0.447	0.076
Local contrast	dictator_offer	0.428	0.091
Local contrast	think_hum	0.395	0.108
Local contrast	phys_sim	0.395	0.118
Local contrast	like	0.347	0.185
Local contrast	plan_indep	0.264	0.304
Local contrast	touch	0.267	0.309
Local contrast	want_to_have	0.190	0.466
Local contrast	creepy	-0.125	0.612
Local contrast	vacuum	0.014	0.798

Table 14: All Pearson correlation coefficient scores from the Netherlands from the group method Part 1

Feature	Question	Pearson	P-value
avg-shannon20-80	vacuum	-0.556	0.017
avg-shannon20-80	creepy	0.434	0.080
avg-shannon20-80	want_to_have	-0.385	0.123
avg-shannon20-80	touch	-0.394	0.130
avg-shannon20-80	dictator_offer	-0.341	0.174
avg-shannon20-80	like	-0.313	0.240
avg-shannon20-80	care_family	-0.238	0.353
avg-shannon20-80	friendly	-0.248	0.363
avg-shannon20-80	cook	-0.219	0.407
avg-shannon20-80	plan_indep	0.124	0.615
avg-shannon20-80	feel_emo	-0.010	0.783
avg-shannon20-80	phys_sim	-0.007	0.822
avg-shannon20-80	think_hum	-0.023	0.837
avg-shannon80-160	vacuum	-0.468	0.054
avg-shannon80-160	creepy	0.391	0.110
avg-shannon80-160	touch	-0.313	0.227
avg-shannon80-160	want_to_have	-0.275	0.272
avg-shannon80-160	dictator_offer	-0.209	0.409
avg-shannon80-160	like	-0.195	0.457
avg-shannon80-160	plan_indep	0.184	0.483
avg-shannon80-160	cook	-0.105	0.607
avg-shannon80-160	care_family	-0.134	0.609
avg-shannon80-160	friendly	-0.125	0.652
avg-shannon80-160	phys_sim	0.096	0.709
avg-shannon80-160	feel_emo	0.073	0.777
avg-shannon80-160	think_hum	0.064	0.783
avg-shannon160-240	vacuum	-0.567	0.020
avg-shannon160-240	creepy	0.427	0.082
avg-shannon160-240	want_to_have	-0.393	0.114
avg-shannon160-240	touch	-0.359	0.159
avg-shannon160-240	dictator_offer	-0.272	0.287
avg-shannon160-240	like	-0.268	0.309
avg-shannon160-240	plan_indep	0.168	0.511
avg-shannon160-240	feel_emo	0.096	0.621
avg-shannon160-240	friendly	-0.139	0.632
avg-shannon160-240	phys_sim	0.109	0.673
avg-shannon160-240	cook	-0.065	0.710
avg-shannon160-240	care_family	-0.101	0.723
avg-shannon160-240	think_hum	0.084	0.747

Table 15: All Pearson correlation coefficient scores from the Netherlands from the group method Part 2