HAL
archives-ouvertes.fr

Aversions to trust

Anne Corcos François Pannequin, Sacha Bourgeois-Gironde

▶ **To cite this version:**

Anne Corcos François Pannequin, Sacha Bourgeois-Gironde. Aversions to trust. Recherches Economiques de Louvain - Louvain economic review, De Boeck Université, 2012, 78 (3/4), pp.152-173. <ijn_00734564>

HAL Id: ijn_00734564
https://jeannicod.ccsd.cnrs.fr/ijn_00734564

Submitted on 23 Sep 2012

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Aversions to trust

Recent economic conceptualizing of trust focuses of two distinct aspects of the notion. On the one side, the stress is put on the *individual decision* to trust another individual, on the other side, due to the fact that trust is displayed through a certain type of social interactions, the analysis will primarily insist on the exchange structures, institutions, and *pro-social preferences* through which trust emerges. The two aspects, of course, are not incompatible. It is first and foremost a matter of taste and focus in the analysis that lead to insist alternatively on the analogy between risk and trust or to spell out what in the decision to trust is precisely not reducible to risk taking but takes its root in social considerations.

Economists following the individualistic approach have then often thought in terms of an analogy between trust and risky investments. The now very popular experimental measure of trust initially proposed by Berg, Dickhaut and McCabe (1995), initially called the "investment game", but progressively subsumed under the label of the "trust game". A "trust game", in its original sense, qualifies an extensive game-theoretical structure which implements the decision to trust in terms of investing a variable amount of money by transferring it to a partner in a context where the trustee might defect. This view of trust as a risky investment made its way beyond experimental economics and was instrumental, for instance, in Guiso *et al.* (2008). Central in that approach is the idea of an individual, an investor, endowed with reliable subjective beliefs with respect to a partner's behavior. Where this reliability stems from is of course a crucial question to address. But, in principle, the trust game can be seen as an arbitrage between owing an asset with certainty and investing in a risky asset (trusting the trustee). The information we acquire concerning the behavior of our partner is tantamount to learning the objective return characteristics of a risky investment. We learn or anticipate those returns on the basis of a more or less predictable profile of an asset/partner. In fine, there is nothing substantially distinct in the analysis of trust (as stylized in the trust game, at least) and the analysis of individual risk.

Differently, though, when the emphasis is put on the social dimension of trust, a more complex array of motives and incentives underlying trusting behavior arises. Some authors make central, for instance, the altruistic aspect of trust (Cox, 2004). Our own analysis will also tend to refute the sheer assimilation of trust to a risky investment and will join with accounts of trust that negatively understands it in terms of overcoming betrayal aversion (Bohnet and Zeckhauser, 2004) rather than the expression of an altruistic or fundamentally pro-social tendency. However, and for the very reason that we will insist on our fear of betrayal in social interactions and the understanding of successful social interactions in terms of an exogenous (due to some explicit contract or signal) or endogenous (due to repetition, actual or virtual) decrease of that fear, informational and probabilistic considerations are not absent from our analysis. The main tenets of our analysis, then, are the following:

- The analogy between trust and risk investments is not grounded.
- Overcoming betrayal aversion is an essential aspect of trust.

- It is worth pursuing the investigation of possible correlations between the decision to trust and particular decision-theoretical contexts, like, especially, decision under ambiguity.

We then clearly focus, in this article, on two main types of "aversion" which seem constitutive of the notion of trust. One social aversion: the fear of betrayal, and a certain aversion to uncertainty: ambiguity aversion. We furthermore elaborate on the possible intrinsic connection, at conceptual, empirical and neurobiological levels, of betrayal aversion and ambiguity aversion. We leave aside possible proximal positive motivations to trust such as altruism. We try to isolate, from a methodological point of view, what the relative impact of these two aversions on invested amounts of money in one-shot trust games is. We therefore suggest a way to disambiguate negative social (betrayal aversion) and non-social (ambiguity aversion) factors in trust game data analysis.

Non-social determinants of trust can be defined as the subjective assessment, by the decision-maker, of the incurred risk of perceiving, in the trust-game, smaller returns than the invested amounts (on the discrimination between social and non-social determinants of trust, see e.g. Bourgeois-Gironde and Corcos, 2011). In that case the investor would have better kept his initial endowment, which always makes a sure gain, and the trade-off is here tantamount to a classical measure of risk aversion in terms of what certainty equivalent of a risky option it is worth foregoing to take that risk. It is that formal analogy between the trust game and the standard measure of risk aversion that certainly prompted the investigation of possible extant correlations between risk aversion and investments in the trust game in the first place. The prediction that higher risk aversion correlates with lesser trust indeed sounds perfectly reasonable. However, we will argue that in the trust-game, and crucially in its basic one-shot version under anonymity, given that incurred risks are not explicitly known in the trust game, missing information in itself will negatively affect trust more specifically in the guise of ambiguity aversion.

Social aspects of trust, about which the question whether they correlate with uncertainty aversion (in the specific distinct forms of risk aversion or ambiguity aversion) also affect investments made. Individuals presumably fear betrayal in contexts (one-shot, anonymity, hence absence of obvious reputational effects) wherein strategic and opportunistic considerations might prevail on the part of the trustee. Trust, as still measured in terms of levels of invested amounts, certainly depends on individuals' degrees of betrayal aversion. We restrict our consideration of various social factors potentially influential in trust game behaviors to this particular one. Our main goal in this article is then to assess what some recent works in experimental economics and neuroeconomics have to say on the correlation between betrayal and uncertainty aversions in the context of trust-game experiments, and whether possible lessons on the nature of trust are to be drawn from there. Numerous studies in these joint fields address the issues of betrayal aversion and risk aversion, but still few, on which we will then bear special attention, address the possible correlation between betrayal aversion and ambiguity aversion.

The outline of this article is the following. In the first section 1.1 we report on the absence of observable correlations, across some crucial experiments by Eckel and Wilson (2004). Apparently, this result seems to indicate that there is no obvious correlations, with respect to the involvement of psychological attitudes toward risks, between trade-offs among variably risky lotteries and trade-offs between certain payoffs and a risky investments in the trust-game. This leads to investigate, in

section 1.2, the plausibility that a purely social dimension of trust might be prevalent in the determination of trusting behavior. But this social dimension is considered in the socially negative aspect of betrayal aversion limiting offers rather that the positively pro-social behavior understood in terms of altruism or other pro-social dispositions. We suggest that betrayal aversion itself relies on a more basic psychological aversion, crossing over the individual and social decision-making realms, which is our distaste for ambiguous signals. In section 2.1 we remind what aversion to ambiguity conceptually and experimentally amounts to and in section 2.2 we report on an experiment the authors have run evidencing some actual negative correlation between amounts invested in the trust game and levels of ambiguity-aversion elicited on a lotteries-choice menu. Section 3.1 summarizes important studies in neuroeconomics on the neural bases of ambiguity processing, showing the major role of brain structures such as the orbitofrontal cortex and the amygdala. Interestingly, and leading to some, hopefully cautious, extrapolation on our own, we likewise report in section 3.2 on the involvement of these structures in the display and inhibition of betrayal aversion in socio-economic contexts.

1.1. An absence of correlation between trust and risk aversion

Can trust behavior be predicted in a Trust Game (TG) by measuring attitudes towards risk? Does trust present analogies with a risky choice or does it fall into another category? In order to answer these two related questions, several experiments have been carried out. Their results suggest a negative answer: the amounts sent by an investor in a one-shot TG do not appear to be correlated to various measures of individuals' risk aversion.

Eckel and Wilson (2004) focus on the relationship between attitudes toward risk and a trust decision vis-à-vis an anonymous partner. The authors used two approaches to measure risk aversion. The first was based on a survey, while the other consisted in the applying the measure by Holt and Laury (2002). The subjects had to take part in one of three treatments planned for the Trust Game: no information about the partner; information about some of the partner's characteristics (e.g. sex, favorite color); or being shown a photo of the partner. In addition to participating in the Trust Game, subjects had to take part in another card game entailing a risky decision, through which they had to choose between a sure gain and a lottery aimed at reproducing the gains of a standard Trust Game ($20, 1/10; $15, 2/10; $10, 4/10; $5, 2/10; $0, 1/10). The observations made by Eckel and Wilson (2004) are characterized by a lack of significant correlation between measures of attitudes towards risk and the trust decision of investors in the Trust Game. The risk aversion coefficient, measured by using the measure developed by Laury and Holt (2002), was not correlated to the trust decision. Likewise, no correlation between risky decisions made during the card game and trust decisions observed during the various Trust Game treatments could be reported.

While the results by Eckel and Wilson (2004) show that trust decisions cannot be predicted neither by the Holt/Laury (HL) indicator nor by decisions observed in a risky game, similar in their mathematical expectation structure to the Trust Game, they also incidentally reveal that this same HL indicator does not predict decisions observed in this risky game. If Holt and Laury's measure of risk aversion proves unable to predict decisions in a risky context analogous to the Trust Game, it becomes difficult to interpret the failure of the HL indicator to predict trust decisions. It sheds

suspicion on the fact that this indicator is liable to be convincingly used in order to demonstrate that trust is distinguishable or analogous to risk. To overcome this state of indeterminacy, Houser *et al.* (2010) studied measures of individual attitudes to risk in a series of four Trust Game treatments designed to control systematically what distinguishes a context of risk from one of trust: prosocial motivations and ambiguity in the distribution of gains. Houser *et al.* (2010) based their experiment on four treatments of the investment game: two trust treatments and two risk treatments. These treatments enabled a gradual shift from a context of risk to a context of trust. Thus, in the first treatment, the investor was faced with a standard individual decision-making problem, whereas in the second treatment, the same decisions made by the investor affected the situation of a passive trustee. The third treatment replicated the BDM "social history" protocol (Berg, Dickhaut, McCabe, 1995), in which the subjects are informed of how individuals have played in the past. The fourth treatment was an exact replication of the BDM investment game procedure. Houser *et al.*'s (2010) main objective was to test the following hypothesis: the HL measures of attitudes to risk predict decisions in "risk" treatments, but not in "trust" treatments.

Similarly to Eckel and Wilson, Houser *et al.* (2010) failed to find a correlation between measures on risk aversion on the Holt and Laury indicators and investments observed in the various treatments, even though the relation pointed in the positive direction without reaching a significance threshold. Generally speaking, trust decisions are distinguished by greater dispersion: while risky investment decisions are more concentrated around a threshold at approximately the half-way point of the endowment, trust investments are more frequently made at extreme positions (investing zero or the full endowment).

Kanagaretnam *et al.* (2009) also fit in the process of identifying or dissociating the role of preferences with regard to risk from the role of social preferences in explaining behavior observed in the Trust Game. They base their experiment on a "social value orientation" indicator that enables them to classify individuals according to how pro-social their preferences are. Their statistical results show that trust increases with the degree of social orientation. Conversely, preferences with regard to risk, measured using the mechanism developed by Becker *et al.*(1964)[1], show no correlation with investment behaviors in the Trust Game.

Schtecher (2006) reaches a very different conclusion than the ones published by Eckel and Wilson (2004), Houser *et al.* (2010) and Kanagaretnam *et al.* (2009). Her experiment was carried out in rural Paraguay. It aimed at separating the dimension of trust (i.e. measured in terms of the subjective probability of expected returns) and risk aversion in the amount sent. To achieve this, two games were used: a lottery (which is supposed to replicate the gains of a real Trust-Game) and a Trust-Game for each subject. The analysis used three regressions to explain the amounts invested in the lottery, and the amounts sent in the Trust-Game, controlling for the amounts invested in the lottery or not. Her results tend to show that risk aversion partly explains the amount sent to the trustee.

[1] Kanagaretnam *et al.* (2009) applied the "two-stage lottery" mechanism. They elicited the certainty equivalents of 24 lotteries in order to identify whether subjects were more or less risk-adverse using an indicator based on the number of times that the certainty equivalent was higher than the mathematical expected gain.

Yet several aspects of the protocol may have parasitized the amounts sent in the TG and altered the scope of the results. First, the gains of the lottery were communicated to participants before they took part in the TG, which may have triggered a loss-aversion behavior (a wealth effect) or conversely a risk-taking behavior (house money effect), and introduced a bias in the relationship between subjects' appetite for risk and their actual risk-taking. Moreover, the fact that the lottery was systematically played before the TG may have led the subjects to consider the TG to be a lottery game as well. Ultimately, the protocol may have introduced a bias tending to increase the amount sent in the TG. The TG was not double-blind which could also have modified the average amount sent to a substantial degree (by increasing it). As this study was different from others in terms of both results and experimental conditions, it appears difficult to draw clear lessons from it.

Overall, research tends to conclude that there is no correlation between individual preferences for risk and the trust that individuals put in their trust game partners. It therefore makes sense to investigate beyond the non-social dimension of trust in order to apprehend it in a more general context of inter-individual relationships.

1.2. Betrayal aversion

Bohnet and Zeckhauser (2004) and Bohnet *et al.* (2008) have analyzed one central aspect of the social dimension of trust – conceived of as an interpersonal relationship based on beliefs and individual preferences, in particular aversion to betrayal. For Bohnet and Zeckhauser, by contrast with their behavior when they face lottery choice, individuals require greater expected returns in a TG to hedge betrayal risk. Neuroscientific findings seem to validate the prevalence of this particular social dimension of trust.

The objective of Bohnet and Zeckhauser (2004) was to find out whether the decision to invest – in the Trust Game – can be assimilated to a risky choice, or whether this decision involves an additional risk premium in order to compensate for the costs of a breach of trust. Their experiment involved three different treatments, each systematically offering the same three potential gains: G > S > B. All the subjects were thus forced to arbitrate between these gains in three contexts. The first context involved an individual decision, whereby the individual chose between a sure gain (S=10) and an ambiguous lottery – i.e. with unknown probabilities – entailing two outcomes (G=15 and B=8). The second context was a "risky Dictator Game", in which the individual chose between the same sure gain (S=10) and the same ambiguous lottery (G=15 or B=18, with unknown probabilities). But in this latter case, the choice made had consequences for an anonymous partner paired with the individual: when the decision maker chose to keep S, the partner received S as well; when the DM chose the lottery, the gains of both players were jointly defined by the lottery: either G=15 and H=15 respectively for the DM and his partner, or B=8 and C=22. It should be pointed out that the lottery shares, for each outcome, a same amount of 30 units (G+H=30, B+C=30). The third context was a "one-shot Trust Game", with the decision-maker choosing between keeping 10(S) for himself and trusting an anonymous partner who could send back 15 (G) or 8 (B) (while keeping respectively 15(H) or 22(C)). In the event of the trustee's betrayal – returning 8 (B) – investor received an amount lower than the amount 15 (G) he would have received from a loyal partner, and also lower than if he had not trusted his partner 10 (S).

These three contexts are all characterized by a situation of uncertainty (or ambiguity), with the third context adding a risk of betrayal to the ambiguity of the outcome. The second context places the individual in a social context with no risk of betrayal from his partner, unlike the third context. Significant differences between these treatments highlight the influence of betrayal aversion. In order to test the hypothesis that there is a specific aversion to betrayal, Bohnet and Zeckhauser (2004) elicited, for each of the three treatments, minimum acceptable probabilities (MAPs) of winning G from which the subjects choose to expose themselves to "risk" (i.e. lotteries or the Trust Game) rather than to guarantee for themselves the sure amount S. If the individuals apprehend trust as a purely risky choice, the minimum acceptable probability is unlikely to be affected by the decision-making context. Their findings show that minimum acceptable probabilities do not vary significantly in the contexts of choosing between the lottery and the risky Dictator Game. However, MAPs are significantly higher in the Trust Game context. According to Bohnet and Zeckhauser (2004), the higher risk premium in the Trust Game context supports the hypothesis that a betrayal of trust represents an additional cost linked to the social dimension of trust. The results of neuroscience experiments lend credence to the thesis of trust as a social factor, as we think is shown in particular by Kosfeld *et al.* (2005) and Baumgarter *et al.* (2010) as discussed in section 3.2 below.

2.1 Trust and ambiguity

The role of the social risk of betrayal (and player 1's aversion to that risk) in trust is apparently established. Although the effect of risk aversion proper is not demonstrated, it nevertheless would appear difficult to definitively rule out that individual non-social factors play a decisive role in trust. Instead of a risk/trust analogy, we should instead consider a parallel with an ambiguous decision situation.

A primary reason to think that trust, in its non-social dimension, would be determined more by an aversion to ambiguity than by an aversion to risk is that the probability distribution of expected returns in the trust-game, especially in its one-shot version, is unknown to the investor. A second reason, empirically based, is that previous studies – either behavioral or neuroeconomic – on the risk-trust relationship do not refute this hypothesis and might on the contrary support a trust-ambiguity negative correlation hypothesis. Studies of a potential link between risk and trust rely on the hypothesis that individuals make subjective anticipations of the probability of being betrayed. Yet, there is no reason to think that individuals would be able to accurately assess their risk of being exposed to betrayal. Instead of subjective probabilities (defining the notion of risk as it is used by Knight), it is reasonable to think they more simply make guesses as to the level of uncertainty that they face.

Therefore, risk, in the Knightian (1921) sense[2], may not be the most adequate informational characterization of the situation faced by investors in a Trust Game. Indeed, in the one-shot Trust Game, the investor interacts with an anonymous trustee, and his decision to trust is an ambiguous one (Ellsberg 1961) rather than a risky one due to the fact that, indeed, human partners are

[2] Risk, understood properly, relates to an informational context in which information about payoffs and their probabilities is available.

essentially unpredictable. What they lack precisely is the information about the probability of what variable counter-offers could be made by trustees, in the case they decide to trust them. The probabilities of returns are unknown and trust behavior may then correlate more specifically with aversion to ambiguity than with aversion to risk. Loosely speaking, trust in this situation is certainly a risky gamble, but speaking in more proper informational terms, it is instead tantamount to decision-making in a context of ambiguity. Secondly, one could further argue that past social interactions shape the investor's expectations on reciprocation when he enters the Trust Game. These expectations most likely take the form of imprecise and general beliefs, and an investment decision in the Trust Game is therefore analogous to a choice between a sure gain and an ambiguous lottery.

The notion of distrust provides us with a new perspective to apprehend trust as a phenomenon based on uncertainty aversion. On financial markets, a rise in uncertainty increases distrust among investors and shrinks financial trade (Guiso *et al.* 2008; Olsen 2008). The natural corollary is that a reduction of perceived uncertainty will restore a trusting climate conducive to trade. Trust and subjective uncertainty appear to be closely negatively linked. The home bias provides a strong illustration of how perceived uncertainty and trust are indeed correlated. National, local and employer's assets appear to be more trustworthy since they seem familiar and therefore less uncertain.

Previous research, far from refuting this hypothesis, lends it some credence. Firstly, the lack of correlation between risk aversion and trust can obviously only support the hypothesis that ambiguity aversion could be a determinant of trust. More interestingly, it is worth noting that hypotheses of betrayal aversion and ambiguity aversion are not incompatible. In the experiment conducted by Bohnet and Zeckhauser (2004), for the TG context, the decision-maker had to indicate the minimum acceptable probabilities (MAPs) of "trustworthy trustees", i.e. those he was prepared to trust. While we cannot deny the reality of the risk of betrayal, which is compensated for by the additional risk premium, we cannot rule out the fact that the individual may also understand this choice as an ambiguous one. The MAP of the TG was compared to the average percentage of player 2s in the group that stated that they wished to behave in a trustworthy fashion. If the MAP was higher than this percentage, the subject kept his endowment; otherwise, he played in the TG without knowing whether his partner was a trustworthy player or not. We could thus wonder whether this risk premium is not also fair compensation for the uncertainty in which the players find themselves. Indeed, if the probability of a gain from the lottery is unknown but definite, characterizing a risky situation, the gain of the TG is only defined by the group average, and will thus remain affected by the inherent uncertainty of the partner's play. In this hypothesis, subjects would react according to their aversion to ambiguity and their aversion to betrayal, by adding a risk premium when social circumstances imply an additional cost. Therefore, the existence of a betrayal aversion does not preclude the existence of an ambiguity aversion, which can also appear and affect trust decisions. Even if all individuals show a betrayal aversion, they may also be more inclined to invest in trust as they have a lower ambiguity aversion.

2.2 Demonstrating a negative correlation between trust and ambiguity aversion

Corcos *et al.* (2012)'s specific aim was to investigate potential relations between measures of ambiguity aversion and investments in a one-shot trust game. Resorting to Chakravarty and Roy's

(2009) methodology with respect to the measure of ambiguity aversion and that of Holt and Laury (2002) to elicit risk aversion, In Corcos et al. the experimental procedure allows to explore possible correlations between these measures and the behavior of the investor in the one-shot trust game. The authors first proceeded to the measure of risk and ambiguity attitudes among participants by eliciting trade-offs between series of pairs of lotteries presenting contrasts between these alternative probabilistic backgrounds. Then, all pairs of subjects were faced to a one-shot trust game environment. The amount engaged by the investor is a direct measure of the risk taken vis-à-vis an anonymous player, which is certainly a way of conceiving of trust. But a more precise apprehension of that decision context likens the one-shot trust game to a choice under ambiguity, participants having no precise information about the probabilities of how their offer will be reciprocated. Since the potential states of nature are known, but not their associated probabilities, one can suggest that the decision to trust takes place within a context of ambiguity. Under this assumption, a negative impact of aversion on trust is expected. The authors also checked the non-significance of risk aversion indicator with offers in the OSG, in line with results in the literature we have discussed above.

Figure 1: Interplay between offers and risk aversion

Figure 2: Interplay between offers and ambiguity aversion

No correlation could be found between attitudes towards risk and the behavior of investors in their one-shot trust game (see figure 1 above). However, one-shot decisions to trust were correlated with attitudes towards ambiguity (see figure 2 above).

Table 1: Models Estimations.

Variable	Models: OSG Coefficient (p-value)	
	H1	H2
RR	-0.126 (0.449)	
RA		-0.553 (0.021)*
Constant	5.079 (0.000)**	7.359 (0.000)*
Log Likelihood	-212.897 (0.448)	-210.484 (0.0202)*

Pseudo R2	0.0013	0.0126

*:1% significant
(): p-value
Number of observations=93

Besides, the analysis used a double censored model explaining investor's offer by reference to an indicator of aversion to risk (RR) or to ambiguity (RA). We found –see Table 1 above—that the coefficient of RR was not significant –showing that a trust decision cannot be reduced to a risky choice—while that of RA was significantly negative meaning that investments in OSG decrease with aversion to ambiguity) and brings support to our hypothesis.

Together, both results give support to the idea that individuals subjectively perceive trust as a strategic decision involving ambiguity.

3.1 Neural bases of ambiguity aversion

Several studies have dealt with the issue of the neural bases of ambiguity vs. risk. Our goal is not to comprehensively review that sub-field of decision-neurosciences but to select hints and evidence that would consistently point toward the fact that neural bases of ambiguity, when elicited in the context of individual decision-making over ambiguous lotteries might also be involved in social ambiguous decisions. Levy *et al.* (2010)'s starting points, for instance, in their fMRI study are very close to our present motivation. They remind that risk aversion and ambiguity aversion are largely uncorrelated phenomena across subjects. They then insist on the difference between risk and ambiguity and their differentiated neural substrates. However, their more complete brain-imaging strategy is to look beyond these differences and investigate whether, in spite of distinct neural activities associated with these two distinct probabilistic settings, a neural system common to both can be evidenced and what its function would then be. fMRI reveals that this common system, consisting of at least the striatum and the medial prefrontal cortex, was found to represent subjective value under both conditions of risk and ambiguity. It is interesting to note that values can be neurally processed in abstraction of risk vs. ambiguity, i.e. of the processing of the nature of probability contexts in which values are embedded.

The fact that value and probability are distinctly implemented in the brain is important to our present purpose, precisely because we could conceive of increasing trust (in terms of augmentation of reciprocal payoffs in the repeated trust game) as an informational transformation, in the partners' brains, from ambiguity to risk. Hsu *et al.* (2005) have run, with that respect, a useful and pioneering study. Their starting point is Ellsberg's paradox, reminding that in decision theory, ambiguous information about probabilities should not affect choices. Using functional brain imaging, they show that the level of ambiguity in choices correlates positively with activation in the amygdala and orbitofrontal cortex. Neurological subjects with orbitofrontal lesions were insensitive to the level of ambiguity and risk in behavioral choices. These data suggest a general neural circuit responding to degrees of uncertainty.

In Hsu and colleagues' experiment Ellsberg's paradox is implemented in a simple graphic way. Each stimulus screen presents on one side a gamble on two cards-decks and two pieces of information: i) in the risky condition, the number to total cards in the two decks and the gamble payoff and ii) in the

ambiguity condition, a cancellation of one of this piece of information, and on the other side a sure payoff on the left that the subject could prefer to the risky or ambiguous gamble. It has to be noted that sure payoffs are degenerate risky payoffs and that ambiguity aversion, according to the authors, should not be affected by this fact. It makes, however, their stimuli close to a branch (the typical A-B one) of the Allais paradox, which has been treated from the same neurobiological perspective in terms of the determinant role of orbitofrontal structures (Bourgeois-Gironde, 2012).

From the perspective of post-Savagean decision-theory, the question is to know whether such an experimental implementation of ambiguity will discriminate between extant accounts of ambiguity aversion. From our present perspective, it makes a difference which psychological and neural mechanisms are probed in this experiment to carry over analyses of potentially similar mechanisms involved in trust or distrust behavior. For example, if ambiguity aversion underlying psychological and neural mechanisms is intuitively tantamount to following a Choquet integral procedure, Hsu *et al.* expect that this should be reflected in levels of brain activities in uncertainty-related brain areas, namely ambiguity aversion being subtracted from risk aversion in those areas in terms of amount of neural activities. Differently, if ambiguity aversion is an altogether different phenomenon from risk aversion, stimuli presenting ambiguous gambles should involve neural structures that are not observed to be activated in the case of risky gambles. According to which prediction is realized we will be able to give a positive or negative comparative assessment of neural overlap between uncertainty and trust aversions.

As we said above, areas that were more active during the ambiguous condition relative to the risk condition included orbitofrontal cortex (OFC) and amygdala. OFC is a brain-area that has mustered much attention among neuroeconomists, stemming back to Damasio's seminal studies with ventral medial patients presenting emotional deficits that consequently made impossible bottom up emotional signals to give impetus to intact cognitive appraisals of the contingencies in a situation in view of an appropriate rational course of action. The OFC is precisely seen by many researchers (see e.g. Critchley *et al.* 2001; Camille *et al.* 2004) as the locus of integration between emotion signals and cognitive treatment in decision-making. Observation of significant OFC activity in Hsu *et al.*'s task certainly indicates that cognitive and emotional processes converge in shaping an observable course of action when the subject faces decisions under ambiguity. Emotional signals are initially processed in the amygdala and can be said to correspond, in the present context, to the processing of uncertain events. It is notable that amygdala activities are also correlated with missing information in the context of social contexts as documented by Phelps *et al.* (2000).

We can see how interdisciplinary endeavors may help shed light on the heterogeneous determinants of trust. Yet it is certainly too early to speculate on more specific neural overlaps between ambiguity aversions arising when facing decisions under ambiguity (in the decision-theoretical strict sense) and ambiguous social situations (in their less rigorous characterization in terms of unfamiliarity and ensuing distrust). There are two distinct ways to think about the potential overlap, and its rationale, of neural signatures in the context of individual or social decision-making under conditions of ambiguity, the latter being defined more or less rigorously. Are amygdala activities observed in individual decision-making over ambiguous Ellsberg-like lotteries a particular instance of a neural system that was shaped in co-evolution with the shaping of our social brain, or, differently, are the

activities that are observed in social contexts wherein unfamiliarity and uncertainty about others' behavior loom derived from a general probabilistic learning mechanism in the brain?

Understanding the neural basis of choice under uncertainty, in the broader sense including both risk and ambiguity, is important, whatever the answer we give to the previous question. Uncertainty is a basic feature of our economies and we could imagine that our brain uses resources to process, possibly in refined way, distinct levels and types of uncertainty, and that those resources encompass social and non-social signals. In fact, both the amygdala and the OFC are known to process salient ambiguous stimuli. This function has been more especially associated with the amygdala (Wahlen, 1998), a brain structure which is alerted when information is missing and that decision made on the information available entails unknown and potentially dangerous consequences. At this juncture our speculative question above can be made more precise. Can we consider that a default-mode common to individual and social decision-making is one that is specialized in processing ambiguous situations, rather than ones in which risk and consequences are known to the individual? In that sense neither risk nor ambiguity would derive one from the other in terms of a differential treatment within a single brain system of levels of probabilistic information available in a given situation, but, rather, a default neural appraisal of choice contexts in terms of missing information or ambiguity (it is reasonable to observe that agents seldom have full probabilistic knowledge of the outcomes of their individual or strategic decisions) has been prioritized in the brain.

A tighter attempt to suggest a neural derivation of ambiguity-processing from risk-processing is performed in Huettel *et al.* (2006) but this study precisely concludes in the negative about the potential neural reality of this derivation in the sense that decision making under ambiguity is not shown to represent a special, more complex case of risky decision making, but, again, that these two forms of uncertainty are rather supported by distinct neural mechanisms. Moreover, ambiguity aversion is not correlated to risk aversion according to several authors (since at least Camerer and Weber 1992). This behavioral result is in accordance with studies that show distinct neural bases for decisions under ambiguity and under risk.

But this default-mode, to corroborate our hypothesis, should also be involved in stylized ambiguous social decision-making. In van den Bos *et al*. (2009) a trust-game is implemented and two main situations are studied from the standpoint of the trustor. When the latter refuses to invest (distrust), neural activities are observed in the anterior Medial Prefrontal Cortex (aMPFC). The authors suggest that such activities are specifically correlated with non-social processes, like reward assessment. On the other hand neural activities observed in the right Temporal-Parietal Junction (rTPJ) in the case of cooperation and reciprocation of trust are indications or really social processes taking place. The TPJ is a region that is involved in the theory of mind network. Let's note that Decety and Grèzes (2006) have shown that lTPJ is selectively involved when I realized that I am imitated, whereas rTPJ is selectively involved when I imitate somebody else. In the trust game, reciprocation as encoded in the rTPJ may underpin the fact that the subject adopts an imitative (tit-for-tat) strategy in the game. Interestingly, then, Trust Aversion (or distrust, betrayal aversion) may involve neural mechanisms which (unlike trust itself) are not properly social. It is as if, when these mechanisms detect signals which are potential threats on cooperative equilibria in a repeated situation, a global neural shift took place from social neural networks (one proper to theory of mind in particular) toward

"individualistic" networks. Those individualistic networks, especially the MPFC, are the same ones that Huettel *et al.* (2005) have evidenced in the case of individual decision under ambiguity.

3.2 Do neural bases of betrayal aversion tap into mechanisms associated with ambiguity processing?

Reduction of betrayal aversion is correlated with the modulatory role of the human hormone oxytocin as it has been popularized by a flurry of recent studies in "hormoneconomics". It has been shown, even more closely to our present interests, that oxytocin modulates the link between affective attachment and cooperation through reduced betrayal aversion (De Dreu, 2011). A message should be made clear from the onset, though. Oxytocin acts as a modulator. It is a neuropeptide which supports complex mediations between emotional and social behaviors (maternal attachment, romantic love anxiety, shared emotions, etc.). The affective states induced by this chemical are the necessary mediators to enhance social coordination. There is no "direct" influence on social coordination in absence of those mediating complex neurophysiologic affective states. It is with this reminder in mind that economic experiments focusing on the role of that hormone in correlation with observed trusting trading relationships should be considered.

Seminal experiments of this type have been run by Kosfeld *et al.* (2005) and Baumgartner *et al.* (2008). Kosfeld has shown that, using a trust game, oxytocin increases the willingness to trust of investors by increasing the acceptance of a social risk. We then again note that it is not social signals conveyed by trustees' returns or attitudes that are made more salient by the experimental inhalation of sprayed oxytocin but the willingness to engage in social (potentially risky) intercourses that is enhanced. Or to put it differently, it is an individual affective state (correlated with pro-social decisions) that is doped, that helps overcome an otherwise natural propensity to distrust, or betrayal aversion. So players 1, in oxytocinized trust-games, have still the same perception (still possibly negatively biased) of incurred risk in the trust-game, they still have the same appraisal of the contingencies presented by co-players, but they are more willing to incur these actual or perceived risks, which reflects, from a mere observational viewpoint, to a reduction of betrayal-aversion.

Kosfeld *et al.*'s experiment proceeds as follows. Two main conditions were tested with respect to which the effects of oxytocin intranasal absorption were measured. In the trust-game the risk incurred by the trustor is due to the relative unpredictability of the trustee's behavior. This is in this condition that the authors deem they are in a position to screen out whether oxytocin affects risk aversion or trusting behavior itself. This observation is made possible by the use of a control risk-condition in which the investor faces the same choices as in the trust game but responses to his acts were determined, in full knowledge of it, by a random mechanism. It is plausible to say then that in the social and individual conditions of Kosfeld experimental paradigm subjects face the same risks but the nature or source of the risks incurred is the manipulated variable. Emotional or mood effects produced by oxytocin should be the same across those conditions. Kosfeld *et al.*'s results show that oxytocin did not modify, by comparison with a control group, the willingness to incur risks in the risk-condition but substantially increases the willingness to incur "social risks" in the trust-game condition. The authors do not hesitate to interpret this result in terms of a specific reduction of betrayal-aversion in the trust-game due to augmentation of oxytocin levels in the organism, on the premise that betrayal is not involved in the risk randomized condition. Moreover oxytocin does not

affect reciprocation strategies, simply, again, social bets in condition of missing social information. When the game is repeated, social information is regularly processed and behavior adapts to standard reciprocity. A focus on the investigation of the specific neural bases of betrayal aversion is thus suggested by this seminal experiment. This result was confirmed and further investigated and analyzed by Baumgartner *et al.* (2010).

Baumgartner *et al.* (2010) replicated Kosfeld *et al.* (2005) and supplemented it with fMRI investigations of the functional neural bases of betrayal aversion and its reduction by oxytocin inhalation. Their investigation relies on previous studies that demonstrated the involvement of particular subcortical structures, with a special focus on the amygdala. Lesions of the amygdala are correlated in patients with hypersocial behavior, absence of fear, but also systematic deviation in assessing others' trustworthiness on the basis of facial traits (Adolphs and Spezio, 2006). Baumgartner and his colleagues ask the question to know how oxytocin modulates the amygdala reactivity in situations of potential social risks such as stylized through the trust-game.

Behavioral results of Baumgartner and Kosfeld's experiments are consistent. Oxytocin does not affect reactions to feedbacks in a pure random risk condition but affects responses to social feedbacks by pondering down the impact of negative feedbacks on subsequent trusting behavior. These behavioral results, as underlined by Baumgartner and his colleagues are also consistent with the hypothesis that betrayal aversion is an important underlying psychological parameter in the trust game (Bohnet and Zeckhauser, 2004) and that oxytocin appears to reduce the impact of betrayal aversion (Kosfeld *et al.* 2005). Brain imaging results show significant neural activity differences between subjects having inhaled oxytocin and a control placebo group, only in the case of the social condition, not in the risk condition. Oxytocin clearly seems to modulate our exposition to social risk, not to risk *per se*. More precisely, the modulatory effect of oxytocin was detected in the amygdala at the moment when the participants treated negative feedback in response to trusting behavior on their part. The neuropeptide effect is a reduction of social fear, not a modification of risk perception. The consequence of fear reduction is an immediately observable increased trusting behavior, but it should be clear that trusting behavior is not directly enhanced but follows from the inhibition of betrayal aversion.

As it stands, this result does not make more precise the idea that reduction of betrayal aversion is a modification of the perception of the nature of uncertainty incurred in the trust game, to the extent that, precisely, that risk behavior is not modified through oxytocin. It is also true that the amygdala activity is not modulated by oxytocin inhalation in the risk condition of Baumgartner's experiment, unlike what happens in the trust game condition. But, for the very same reason, this result does not either contradict behavioral results in experimental economics that could not demonstrate any obvious correlation between risk attitudes and trust behavior in the trust game. Risk aversion appears to be shown, by two very distinct research approaches, to be partially disconnected from our appraisal of social uncertainty. This may sound paradoxical as long as one understands the type of uncertainty associated with betrayal aversion in the trust game in terms of the technical notion of risk, or at least the notion of risk that was actually implemented in the experiments we refer to. Namely lotteries proposed in Baumgartner and Kosfeld's experiments, even though they corresponded to the actual counteroffers made in the siding trust-game, were informationally complete from the point of view of the investor. He knew the risk he was taken and the impact of

this determinate knowledge was not modified by oxytocin. It would be another matter to check whether partially missing, or ambiguous information and its impact on individual decision-making would be in fact modified by oxytocin inhalation, but authors, so far, did not test this possibility. As it stands, then, those results cannot either contradict the hypothesis that betrayal aversion is deeply psychologically and potentially neurally (as we will propose more precisely below) correlated with ambiguity-aversion.

Conclusion

We suggest that the treatment of offers and trades in bargaining-games may alternatively resort to two distinct neural networks. Regions involved in social cognition are activated in cases of reciprocation. In this context a proper social dimension emerges in the game that is treated as such by the brain. But in cases of trust or betrayal-aversion, those very same regions do not seem to play a role. Another network is then involved which is common to individual decision-making under ambiguity. It seems then worthy to deepen investigations of a common neural system associated with social (trust) and individual (ambiguity) aversions to uncertainty. It would be tempting, however, to counterbalance this dominant hypothesis that we have discussed by a possible rival or complementary one, according to which bargaining games usually implemented in the context of the neural and behavioral study of trust do not simply carve out social uncertainty. Of course in one-shot and repeated trust games an individual faces distinct levels and types of uncertainty, but it should also be emphasized that those experimental bargaining situations are, normally, ones in which agreements are to be reached. The preference for status-quo, compromise, equilibrium, may actually prevail among most not pathologically antisocial players of these games, and this propensity should be swiftly revealed in the emergence of social situations. It is when the rapidly emerging social equilibria threaten to be broken that an alternative default neural network takes place and inhibits the propensity to negotiation and cooperation. The correlation between trust and ambiguity is far from being systematically addressed in the extant literature. It appeared to us that so far studies in experimental economics and neuroscience have not excluded our hypothesis that betrayal and ambiguity aversions share a lot, from decision-theoretical, psychological and neurobiological standpoints. To that extent, it appeared to us interesting, as economists, to see what neuroeconomics could contribute to that issue. Studies in neuroeconomics have identified neural bases for ambiguity and trust, and significant, yet to be fully interpreted and replicated, some overlap between the two sets of respective brain activities. Further experiments are in order.

References

Adolphs, R., and Spezio, M.L., (2006). Role of the amygdale in processing visual social stimuli. Progress in Brain Research, 156, 363-378.

Baumgartner, T., Heinrichs, M., Vonlanthen, A. Fischbacher, U., and Fehr, E. (2010). Oxytocin shapes the neural circuitry of trust and trust adaptation in humans. Neuron, 58, 639-650.

Becker, G., Degroot, M., and Marschak, J. (1964). Measuring Utility by a Single-Response Sequential Method. Behavioural Science, 9, 226-232.

Berg, J., Dickhaut, J., and McCabe, K. (1995). Trust, Reciprocity and Social History. Games and Economic Behavior. 10, 122-142.

Bohnet, I. and Zeckhauser, R. (2004). Trust, risk and betrayal. Journal of Economic Behavior and Organization. 55, 467-484.

Bourgeois-Gironde, S. (2012). Ellsberg and Allais in the brain – what's next in the neurobiology of decision-theoretical paradoxes? *in* Watanabe S. (ed) *Logic and Sensibility*, Keio University Press.

Bourgeois-Gironde, S. and Corcos, A. (2011). Discriminating strategic reciprocity and acquired trust in repeated trust-game, *Economics Bulletin*. 31, 177-188.

Camerer, C. and Weber, M. (1992), Recent developments in modeling preferences: Uncertainty and ambiguity. Journal of Risk and Uncertainty, 5, 325-370.

Camille, N., Coricelli, G., Sallet, J., Pradat-Diehl, P., Duhamel, J.R., and Sirigu, A. (2004). The involvement of the orbitofrontal cortex in the experience of regret. Science. 304, 1167-1170.

Chakravarty, S. and Roy J. (2009). Recursive Expected Utility and the Separation of Attitudes towards Risk and Ambiguity: An Experimental Study. Theory and Decision, 66, 1999-228.

Critchley, H., Mathias, C., and Dolan, R. (2001), Neural Activity in the Human Brain Relating to Uncertainty and Arousal During Anticipation, Neuron, 29, 537-545.

Corcos, A., Pannequin, F., Bourgeois-Gironde, S., (2012). Is trust an ambiguous rather than a risky decision? Economics Bulletin, forthcoming.

Cox, J. (2004). How to identify trust and reciprocity. Games and Economic Behavior, 46(2), 260–281.

Decety, J. and Grèzes, J. (2006). The power of simulation: Imagining one's own and other's behavior. Brain Research. 1070, 4-14.

De Dreu, K. (2012). Oxytocin modulates the link between adult attachment and cooperation through reduced betrayal aversion. Psychoneuroendocrinology, in press.

Eckel, C. and Wilson, R. (2004). Is Trust a Risky Decision? Journal of Economic Behavior and Organization, 55, 447-465.

Ellsberg, D. (1961). Risk, Ambiguity and the Savage Axioms. Quarterly Journal of Economics, 75, 643-669.

Guiso, L., Sapienza, P., and Zingales, L. (2008). Trusting the Stock Market. Journal of Finance, 63, 2557-2600.

Holt, C. and Laury, S. (2002). Risk Aversion and Incentive Effects. American Economic Review, 92, 1644-1655.

Houser D., Shunk D. and Winter J. (2010). " Distinguishing trust from risk: An anatomy of the investment game ", Journal of Economic Behavior & Organization, 74(1-2), 72-81.

Hsu, M., Bhatt, M., Adolphs, R., Tranel, D. and Camerer, C. (2005). Neural Systems Responding to Degrees of Uncertainty in the Human Brain. Science, 310, 1680-1683.

Huettel, S., Stowe, C.J., Gordon, E.M., Warner, B.T. and Platt, M.L., (2006). Neural Signatures of Economic Preferences for Risk and Ambiguity. Neuron, 49, 765-772.

Kanagaretnam, K., Mestelman, S., Nainar, K., and Shehata, M. (2009). The impact of social value orientation and risk attitudes on trust and reciprocity. Journal of Economic Psychology, 30, 368-380.

Knight, F. (1921). *Risk, Uncertainty and Profit*. New York, Houghton Mifflin.

Kosfeld, M., Heinrichs, M., Zak, P., Fischbacher, U., and Fehr, E. (2005). Oxytocin increases trust in humans. Nature, 435, 673-676.

Levy, I., Snell, J., Nelson, A.J., Rustichini, A., and Glimcher, P. (2010). Neural Representations of Subjective Value Under Risk and Ambiguity. Journal of Neurophysiology, 103, 1036-1047.

Olsen, R.A. (2008). Trust as risk and foundation of investment value. Journal of Socio-Economics, 37, 2189-2200.

Phelp, E., O'Connor, K., Cunningham, W., Funayama, E., Gatenby, JC., Gore, J., Banaji, M. (2000). Performance on indirect measures of race evaluation predicts amygdala activation. Journal of Cognitive Neuroscience, 12, 729-738.

Schechter, L. (2007). Traditional trust measurement and the risk confound: an experiment in rural Paraguay. Journal of Economic Behavior and Organization. 62, 272-292.

Van den Bos, W., Van Dijk, E., Westenberg, M., Rombouts, S., and Crone, E. (2009). What Motivates Repayment ? Neural correlates of reciprocity in the trust game. Social Cognitive and Affective Neuroscience, 4, 294-304.

Wahlen, P.J. (1998). Fear, vigilance and ambiguity: initial neuroimaging studies of the human amygdale. Current Directions in Psychological Sciences, 7, 177-188.