

Facing expectations: Those that we prefer to fulfil and those that we disregard

Christophe Heintz * Jérémy Celse † Francesca Giardini ‡ Sylvain Max §

Abstract

We argue that people choosing prosocial distribution of goods (e.g., in dictator games) make this choice because they do not want to disappoint their partner rather than because of a direct preference for the chosen prosocial distribution. The chosen distribution is a means to fulfil one's partner's expectations. We review the economic experiments that corroborate this hypothesis and the experiments that deny that beliefs about others' expectations motivate prosocial choice. We then formulate hypotheses about what types of expectation motivate someone to do what is expected: these are justifiable hopeful expectations that are clearly about his own choices. We experimentally investigate how people modulate their prosociality when they face low or unreasonably high expectations. In a version of a dictator game, we provide dictators with the opportunity to modulate their transfer as a function of their partner's expectations. We observe that a significant portion of the population is willing to fulfil their partner's expectation provided that this expectation expresses a reasonable hope. We conclude that people are averse to disappointing and we discuss what models of social preferences can account for the role of expectations in determining prosocial choice, with a special attention to models of guilt aversion and social esteem.

Keywords: altruism, expectation, dictator game, guilt aversion.

1 Introduction

Standard theories of social preferences assert that decision-makers care not only about their own material payoffs, but also about the material resources allocated to others. These theories concentrate on the final distribution of material benefits. Thus, inequity aversion, preference for social welfare (Charness & Rabin, 2002), conditional altruism (including strong reciprocity) (Fehr, Fischbacher & Gächter, 2002) and preference for fair outcomes (Rabin, 1993), are distributive preferences: the agent having such preferences prefers not just benefits for herself (self-regarding preferences) but also benefits for others (other-regarding preferences). She cares about how the benefits are distributed between her and others. In this paper, we defend the view that there are other preferences that cause prosocial choice, which are not about how goods or benefits are distributed. Rather, they are about the mental states of the protagonists (Dufwenberg, 2008). We say that they are mind-directed preferences.¹ The aver-

sion to disappointing one's partner is such a preference.² We will argue that it motivates much of the prosocial choice observed in lab experiments and in real life. People, we argue, conceive what is reasonably expected from them and they prefer to fulfil these expectations. This leads them to make prosocial choices.

We experimentally investigate the extent to which others' expectations foster prosocial decisions. To test the effect of expectations on altruistic choice, we use a modified version of the well known Dictator Game, to which we apply the Strategy Method (Brandts & Charness, 2011). In order to assess sensitivity to others' expectations, we asked each dictator to specify what she would transfer for a number of possible expectations. Previous experiments showing sensitivity to expectations have all compared behaviour across subjects in different conditions. We used a within-subject design, which allows us to observe whether individuals vary their transfer when nothing else but expectations vary. This is therefore a more direct and stronger test of the sensitivity to expectations. This method also enabled us to test hypotheses about which types of expectations people prefer to fulfil. A first hypothesis relies on Miceli and Castelfranchi's (Miceli & Castelfranchi, 2002) analysis of

Copyright: © 2015. The authors license this article under the terms of the Creative Commons Attribution 3.0 License.

*Corresponding author. Department of Cognitive Science, Central European University, Oktober 6 street 7, Budapest, 1051, Hungary. Email: heintzc@ceu.edu

†Burgundy School of Business

‡Laboratory for Agent-Based Social Simulation, ISTC-CNR

§Burgundy School of Business

¹Dufwenberg and others talk about "psychological games." They focus on the game theoretic models. We talk about mind-directed preferences because we focus on what truly motivates people and the content of the social preferences.

²This preference has been debated under the name of "guilt aversion" (Battigalli & Dufwenberg, 2007; G Charness & Dufwenberg, 2006; Ellingsen, Johannesson, Tjøtta & Torsvik, 2010; Reuben, Sapienza, & Zingales, 2009; Vanberg, 2008), but as Vanberg (2008) notes, guilt is a negative emotion that can be felt under several conditions other than disappointing others. It can be felt when behaving immorally, for instance. We thus use the term "aversion to disappointing". When we refer to the specifics of Dufwenberg and others' model, we revert to using "guilt aversion".

expectations: they distinguish forecasts about others' behavior depending on their valuations—positive or negative. We argue that people do not have a preference for fulfilling others' forecasts whatever these forecasts are; people prefer to fulfil only the grounded and justified hopeful expectations of their partners. Our second hypothesis concerns the justifiability of expectations: we show that expectations that are unjustifiably high are disregarded.

In the next section, we review the experimental evidence showing that many prosocial choices traditionally explained in terms of distributive preferences are in fact better explained by an aversion to disappointing. In the process, we specify the conditions in which the aversion to disappointing truly motivates prosocial choice, and we formulate hypotheses about which expectations people are willing to fulfil. The third and fourth sections explain our experimental tests of the hypotheses using a version of a dictator game. The last section discusses how the aversion to disappointing can be modeled and how it relates to other mind-directed preferences and to distributive preferences.

2 The aversion to disappointing: hypotheses and their current evidential support

2.1 Theory and models

The aversion to disappointing has been nicely specified with utility functions: in Battigalli and Dufwenberg's (2007) models of "guilt aversion", the utility function ascribes disutility to not fulfilling others' perceived expectations. If Ines and Jack interact with each other, then Jack might expect to derive a benefit (whose measure is e) from this interaction. Ines forms a belief about what Jack expects from the interaction (whose measure is b_e). The theory of guilt aversion asserts that Ines prefers to fulfil Jack's expectations and thus to satisfy b_e . The utility for Ines to make choice C is the utility derived from material gains minus the disutility of disappointing Jack. This disutility is measured as how much Jack's actual benefit fails to be up to what he expected—or, more precisely, what Ines thought he expected. The disutility is b_e minus the actual benefit that j obtains from the interaction (i.e., as a result of Ines making the choice C), or zero if this difference is negative, in which case Jack is not disappointed.³

The important insight of the model of guilt aversion is that agents obtain utility or disutility in function of what they think others think: is the other satisfied or disappointed by my action? The model is about a social preference, since it concerns others' mental states, and this social preference

³Battigalli and Dufwenberg (2007) give a more formalised characterisation.

can motivate pro-social choices. While the model of guilt aversion accounts for many aspects of the motivations that lead people to do what is expected, we will argue that it misses an important process: the assessment of the expectations. Is my partner simply predicting that I will do something or is he hoping that I will? Has he good reasons to hope that I do this thing or is it merely wishful thinking? Some experimental results questioning the model of guilt aversion (Ellingsen et al., 2010; Vanberg, 2008) call for making these specifications. These specifications are all the more needed because aversion to disappointing might well explain many of the experimental results meant to test the extent and conditions in which people make prosocial choices. Saying that people are averse to disappointing is hypothesizing that they have a willingness to do what is expected from them, which can motivate altruistic behaviour even in anonymous and non-repeated interactions. This hypothesis can take more or less radical forms depending on how much of altruism is said to be due to aversion to disappointing and how much is said to be due to other preferences such as the preference for fair distribution. In any case, we need to show that at least an interesting range of prosocial choices is due to aversion to disappointing. We review below experiments showing that much of the prosocial behaviour observed is due to aversion to disappointing rather than distributive preferences. We then we consider the more critical papers.

2.2 No prosocial choice if the partner has no justifiable expectations

The experimental results we now present suggest that an important part of the documented prosociality is better accounted for with aversion to disappointing than with some distributive preferences. Yet, it turns out that the the expectations of others has a motivating strength in some conditions but not in others. The goal of this review is to clarify which are those conditions and then derive hypotheses about the aversion to disappointing.

Dana and colleagues (Dana, Cain & Dawes, 2006) present some experimental results that strongly suggest that transfer in dictator games is motivated by an aversion to disappointing the receiver rather than by a distributive preference such as inequity aversion. Dana et al. first invited subjects to participate in a Dictator Game. Subjects were asked to share what they wished of an initial endowment with an anonymous recipient. The recipient had no choice to make. Right after the dictators had made their choice, the experimenter gave them the option to pay a small amount of their initial endowment (one dollar out of 10) in order to prevent the potential receivers from knowing that a dictator game had been played—this is called "exit option". This option provided a means for dictators to avoid disappointing, not by fulfilling expectations, but by preventing the formation of expecta-

tions of the potential receiver. Furthermore, this means cost only one dollar out of ten, while fulfilling expectations of informed receiver would probably cost much more. 28% of the subjects chose the one dollar exit option. (The results have been replicated with higher rates by Broberg, Ellingsen & Johannesson, 2007, and Hashimoto, Mifune, & Yamagishi, 2014; see also Trachtman et al., 2015, for a protocol where the cost of avoiding expectations is non-monetary.) People choosing the outside option included both people who had initially chosen to make no or very low transfers, who could now avoid disappointing at a sufficiently low cost, and those who had shared equally, who found a way to achieve the same goal—not disappointing the receiver—at a much lower cost. Distributive social preferences (inequity aversion, reciprocity, fairness, etc.) cannot account for this choice. By taking the exit option, subjects choose to have \$9 rather than having \$10 and the opportunity to share it or not. Some people chose to pay one dollar for not having to face expectations, which they knew they would feel like fulfilling. From this, one can hypothesize that sharing behaviour in dictator games is partly motivated by the knowledge that dictators' partners expect the dictators they are paired with to share.

Some further evidence for the role of partners' expectations in giving in dictator games comes from a "natural-field dictator game" (Winking & Mizer, 2013). In this study, the authors ran a dictator game in "natural context"—a bus station in Las Vegas—with subjects not knowing they were taking part to an experiment. The scenario that they implemented⁴ has two consequent differences with standard dictator game: first, the subjects, who played the role of dictators, did not know that they were participating to an experimental game. They consequently did not form beliefs about the experimenter's expectations. Second, and probably more importantly, the dictators believed that the potential receiver did not know that he—the dictator—received an endowment that he could share. They consequently believed that the receiver had no expectations towards them. This is in stark contrast with standard dictator games where the instructions are read aloud and are therefore common knowledge.⁵ The results were striking: none of the 30 sub-

jects transferred anything to the potential recipient. Winking and Mizer conclude that dictators' transfers in standard dictator games are experimental artefacts due to experimenter's demand effects. In fact, Winking and Mizer's experiment reveals more than that: it shows that dictators' transfer occurs only when they believe they are expected to make a transfer—by the experimenter, probably, but mostly and more importantly by the potential receiver.⁶ In an earlier relevant study, Koch and Norman (2008) compared a standard dictator game, in which instructions are common knowledge, with a dictator game where the dictators knew that the recipient did not know a game was played. The number of transfers decreased by half in the latter condition.

Another type of evidence supports the hypothesis that beliefs about partners' expectations drive generous decision-making, specifically, an effect called "hiding behind the small cake" (Ockenfels & Werner, 2012), which is obtained in experimental conditions with asymmetric information. Ockenfels and Werner use a modified dictator game where recipients do not know for sure about the initial endowment of dictators; they know only that dictators' initial endowment can be either one of two amounts. Most importantly, dictators know that recipients' have incomplete information. They are therefore able to infer that recipients can justifiably expect at least a fair share of the small initial endowment, but they cannot expect for sure a fair share of the big initial endowment. Under these conditions, a significant portion of dictators decides to transfer a fair share of the small initial endowment, even when they received the big one. Vranceanu, Sutan and Dubart (2012) obtained a similar effect in a trust game by introducing a very low probability that the money that the truster transfers to the trustee is lost in the process. Trustees receiving the money from trustors know that he or she cannot be sure that the sum has been transferred. Some of the trustees who received the money nonetheless played as if they had not: the rate of transfer back to the trustor was drastically decreased by introducing this small uncertainty. Again, generosity decreased, we think, because the receiver could not be certain about what to expect. A similar effect was obtained by Dana, Weber and Kuang (2007), with a protocol showing that much of the fair behaviour observed in experimental games is most probably due to a dislike for appearing unfair or failing to meet justifiable expectations. Altruism decreases as soon as these expectations lower or are less justified. Also using a trust game, Reuben, Sapienza and Zingales (2009) show that people are quick to behave more selfishly as soon as they believe that little is expected of them. In Reuben et al.'s experiment, trustors' expectations are conveyed to trustees.

⁴The experimenter's confederate gave a set of casino chips to the subjects, who were waiting at a bus station in Las Vegas. The confederate explained that he did not have the time to cash in the chips as he needed to take a plane now, and would have no use of them outside of Las Vegas. He further said "I don't know, you can split it with that guy however you want", while gesturing towards another confederate, the only other person waiting at the same bus station. Importantly, when this happens the confederate is pretending to talk on the phone and does not see the scene of the subject receiving tokens to share.

⁵In Winkler and Mizer's control condition, the instructions are not made common knowledge by reading them aloud in a standard way. Still, the dictator knows that the receiver will be handed an envelope with faked and real tokens and thus be in a position to expect a fair proportion of real tokens with regard to the total number of tokens. The dictator can imagine that sending an envelope with only faked tokens will cause disappointment.

⁶Winking and Mizer recognised the role of the experimenter's expectations but failed to note the role of the receiver's expectations: this is because they did not realize that in standard dictator and in their control condition, but not in their test scenario, dictators believe or know that their paired receiver expects them to share the endowment.

Trustee that face low expectations tend to transfer significantly less than trustee that face high expectations.

The variations on the dictator and trust games reviewed above suggest that choices of dictators and trustees do not reveal distributive preferences but rather aversion to disappointing. Indeed, as soon as the dictator or the trustee is, for one reason or another, not expected by his partner to transfer a given amount, then the dictator does not make the distributive transfer. Removing partners' expectations leads to significant decrease in prosociality. We can also already note that givers, dictators or trustees, tend to consider only the expectations that recipient could justify in all cases.

2.3 Communication and expectations

We now turn to some of the studies that look at the relations between communication and prosocial choices. The question is how much this relation is in fact mediated by the aversion to disappointing. Communication might have an effect on prosociality by conveying information about expectations and about which of others' expectations are deemed justified.

Andreoni and Rao (2011) tested the effect of communication in a simple dictator game. In the condition where the receiver could communicate and justify his expectation, the dictators were much more generous than in the baseline condition (a standard dictator game). This suggests that receivers managed to convey their justified expectations and that this motivated giving. By contrast, in the condition where dictators could explain their transfer, dictators were much less generous than in the baseline condition. This suggests that the opportunity to explain one's choice as a dictator was used to convince their partner that their expectations should be low.⁷ Last, dictators were more generous when they had to play both roles, knowing that only one of the two roles would be randomly taken into consideration for the eventual distribution of money. Presumably, this way to frame decision choices made the expectations of the receiver much more salient, since dictators had to put themselves in the situation of receiver. Communication made the expectations more salient, as did the role-taking condition, but it also enabled justifications. Self-interest affected the choice of justifications: they were usually in favour of equity for those in the receiver's position or in favour of inequity for those in the dictator's position.

In the above situations, the giver prefers the receiver to have low expectations, but there are contexts in which the giver would like the receiver to have high expectations. For instance, when the potential giver would like to be selected

for an economic interaction by the potential receiver. The potential giver might then choose to make promises so as to convince the receiver to enter the economic interaction. The phenomenon of interest for our purpose is that people who promised to make an altruistic choice tend to keep their promise.

Why do people keep their promises even when it is costly for them to do so? Charness and Dufwenberg (2006) note that breaking a promise causes disappointment to the promisee. Aversion to disappointing is therefore likely to motivate keeping one's promise. In their experimental game, they observe that subjects who made promises were much more likely to make the altruistic choice than people who could not make any promise in same experimental context but with no communication allowed. This effect of communication seems to be driven by the knowledge that making promises raise the expectations of the promisee: promising correlated with the belief that the partner has high expectations, which correlated with the likelihood to make an altruistic choice. Furthermore, the difference between the communication and the no communication condition was driven by the change in beliefs about expectations.

Indeed, for similar beliefs about expectations, the rate of altruistic choice is the same in the communication and the non-communication condition. Vanberg (2008) ran a variant of this experiment in which the subjects had a 0.5 probability of interacting with a new partner rather than with the promisee. Thus some subjects made promises to someone but found themselves interacting with someone else. Still, they knew that the new partner had received a promise and that he did not know whether partners had been changed or not. The new person was therefore likely to have the same expectation as the partner to whom they had initially promised. Vanberg found that subjects who made a promise but had to interact with a partner to whom they had not promised were much less likely to make an altruistic choice, keeping their promise, than the subjects who interacted with the subject to whom they had made the promise. This difference occurred in spite of the fact that beliefs about partner's expectations was the same whether the partner had changed or not. From this result, Vanberg concludes that aversion to disappointing is not what motivates people to keep their promises. In fact, Vanberg's results show that people are not especially motivated to fulfil any type of expectations. If A expects B to do X, then C is not especially motivated to do X. Vanberg's experimental conditions do just that: B makes a promise to do X, A consequently expects B to do X, but when C has the opportunity to do X, he does not especially feel compelled to do it. Vanberg's results resemble bystander effects, where a person in need expects others to help but nobody feels that she herself is the one who is expected to help. Vanberg's experiment is therefore a new instance showing that altruistic choice significantly decrease as soon as the expectation of altruism is not well justified.

⁷In the appendix of their paper, Andreoni and Rao (2011) list the type of explanations that people gave. One standard explanation was related to the fact that allocators had the right to benefit from their privileged position as decision maker. If dictators truly had this right, then receivers could not justify high expectations.

2.4 Difference between expecting someone to do something and making a guess about what people do

The above experiments suggest that communication modulate beliefs about expectations, which in turn influence the motivation to make altruistic choices. They also suggest that the aversion to disappointing cannot be expressed as a blunt motivation to fulfil others' predictions whatever these predictions are. The content and valuation of expectations matter: Am I expected to do something? Does the person hope that I will do it? How justified is the expectation?

The last experiment we want to mention was run by Ellingsen, Johannesson, Tjøtta and Torsvik (2010) with the explicit goal of testing guilt aversion. Ellingsen et al. ran a set of experiments where dictators in a dictator game and trustee in a trust game were informed of what the receivers predicted to receive. Dictators and trustees were informed by the "guess" of the subject they were associated with, and decided of their transfer in view of this guess. The rationale is that if dictators and trustee are willing to conform to the receivers' predictions, then there should be a correlation between the transfer and the prediction. But Ellingsen et al. found no such correlations. These results, it must be said, do not refute the theory of guilt aversion: the authors report non-significant results; they did not succeed in finding a statistically significant demonstration of guilt aversion. However, they raise again the need to specify which expectations we are willing to fulfil and which we are happy to ignore. Why did subjects of Ellingsen et al.'s paper seem to ignore others' predictions?

Ellingsen et al. did not differentiate between predictions and expectations. In Ellingsen et al.'s variation on the dictator game, dictators could decide their transfers in view of what receivers had "guessed" the average transfer would be. Thus, dictators were not asked whether they wanted to fulfil receivers' expectations, but only whether they wanted to behave in accordance with what receivers predicted about the average behaviour. Two elements seem to be missing for tapping into the aversion to disappointing: first, there is an element of hope in the expectations that people prefer to fulfil. I can guess that you will wake up before eight this morning but not really care. In that case, my prediction will not motivate you to wake up before eight. Second, someone's expectation motivate someone else's choice especially when it is about the choice of this very person. You are motivated to do X, if I expect you to do X, but less so if I only expect X to happen. Subjects in Ellingsen et al.'s experiment, however, were given predictions of the average population rather than expectations about their own behaviour. Ellingsen et al. (2010) and Vanberg (2008) have used similar manipulations: subjects were given expectations that were not about their own behaviour but about others'. Thus, we argue, expectations about B's choice do not motivate A to make the

expected choice.

The experiment that we present here is a close variation on Ellingsen et al.'s dictator game. However, we have explained why others' expectations in Ellingsen et al.'s experiment did not evoke aversion to disappointing. So when we describe our own protocol, in the next section, we will specify how it differs from Ellingsen et al.'s one.

2.5 A cognitive account of expectations and their motivating effect

The experimental results reviewed above support the hypothesis that people's altruistic choices in experimental games are largely due to beliefs about their partner's expectations and preferences for fulfilling these expectations.

Partners' expectations can easily be computed: Our capacity to compute what others believe and desire provides a firm cognitive ground on which preferences can operate: we can perceive others' expectations and predict the effect of our own behaviour on others' mental states, including being disappointed. While the processes of ascribing mental states is being thoroughly studied in current cognitive psychology,⁸ we here hypothesize about how people come to value others' mental states, specifically, how people perceive and value others' disappointment. Indeed, the above-mentioned experiments show that people do not blindly want to fulfil others' expectations. We can in particular note the following documented effects:

1. Effect of low expectations: if the "giver" is led to think that the "receiver" expects little or nothing, then he tends to give as little as is expected from him (Broberg et al., 2007; Dana et al., 2006; Winking & Mizer, 2013)
2. Effect of justificatory ground: expectations that could not be justified by the expecting partner tend to be disregarded (Dana et al., 2007; Ockenfels & Werner, 2012; Reuben et al., 2009; Vranceanu et al., 2012); expectations that can be justified by appealing to social norms tend to be fulfilled (Cronk & Wasielewski, 2008; Heintz, 2013; Henrich et al., 2004).
3. Effect of salience: the more salient the expectation and its justificatory grounds, the more likely the expectation will be fulfilled (Andreoni & Rao, 2011; G Charness & Dufwenberg, 2006; Rankin, 2006). In particular, expectations must clearly be about a specific individual's actions and choice in order to motivate this individual to fulfil the expectation (Ellingsen et al., 2010; Vanberg, 2008).

⁸It is shown that humans, are able to ascribe beliefs and intentions to others with great ease (e.g. Frith & Frith, 2012; Leslie, Friedman & German, 2004) and they seem to compute others' thoughts automatically from an early age (e.g. Kovács, Téglás & Endress, 2010; Samson, Apperly, Braithwaite, Andrews & Bodley Scott, 2010).

It follows from our discussion of guilt aversion that people not only predict what their partner will do but also associate a valence to their predictions. People might be indifferent about someone's choice, or they might desire that he does something, or they might fear it. Miceli and Castelfranchi (2002) identify different configurations of predictions and goals, which produce various kinds of representations of the future. In their taxonomy, which includes levels of uncertainty, personal concerns and need for prediction, people can believe that something will happen and at the same time want to see that event realized.⁹ Miceli and Castelfranchi coined the term "hope-casts" in order to refer to those mental states that imply both forecasts and a positive valuation of the predicted event—the term is a portmanteau word for "hopeful forecast". By contrast, "fear-casts" refer to the mental states that imply both forecast and the desire that the forecasted event does not happen. Are people intrinsically motivated to comply with others' fear-casts? We hope-cast it is not the case. For instance, your dancing partner might fear-cast that you will step on her feet, but knowing this probably does not motivate you to do it. Predictions that include no personal concern will also not motivate choice. For instance, my partner can predict that I will wear a blue shirt today because she wrongly thinks it is the only clean one I have; this will not necessarily motivate me to do so . . . unless she hopes that I do so for some reason. She might expect me to wear this colour because we go to some fancy ball and she'd like my own shirt to fit the colour of her shoes.

In what follows, we use the term "expectation" as meaning hope-cast. This more specific meaning of "expectation" is similar to a common usage of expectation: to "expect something" often means "to hope for it", and to expect someone to do something often means that we want him to do it. This usage is not the only one: for instance, in most scientific papers, expectations and predictions are used interchangeably. With the distinction between expectations as hope-casts and prediction in mind, we can point out the effects of predictions, such as self-fulfilments, which are mainly obtained because predictions, when made, publicly convey information; self-fulfilments are not obtained because one is motivated to make them true. Expectations as hope-casts are special cases: those who are expected to do something can be motivated to make these expectations true because they do not want to disappoint the expecting person.

The literature also suggests that people evaluate the justification of others' expectations. For instance, receiving a promise provide firm grounds for justifying expectations, but if the promise was made by someone else, as in Vanberg's (2008) experiment, then the partner can always argue

that he did not make any promise to the expecting partner. Likewise when the endowment to be shared is uncertain, the "giver" could always argue back against a request for a faire share of the high endowment: "hiding behind the small cake" is possible because the expecting partner could not justify high expectation. Misyak et al. (2014) argue that in many social interactions, "virtual bargaining" is taking place: people commonly know what would be the result of explicit bargain and make their choice as if the explicit bargain had taken place. In our case, we do not need to go as far as saying that people make a virtual and implicit bargain, but it seems that they do assess the reasons that their partner might have to justify their expectations. When fulfilling an expectation is costly, it is indeed an adaptive thing to do. So our hypothesis is that the aversion to disappointing includes mechanisms assessing whether the partner is expecting more than what already makes one be a valuable cooperator. When it is the case, then we are better off disregarding the expectations. Figure 1 is a simplified picture of the hypothesized process that ascribe utility to fulfilling others' expectations.

In the experiment we report, we provided subjects with information meant to influence their beliefs about what their partner hoped they would do. We predicted that subjects in the role of dictators would be sensitive to expectations and make their transfer dependent on expectations. More precisely, we predicted that people would tend to fulfil expectations as long as these are justifiable but would otherwise disregard them. We presented low and high expectations with the idea that only low expectations would be thought of as justified expectations.

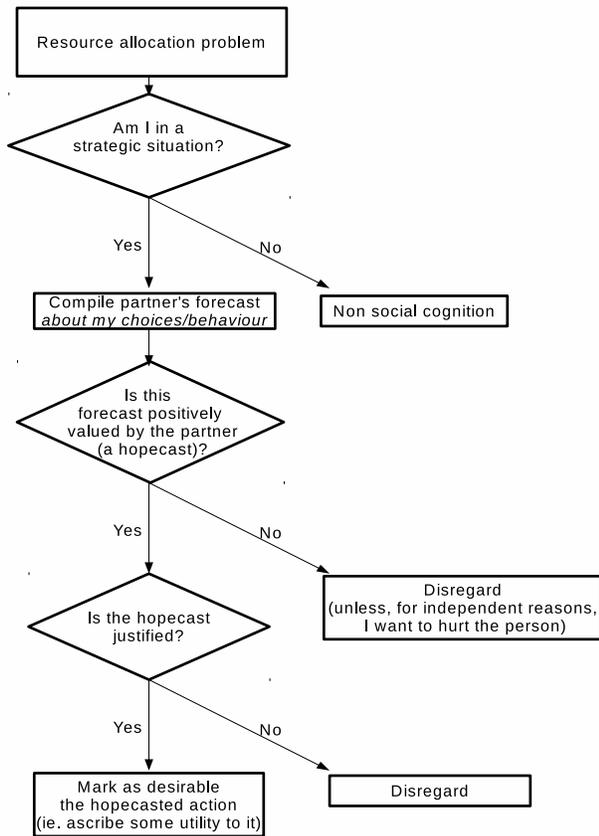
3 Experimental design

The experiment is a simple variation of a dictator game, with three conditions. In the test condition—the expectation condition (EC)—we asked dictators how much they wanted to transfer of their initial endowment to the subject they were paired with, but we also provided dictators with the opportunity to decide how much to transfer in view of their paired subject's expectation. The baseline condition (BSL) was a standard dictator game, and a control condition provided dictator with the opportunity to decide how much to transfer in view of irrelevant information (Irrelevant Information Condition, IIC).

Subjects were randomly assigned to the role of "dictator" or "receiver" (labelled A and B during the experiment). Each dictator was given an endowment of 10 Euros and the opportunity to transfer a share of this endowment to the receiver. When subjects entered the laboratory, they were invited to seat in private boxes in which they received printed instructions of the game. Subjects were told to read these instructions privately. After ensuring that all subjects had read

⁹The existence of this motivational component and its effect on decision making is illustrated by Giardini et al. (2008) who show that people tend to be more confident in the occurrence of favourable events, with little or no regard for their objective likelihood.

Figure 1: Processes ascribing utility to fulfilling expectations



the instructions once, a monitor read the instructions aloud so as to guarantee common knowledge.¹⁰ Subjects received the decision sheet for their role and indicated their decisions on that sheet. Once all subjects filled in their sheets, the experimenter collected them so as to compute subjects' final payoffs.

The experimental sessions were conducted at the Burgundy School of Business in 2011. They were conducted in French (BSL and EC) and English (EC and IIC). All sessions were implemented using pen and paper. On aggregate, 284 subjects were recruited from introductory courses: 100 subjects (i.e. 50 dictators) participated in the Baseline (BSL), 94 subjects (i.e. 47 dictators) in the Expectation Condition (EC) and 90 subjects (i.e. 45 dictators) in the Irrelevant Information Condition (IIC). EC and BSL were run first and IIC was run some six months later with a new pool of students. Subjects' average age was 20.8 (S.D.= 1.206).

¹⁰We did not check subjects' understanding of the instructions but subjects were all informed that they could ask privately questions of the monitor by raising their hands at any moment of the experiment.

3.1 Framing

Ellingsen et al. (2010) ran an experiment that also allowed dictators to decide how much to transfer in view of the recipients predictions. But the team obtained no results. We therefore emphasize how our protocol differs from Ellingsen et al. and why.

As we argued, a hopeful expectation contains a predictive element but is not reducible to it. We wanted dictators to understand that receivers expressed their hopes. Contrary to Ellingsen et al., we did not incentivise receivers to state accurate predictions and we framed the questions with the term “expect” rather than “guess”.¹¹ Although the structure of the game remains the same, the different framing is likely to be consequential¹² due to common sense understanding of “expect”, which conveys a sense of hopeful prediction. With this change of frame, subjects make inferences about others' mental states. Using the term “expectation” rather than “guess” is probably a better way to tap into the aversion to disappointing and thus observe its effects on choices.

In our own experiment, dictators had to decide of their transfers answering a question starting with: “If the player B I am paired with expects to receive ...”. This sentence makes it clear to a given dictator that his paired receiver has expectations about him. We therefore used a frame that, we believe, tapped into the aversion to disappointing because it conveyed that the partner was hoping for a given transfer from his paired subject.

3.2 The strategy method

In order to investigate the effect of all potential expectations, we used the Strategy Method: in the Expectation Condition (EC) condition we asked dictators to indicate how much they were willing to send for each of the 11 (from €0 to €10) possible expectations of the receivers. Dictators were asked to answer the following question, repeated for each of the 11 possible amounts with X taking the values from 0, 1, 2, 3 till 10.

If the player you are associated with expects you to send him € X, you will send €__.¹³

Final payoffs were calculated by randomly forming pairs and allocating to the receiver of each pair what her paired dictator wanted to transfer for the expectation that the receiver actually reported.¹⁴ In this way, we were able to

¹¹In French, we used the expression “s’attendre à” and “souhaiter”.

¹²The dictator game is sensitive to changes of frame: for instance, framing the dictator game initiating a sense of entitlements change subjects' choices (e.g. List, 2007), as does referring to social norms (Lesorogol, 2007). But see Dreber et al., (2013) for a sceptical view.

¹³In the French version: “Si le participant B qui compose ma paire s’attend à recevoir €X, je lui envoie €__.”

¹⁴Of course, dictators had no direct access to recipients' expectation. They could form their beliefs only in view of the answer given by the re-

gather data about dictators' choices for all types of expectations, including those that receiver would not usually form (for instance, very few receivers expected to receive the full amount or nothing). By using the Strategy Method in EC, we were able to observe whether and to what extent receivers' expectations modulate dictators' decisions.

The strategy method does not generally change the results obtained by a direct method (e.g., Brandts & Charness, 2011). There are, however, good reasons to run a control checking the effects that are due to the strategy method. Several factors other than sensitivity to expectations can lead subjects to vary their answers when using a strategy method. One is the way information is displayed: the mere presence of numbers in relation to allocations might work as a reference point, influencing dictators' choice due to some version of an anchoring effect. Another confound is the "experimenter effect": with the strategy method, the subject might be lead to think that, since the experimenter asks several questions, he must expects to obtain different answers. In order to control for such external factors, we implemented a condition—the irrelevant information condition (IIC)—where dictators had the opportunity to vary their transfer in function of some irrelevant numbers. The irrelevant figures corresponded to the last number of the receiver's student card. More precisely, in the Irrelevant Information Condition (IIC), dictators were asked to state their allocation conditional on their partners' ID last digit number. Receivers had to indicate their ID last digit instead of indicating how much they expect dictators to send them, as subject did in the Expectation Condition. In the meanwhile, each dictator had to indicate how much she was willing to send for each of the 10 (from 0 to 9) possible final digits of the receiver they were matched with. Dictators were asked the following question, repeated for each of the 10 possible numbers:

If the player you are associated with has an ID card ending with X, you will send €__.

The final allocation was computed by matching the ID's digit indicated by the receiver with the dictator's allocation for that specific number. By comparing dictators' decisions in EC and in IIC, we were able to measure the impact of others' expectations in dictators' transfers. (See the supplementary material for the instruction and the decision sheets).

ipient. Dictators could interpret the recipient's answer as strategic rather than a genuine report about their expectations. One option would have been to hide from the recipient that their answer will be known to dictators and let dictators know that. However, this would have meant somewhat deceiving the recipients and adding a level of cognitive complexity for the dictators, so we did not implement this option.

4 Results

4.1 Replications of previous results together with different pattern of choices between low and high expectations

In BSL, the average amount sent by dictators was €3.84 (S. D = 7.892). In line with previous Dictator Games, we found in BSL that the huge majority of dictators decided to transfer positive amounts to receivers. Only 20% of dictators chose to send nothing to receivers whereas 36% of dictators shared their endowment equally with receivers. In the expectation condition (EC), dictators sent an average amount of €2.55 (Std. Dev = 7.441) to receivers.

We analysed the correlation coefficients between transfer and expectations. According to the model of guilt aversion, dictators' transfers should have correlated with partners' expectations: the higher the expectation, the higher the transfer should have been. This is because the higher the expectation is, the more one needs to transfer in order to reduce the guilt induced by the difference between partner's expectation and transfer. The correlation coefficient need not be one, since selfish motives and not just guilt avoidance determine choice; but it should be positive.

The correlation coefficient between transfer and expectation is not significantly different from the null hypothesis of a zero coefficient. Ellingsten and colleagues also report a correlation coefficient that is not significant (Pearson correlation coefficient of -0.075 , $p = 0.497$; and Spearman correlation coefficient of -0.044 , $p = 0.689$). We obtained a Pearson correlation of 0.166 with a much "better" but still not significant p-value: $p = 0.058$. The non-parametric Spearman coefficient in our experiment is 0.000 . A look at Table 1 suggests that there is no correlation because, when the expectation is higher than 5, subjects tended to give less and less as expectation increases.

Contrary to the model of guilt aversion, we did not predict correlations on the zero to 10 range: our hypothesis is that people will fulfil expectations only if these are justifiable, and thus, in the dictator condition, less to half of the initial endowment. We can confirm, however, that the prediction of the guilt aversion model that there would be a strictly positive correlation between transfer and expectations, independently of the content of the expectations, is not supported.

In order to check which of the expectations affected transfer, and more precisely, which expectations people preferred to fulfil and which they disregarded, we analysed the responses across subjects to any given expectations. In particular, for any given expectation, the analysis reveals how many subjects decide to fulfil the expectation. Table 1 shows the proportion of dictators that chose a given transfer (y-axis) for a given expectation (x-axis). The most frequent transfer, across expectations, is zero, followed by 5: the "rational selfish" choice and the "fair" choice. Sensitivity to

Table 1: Proportion of dictators in each column choosing the transfer for the expectation specified on the horizontal axis.

	10	4	2	2	2	2	2	4	2	2	6	8
	9	0	0	0	0	0	0	0	0	6	2	0
	8	0	2	2	0	0	0	2	4	2	0	4
	7	0	0	0	2	2	4	6	6	0	0	0
	6	2	0	0	2	6	4	4	2	6	4	4
Transfers	5	14	10	12	21	14	38	14	17	17	19	14
	4	0	2	6	6	17	8	12	6	6	4	4
	3	4	12	12	23	4	0	4	12	6	2	0
	2	10	8	21	8	12	6	6	6	12	6	4
	1	12	25	8	2	6	6	8	8	6	14	12
	0	51	36	34	31	34	29	36	34	34	40	46
		0	1	2	3	4	5	6	7	8	9	10
		Expectations										

expectations is visible in the following patterns:

- The frequency of transferring nothing decreases up to five then increase anew. This pattern is inverted for the frequency of transferring 5: increasing till five then decreasing. This suggests that people were more likely to fulfil expectations when those expectations appeared reasonable or fair enough. Otherwise, they are more likely to transfer nothing.
- There is a smaller but significant amount of choices that equal expectations (line $y=x$) until expectations reach five, then a small but significant amount of choices that, for expectations from 6 to 10, are proportionally inverse to expectations (equal to the initial endowment minus the expectation).

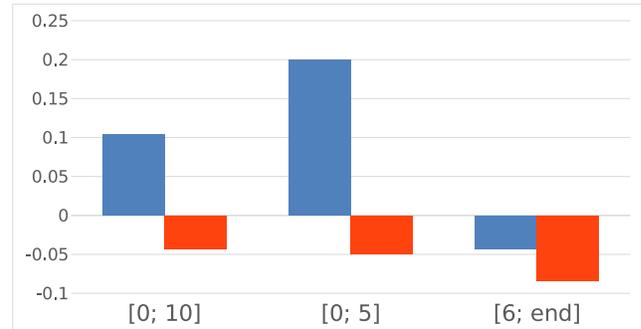
For instance, when the receiver expects nothing, 51% of the dictators choose to transfer nothing, but this number drops to 29% when the receiver expects half of the endowment. When facing an expectation of half of the endowment, 38% choose to transfer it.

4.2 People are willing to fulfil low but not high expectations

The transfers differ a lot for low expectations vs. low irrelevant numbers, but differ little for high expectations vs. high irrelevant numbers. We therefore examined transfers for expectations from 0 to 5, which we compared with transfers for irrelevant information from 0 to 5, and transfers for expectations from 5 to 10. We note these conditions respectively EC[0,5], IIC[0,5] and EC[6,10].

We used Wilcoxon tests (within-subjects correlations) to compare our results in The correlation coefficients between subjects' transfers and expectations on EC[0,5] to the null

Figure 2: Average correlation between transfers and expectations (EC) and transfers and irrelevant information (IIC), comparing different ranges of expectations/irrelevant information.



hypothesis according to which the correlation is zero and to the IIC0, 5 condition (Figure 3). The average of the is 0.200 (correlation between the vector of all transfers and the vector of expectations, $p = 0.0007$).¹⁵ We also compare the distributions of subjects' correlation coefficients between expectation and transfer on factors, attributing correlation factor 0 to subjects that show no variance: EC[0,5] is 0.200. The distribution of correlation coefficient is significantly different from zero ($p = 0.0001$, two-tailed).¹⁶ Furthermore, it is significantly different from IIC[0,5], ($p = 0.0030, 0.002$, two-tailed). The¹⁶ IIC[0,5] does not have a distribution of coefficient correlations that is significantly different from the null hypothesis. Our hypothesis that others' expectations are motivating as long as they are not "unreasonable" is therefore supported.

When the expectations are above 5, by contrast, we find

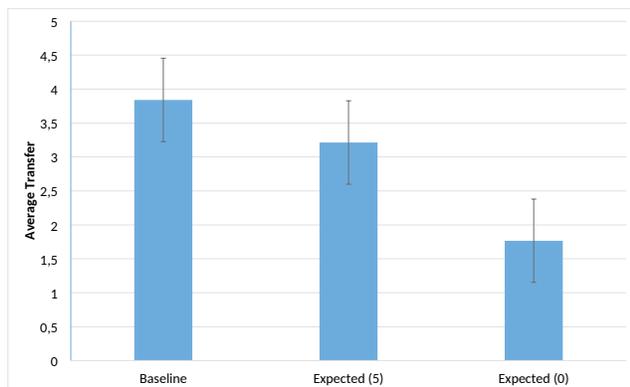
¹⁵We observe a significant correlation also if we run a multilevel regression analysis with transfer vector as dependent variable and expectancy vector as predictor, allowing for a random intercept by participant. This point and several others in this subsection, were made by Mark Jeckel, who provided highly valuable guidance for the statistical analysis.

This significant correlation on [0;5] is worth comparing with Ellingsen et al.'s non-significant correlation. Indeed, Ellingsen et al.'s data points for the correlation analysis are mostly in the range of "reasonable" expectations: from zero to half of the endowment. This is because Ellingsen did not use a strategy method. Consequently, he could gather data only on the transfers made for expectations that had been actually expressed. As we hypothesized, we obtain significant correlation by using the right framing.

¹⁶It is this significant correlation on 0;5 that is worth comparing with Ellingsen et al.'s non-significant correlation. Indeed, Ellingsen et al.'s data points for the correlation analysis are mostly in the range of "reasonable" expectations: from zero to half of the endowment. This is because Ellingsen did not use a strategy method. Consequently, he could gather data only on the transfers made for expectations that had been actually expressed. By contrast, we gathered data also for extreme possible expectations that are above half of the endowment and which are rarely or never expressed.

¹⁶We obtain similar results if we perform the analysis by removing the subjects with no variance rather than ascribing them a zero coefficient. This holds here and for the analysis of EC[6,10] vs. IIC[6,9] reported in the next paragraph.

Figure 3: Average transfer for expectation 0, expectation 5 and baseline



no significant difference of correlation coefficients between EC[6,10] and the null hypothesis ($p = 0.489$, two-tailed) and 0.5091). We observe no significant difference between the distributions of correlation factors of EC[6,10] and IIC[6,9] ($p = 0.3180, 3285$; attributing a correlation 0 to subjects that show no variance). This therefore suggests that expectations are taken into consideration when corresponding to less than half of the endowment, but that they are treated as irrelevant, on the average, when exceeding half of the endowment.

Unlike Ellingsen et al.'s (2010) analysis, our results show that people are sensitive to others' expectations. In favour of Ellingsen et al., however, it turns out that people do not willingly fulfil all type of expectations. In our experiment, it is only low expectations that are, often enough, being fulfilled.

The controversial claim that generosity observed in dictator game is mainly due to aversion to disappointing rather than other prosocial preferences such as inequity aversion is further corroborated by the fact that the average transfer for the most common expectation, the expectation of 5 (31.81% of receivers expect to receive half of the endowment) in the EC condition is a close approximation of the transfer in BSL: €3.21 for the EC condition and €3.84 for BSL.¹⁷ This similarity can be contrasted to the significant difference of transfer between when the expectation is zero and when the expectation is 5 (EC[5] vs. BSL: two-sided Mann-Whitney, $p = 0.361$ and EC[0] vs. EC[5]: $p = 0.006$. The confidence intervals for BSL and Expectation 5 are respectively: [3.041; 4.638] and [2.445; 3.979] showing no significant differences across the two variables. see Figure See figure 3).

¹⁷ This average transfer is slightly bigger than the average transfer found by Engel's meta study: 28.35% (Engel, 2011).

4.3 Types of reaction to receivers' expectations

Thanks to the Strategy Method we used in EC, we defined behavioural patterns by observing vectors of decisions from dictators, i.e., by grouping dictators in accordance with their decisions for ranges of possible expectations. We identified three patterns: Constant, Positive Fulfiller, Negative Fulfiller.

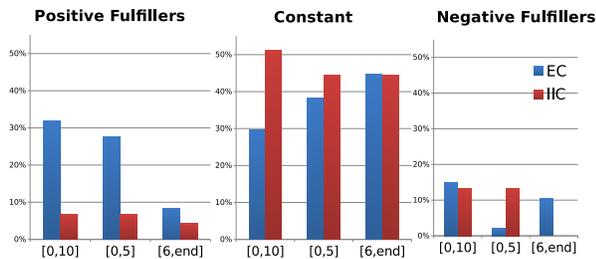
To fall into the category Constant, dictators had to send the very same amount of money for at least 80% of her choices. A constant player is therefore someone who decides on what to transfer disregarding the information provided—whether it is about expectations, in the EC, or irrelevant, in IIC. Still, we allowed for some errors (20%) because the strategy method does provide a frame that facilitates making these kind of little motivated changes. A dictator exhibits a Constant behaviour if she makes the very same transfer for at least 9 of the 11 possible expectations of the receivers on EC[0,10]. On EC[0,5] a behavioural pattern is Constant if the dictator makes the very same transfer for at least five of her six possible choices, and on EC[6,10] the pattern is Constant if she makes the same transfer for at least 4 of her 5 possible choices. The same applies for IIC[0,9], IIC[0,5] and IIC[6,9] relative to the irrelevant information 0 to 9, 0 to 5 and 6 to 9. On EC[0,10], 29.79% of the dictators fell into the category Constant. In IIC, more than half of the subjects (51.11%) sent a Constant amount. There are significantly more dictators belonging to the Constant category in IIC[0,9] than in EC[0,10] ($p = 0.0377$, $\chi^2 = 4.348$). This suggests, again, that people do not consider others' expectations as irrelevant information.

Proportion of subjects for behavioural patterns in EC and IIC:

We consider a dictator to be sensitive to others' expectations if her decisions were significantly correlated (either positively or negatively) with the receivers' expectations, using Spearman Rank Correlation coefficients between the decisions of dictators and the expectations of receivers. In EC[0,10], 46.81% of subjects exhibited sensitivity to others' expectations. Thus, the decisions of almost one dictator out of two were significantly correlated (either positively or negatively) with receivers' expectations. This is significantly more than in ICC[0,9], in which only 20% of the dictators show sensitivity to irrelevant numeric information ($p = 0.01$, $\chi^2 = 6.556$).

We refine the category of sensitive players by distinguishing between Positive Fulfiller and Negative Fulfiller. An individual is considered as a Positive Fulfiller if her decisions are significantly (at a 0.05 level) and positively correlated with receivers' expectations. Positive Fulfillers transfer more and more as receivers' expectations increase. Conversely, an agent is classified as a Negative Fulfiller if her decisions are significantly (at a 0.05 level) and negatively

Figure 4: [Proportion of subjects for behavioural patterns in EC and IIC.](#)



correlated with receivers' expectations. Negative Fulfillers transfer less and less as receivers' expectations increase. They reward low expectations (inferior to half their endowment) and punish receivers with high expectations by decreasing their transfers.

In EC[0,10], 31.91% of agents belonged to the category Positive Fulfiller, while in IIC[0,10] only 6.67% belonged to that category. Again, we find a significant difference between the two conditions ($\chi^2 = 9.312$, $p = 0.002$). Unsurprisingly, this significant difference is found on EC[0, 5] vs. IIC[0, 5] ($\chi^2 = 7.052$, $p = 0.007$) but not on EC[6, 10] vs. IIC[6, 10] ($\chi^2 = 0.623$, $p = 0.429$).

In EC [0,10] 14.89% of agents were classified as Negative Fulfiller, and 13.33% in IIC[0,10]: there are no significant difference when we examine the full range from 0 to 10 ($\chi^2 = 0.046$, $p = 0.829$) but when we consider the ranges from 0 to 5 and from 6 to 10, it turns out that there are fewer Negative Fulfiller in EC[0, 5] than in IIC[0, 5] ($\chi^2 = 4.601$, $p = 0.042$) but more in EC[6, 10] than in IIC[6, 10] ($\chi^2 = 5.062$, $p = 0.024$). We interpret this finding as expressing the fact that dictators withdraw their concern when interacting with people with too high expectations. People with unreasonable expectations are not worth satisfying. Alternatively, people might disapprove those who communicated having unreasonable expectations, taking it as an attempt to exploit generous dispositions. For any of these two reasons, dictators withdraw their concern for their partner and act more selfishly. Figure 4 sums up these results.

5 Discussion

The experiment provides further evidence that a mind-directed preference, the aversion to disappointing, better predicts when people act generously in dictator games than do distributive preferences alone (inequity aversion or preference for fairness) because distributive preferences cannot predict the dependence of transfers on expectations. Our simple experiment shows that there often is a dependence between transfer and expectations within each subject.

While previous studies show that transfers vary with oth-

ers' expectations, they do not allow a comprehensive investigation of the effect of various expectations. The more comprehensive investigation that we did is revealing: not all of others' expectations turn out to be worth fulfilling. In the discussion below, we ask what model of preferences can account for our findings.

5.1 The model of guilt aversion and its psychological implementation

Sensitivity to others' expectations is definitively a point in favour of the model of guilt aversion. In complementing Ellingsen et al.'s (2010) analysis and making some variations on their initial experiment, we have been able to show, against their claim, that expectations do play a role on altruistic choices. We obtained behaviours in the expectation condition different from behaviours in the irrelevant information condition because we framed the questions so as to convey that the receiver hoped that her partner would behave in a specified way. Ellingsen et al. did not refute the model of guilt aversion but helped us to clarify how to tap into the aversion to disappointing. To be fair, it should be acknowledged that the disutility calculated by Battigalli and Dufwenberg's (2007) model of guilt aversion does express more than a blunt willingness to conform to others' predictions. Disutility of i 's guilt for disappointing j is calculated as follows:

1. Measure j 's expected payoff for her strategy, given her beliefs about what others are likely to do.
2. Measure j 's disappointment as equal to the difference between the above measured expected payoff and her actual payoff.
3. Measure the extent to which i is responsible for disappointing j : this is done by calculating the smallest disappointment that i could generate for j , given i 's possible strategies. This number represent the amount of j 's disappointment that i could not reduce by any strategy. Then, in order to obtain how much disappointment is due to i , subtract this number to the "total" disappointment calculated above.
4. The total disutility for i brought by aversion to disappointing is the sum of disappointments that others have due to his choice and multiplied by a factor reflecting how much disappointing is disliked.

Thus, the disutility of guilt is calculated as being in proportion to how much less an agent's obtained utility is, compared to what she had predicted (her calculation of expected utility). This aspect of the model adequately captures the element of hope in expectations, upon which we insisted and which Ellingsen et al. had ignored. For instance, you will not feel guilty if you behave more generously than what your

partner predicted: you have not behaved according to his prediction, but you have not decreased his utility compared to his predictions. More generally, any unpredicted action that does not decrease others' utility is unlikely to generate guilt. A further element that we insisted on is that others' expectations are motivating only to the extent that they are directed at oneself. This aspect did not figure in Ellingsen et al.'s test of guilt aversion but it figures in some form in the point (3). Still, the dependence of transfer on expectations that we observe is not exactly what models of guilt aversion would predict. We observe that it is necessary to distinguish low from high expectations: only the former are being fulfilled. Current models of guilt aversion fail to predict which hopecasts are disregarded (because unjustified) and it does not seem to us possible to derive the distinction by tinkering with the variables of the model. In the model, disutility from disappointing is a strictly increasing function of others' expectations.¹⁸ Applied to our experiment, this means that the higher the expectations of recipients, the higher the transfer of dictators predicted by the model. But we observe that expectations above the cut-off point are just not worth considering. They even motivate non altruistic choices, like teaching the person who has the too high expectations that his expectations are too high. In our experiment, this was done by systematically lowering the transfer as the expectations grew. Expectations need to be sufficiently low or else partners will not be motivated to fulfil them. This result replicates the "hiding behind the small cake" phenomenon and some other experimental results mentioned in section 2, where subjects decrease altruistic transfer as soon as high expectations lack justificatory grounds. It thus confirms that this decrease in generosity results from the subject selecting, with some guile, the possible expectations that are the least costly to fulfil. How do people distinguish between the expectations that they prefer to fulfil and those that they can disregard? Models of guilt aversion do not say. We look at several options in the next subsection.

5.2 Justifiable expectations

The first and most conservative hypothesis is that the cut-off point between expectations being worth fulfilling and expectations that can be determined is determined by a shared sense of fairness (Baumard, André & Sperber, 2013; Binmore, 1998, 2005). This hypothesis is compatible with our data, since the cut-off point we observe is half of the initial endowment, which most people think of as the fair way of sharing. Theories of fairness then need to be adjusted as

¹⁸For instance, the disutility function of guilt could be concave so that it becomes, at some level, not worth avoiding guilt because it would require a too high cost. There are cases where we choose a course of action because it provides immediate material benefits even though we know we'll feel guilty: not working as hard as your colleagues expect you to, for instance. But this type of tinkering will not predict our experimental results: there is a cut-off point where expectations are just disregarded.

follows: fair choice is not, or is only partially, motivated by an intrinsic preference for fair distributions. This accounts has the benefit of explaining why it is that altruistic choice is enhanced when expectations are more salient. There is a modest form of this hypothesis according to which there is a motivation to act fairly that is supplemented with an aversion to disappointing, and a strong one according to which the desire to act fairly derives from a primary (evolved) aversion to disappointing. The latter hypothesis has the advantage of reducing the number of primary social preferences and thus offloading evolutionary psychology from the task of accounting for yet another pro-social preference—which always raise difficult issues.¹⁹ But it raises further questions: why would people sometimes act generously even in the absence of expectations? A number of potential explanations come to mind: it is the mental representations of others' expectations that have a causal role, and these can be produced even if others' do not actually have expectations. They can be produced via perspective taking and counterfactual thinking ("If I were in her situation, this is what I would expect").

A more "eliminatory" hypothesis is that the cut-off point at which expectations are disregarded is an equilibrium point where willingness to fulfil expectations and actual partners' expectations tend to meet. The difference between the above hypothesis and the last one is that the equilibrium need not be fixed by an evolved sense of fairness. It results from learning from day-to-day interactions, from perspective taking and counterfactual thinking, and from understanding the coordinating role of cultural norms.²⁰ The implicit negotiations and virtual bargaining that this would request has recently been hypothesized to actually occur (Misyak et al., 2014). Likewise, DeScioli and colleagues have advocated a strategic model of moral judgement, according to which fairness judgements are very susceptible to self-interest yet bounded by possible justifications (DeScioli, Massenkoff, Shaw, Petersen & Kurzban, 2014).

¹⁹Preferences are unlikely to evolve unless they increase fitness in some way or another, which is far from obvious for social preferences that motivate altruistic choices (West, El Mouden & Gardner, 2010).

²⁰In particular, it can be constrained by aspects of a market where potential cooperators have outside options. An individual with outside option would be disappointed to get less from an interaction than they would have gotten from the outside option—not interacting. For Binmore, Baumard and others these constraints have had an effect during phylogeny so that humans are endowed with a genetically canalysed sense of fairness. With this third hypothesis, we suggest that these aspects of the social environment have a causal effect during phylogeny but only for the evolution of a universal aversion to disappointing. What can be justifiably expected in the specific contexts of interactions is, however, learned. We thus acquire a sense of fairness from a learning process that is backed up by an evolved aversion to disappointing.

5.3 Aversion to disappointing and the social esteem model

Another important question that arises is whether the aversion to disappointing is any different from a preference for maintaining a good image in the eyes of others (Ariely, Bracha & Meier, 2009; Bénabou & Tirole, 2006, 2009). It might be that people avoid disappointing not because they are averse to it but because it is a cheap means for maintaining a good image in the eyes of others. People would be willing to fulfil expectations so that others know they do, and thus consider them as good cooperators.

A social esteem model, however, would need some auxiliary hypotheses to predict the aversion to disappointing that we observed. In particular, we see that the aversion to disappointing is expressed in an anonymous context, which does not directly fit the function of a social esteem model. There are plausible ad hoc hypotheses answering this challenge: for instance, it might be that anonymity in experimental contexts is understood at the reflexive level but not fully processed as such. The social esteem model could also be extended so as to include self esteem: in anonymous context people are generous because they want to think of themselves as good cooperators; after all, their actions is not kept anonymous to themselves.²¹ The reduction, in any case, is not straightforward and would need to deal with a few difficulties: for instance, the self-esteem model still cannot directly account for hiding-behind-the-small-cake effect. And there are many behaviour that would be explained by a social esteem model but not an aversion to disappointing—conspicuous consumption, for instance. At the experimental level, the fact that prosociality increases with the feeling of being observed (e.g., Andreoni & Bernheim, 2009; Haley & Fessler, 2005) is clearly related to a preference for signalling but not an aversion to disappointing. At this point, we think it is best to consider the aversion to disappointing as a preference on its own that would have evolved in view of maintaining a good reputation at the lowest possible cost.

5.4 How much of altruistic behaviour is explained by mind-directed preferences?

Note that a preference for maintaining a good image is also a social preference and a mind-directed one: people would prefer that others think well of them. It is not to be confused with a strategic investment in reputation done in view of increasing future gains. We therefore have a hierarchy of nested intentions and different ultimate goals that potentially explain prosocial choice. Prosocial choices can be motivated by:

A. Preferences about the distribution of material goods—

preference for a fair distribution, for instance. And they can be motivated by some further intended effect of the distribution:

- B. not disappointing the partner (by means fulfilling his or her expectation about the distribution);
- C. being well thought of (by means of not disappointing the partner, by means of fulfilling his or her expectations about the distribution);
- D. increasing expected payoff (by means of being well thought of, by means of not disappointing the partner, by means of fulfilling his or her expectations about the distribution).

We have argued that A fails to explain some of the observed behaviour, and D, even though it obviously motivate many choices, falls short of explaining generous transfer in one shot anonymous games. B and C, i.e., social esteem models and aversion to disappointing, are mind-directed preferences that have much potential for explaining altruistic choices. Even though the literature on social preference has bloomed in experimental economics in the last decades, there has been a large focus on distributive preferences. Hopefully, this paper will contribute to switching the focus towards mind-directed preferences and psychological games (Geanakoplos, Pearce & Stacchetti, 1989). It provides evidence that people are not keen to disappointing others but that they consider, with some guile, which expectations are worth fulfilling and which they can disregard.

References

- Andreoni, J., & Bernheim, B. (2009). Social image and the 50–50 norm: A theoretical and experimental analysis of audience effects. *Econometrica*, 77(5), 1607–1636.
- Andreoni, J., & Rao, J. M. (2011). The power of asking: How communication affects selfishness, empathy, and altruism. *Journal of Public Economics*, 95(7), 513–520.
- Ariely, D., Bracha, A., & Meier, S. (2009). Motivation and Monetary Doing Good or Doing Well? Image in Incentives Behaving Prosocially. *The American Economic Review*, 99(1), 544–555.
- Battigalli, P., & Dufwenberg, M. (2007). Guilt in Games. *The American Economic Review*, 97(2), 170–176.
- Baumard, N., André, J.-B., & Sperber, D. (2013). A mutualistic approach to morality: The evolution of fairness by partner choice. *Behavioral and Brain Sciences*, 36(01), 59–78.
- Bénabou, R., & Tirole, J. (2006). Incentives and prosocial behavior. *The American Economic Review*, 96(5), 1652–179.
- Bénabou, R., & Tirole, J. (2009). Intrinsic and Extrinsic Motivations. *The Review of Economic Studies*, 70(3), 489–520.

²¹Kurzban (2010) convincingly argues that the evolved function of self-esteem is to foster behaviors that are good for one's social reputation.

- Binmore, K. (1998). *Playing fair*. MIT Press.
- Binmore, K. (2005). *Natural Justice*. New York: Oxford University Press.
- Brandts, J., & Charness, G. (2011). The strategy versus the direct-response method: a first survey of experimental comparisons. *Experimental Economics*, 14(3), 375–398.
- Broberg, T., Ellingsen, T., & Johannesson, M. (2007). Is generosity involuntary? *Economics Letters*, 94(1), 32–37.
- Charness, G., & Dufwenberg, M. (2006). Promises and Partnership. *Econometrica*, 74(6), 1579–1601.
- Charness, G., & Rabin, M. (2002). Understanding Social Preferences with Simple Tests. *The Quarterly Journal of Economics*, 117(3), 817–869.
- Cronk, L., & Wasieleski, H. (2008). An unfamiliar social norm rapidly produces framing effects in an economic game. *Journal of Evolutionary Psychology*, 6(4), 283–308.
- Dana, J., Cain, D. M., & Dawes, R. M. (2006). What you don't know won't hurt me: Costly (but quiet) exit in dictator games. *Organizational Behavior and Human Decision Processes*, 100, 193–201.
- Dana, J., Weber, R. a., & Kuang, J. X. (2007). Exploiting moral wiggle room: experiments demonstrating an illusory preference for fairness. *Economic Theory*, 33(1), 67–80.
- DeScioli, P., Massenkoff, M., Shaw, A., Petersen, M. B., & Kurzban, R. (2014). Equity or equality? Moral judgments follow the money. *Proceedings of the Royal Society B: Biological Sciences*, 281(1797), 20142112–20142112.
- Dreber, A., Ellingsen, T., Johannesson, M., & Rand, D. G. (2013). Do People Care About Social Context? Framing Effects in Dictator Games. *Experimental Economics*, 16(3), 349–371
- Dufwenberg, M. (2008). Psychological games. *The New Palgrave Dictionary of Economics*, 2(1988).
- Ellingsen, T., Johannesson, M., Tjøtta, S., & Torsvik, G. (2010). Testing guilt aversion. *Games and Economic Behavior*, 68, 95–107.
- Engel, C. (2011). Dictator games: A meta study. *Experimental Economics*, 14, 583–610.
- Fehr, E., Fischbacher, U., & Gächter, S. (2002). Strong reciprocity, human cooperation, and the enforcement of social norms. *Human Nature*, 13(1), 1–25.
- Frith, C. D., & Frith, U. (2012). Mechanisms of social cognition. *Annual Review of Psychology*, 63, 287–313.
- Geanakoplos, J., Pearce, D., & Stacchetti, E. (1989). Psychological games and sequential rationality. *Games and Economic Behavior*, 1, 60–79.
- Haley, K. J., & Fessler, D. M. T. (2005). Nobody's watching? Subtle cues affect generosity in an anonymous economic game. *Evolution and Human Behavior*, 26(3), 245–256.
- Hashimoto, H., Mifune, N., & Yamagishi, T. (2014). To Be Perceived as Altruistic: Strategic Considerations That Support Fair Behavior in the Dictator Game. *Letters on Evolutionary Behavioral Science*, 5(2), 17–20.
- Heintz, C. (2013). What can't be inferred from cross-cultural experimental games. *Current Anthropology*, 54(2), 165–166.
- Henrich, J., Boyd, R., Bowles, S., Camerer, C. F., Fehr, E., & Gintis, H. (2004). *Foundations of Human Sociality: Economic Experiments and Ethnographic Evidence from Fifteen Small-Scale Societies*. (J. Henrich, R. Boyd, S. Bowles, C. Camerer, E. Fehr, & H. Gintis, Eds.). Oxford University Press.
- Koch, A. K., & Normann, H.-T. (2008). Giving Dictator Games: Regard for Others or Regard by Others? *Southern Economic Journal*, 75(1), 223–231.
- Kovács, Á. M., Téglás, E., & Endress, A. D. (2010). The social sense: Susceptibility to others' beliefs in human infants and adults. *Science*, 330, 1830–1834.
- Kurzban, R. (2010). *Why everyone (else) is a hypocrite*. Princeton University Press.
- Leslie, A. M., Friedman, O., & German, T. P. (2004). Core mechanisms in “theory of mind”. *Trends in Cognitive Sciences*, 8(12), 528–33.
- Lesorogol, C. (2007). Bringing Norms In: The Role of Context in Experimental Dictator Games. *Current Anthropology*, 920–926.
- List, J. J. A. (2007). On the interpretation of giving in dictator games. *Journal of Political Economy*, 115(3), 482–493.
- Miceli, M., & Castelfranchi, C. (2002). The mind and the future: the (negative) power of expectations. *Theory and Psychology*, 12(3), 335–366.
- Misyak, J. B., Melkonyan, T., Zeitoun, H., & Chater, N. (2014). Unwritten rules: virtual bargaining underpins social interaction, culture, and society. *Trends in Cognitive Sciences*, 1–8.
- Ockenfels, A., & Werner, P. (2012). “Hiding behind a small cake” in a newspaper dictator game. *Journal of Economic Behavior & Organization*, 82(1), 82–85.
- Rabin, M. (1993). Incorporating Fairness into Game Theory and Economics. *The American Economic Review*, 83(5), 1281–1302.
- Rankin, F. (2006). Requests and social distance in dictator games. *Journal of Economic Behavior & Organization*, 60, 27–36.
- Reuben, E., Sapienza, P., & Zingales, L. (2009). Is mistrust self-fulfilling? *Economics Letters*, 104(2), 89–91.
- Samson, D., Apperly, I. a, Braithwaite, J. J., Andrews, B. J., & Bodley Scott, S. E. (2010). Seeing it their way: evidence for rapid and involuntary computation of what other people see. *Journal of Experimental Psychology. Human Perception and Performance*, 36(5), 1255–66.
- Trachtman, H., Steinkruger, A., Wood, M., Wooster, A., Andreoni, J., Murphy, J. J., & Rao, J. M. (2015). Fair weather avoidance: unpacking the costs and benefits of

- “Avoiding the Ask.” *Journal of the Economic Science Association*, 1(1), 8–14.
- Vanberg, C. (2008). Why Do People Keep Their Promises? An Experimental Test of Two Explanations. *Econometrica*, 76(6), 1467–1480.
- Vranceanu, R., Sutan, A., & Dubart, D. (2012). Trust and Financial Trades: Lessons from an Investment Game Where Reciprocators Can Hide Behind Probabilities. *The Journal of Socio-Economics*, 41(1), 72–78
- West, S. A., El Mouden, C., & Gardner, A. (2010). Sixteen common misconceptions about the evolution of cooperation in humans. *Evolution and Human Behavior*, 32(4), 231–262.
- Winking, J., & Mizer, N. (2013). Natural-field dictator game shows no altruistic giving. *Evolution and Human Behavior*, 34(4), 288–293.