Objective and Subjective Compliance: How 'Moral Wiggle Room' Opens

Kai Spiekermann* & Arne Weiss†

Abstract

We propose a cognitive-dissonance model of norm compliance to identify conditions for strategic information acquisition. The model distinguishes between: (i) objective norm compliers, for whom the right action is a function of the state of the world; (ii) subjective norm compliers, for whom it is a function of their belief. The former seek as much information as possible; the latter acquire only information that lowers, in expected terms, normative demands. The source of 'moral wiggle room' is not belief manipulation, but the coarseness of normative prescriptions. In a novel experimental setup, we find evidence for such strategic information uptake.

JEL Classification: C91, D63, D83

Keywords: norm compliance, cognitive dissonance, uncertainty, self-serving biases, strategic learning, dictator games

Consider a company owner who has to decide on the size of a voluntary performance bonus for a leaving employee. Suppose employer and employee agree on the social norm that applies to this decision: 'if an employee performed well, he ought

*(Corresponding author), London School of Economics Department of Government, Houghton Street , London WC2A 2AE, United Kingdom, Email: k.spiekermann@lse.ac.uk

†Staatswissenschaftliches Seminar für Experimentelle Wirtschafts- und Verhaltensforschung, Aachener Straße 217, D-50931 Cologne, Germany

to be rewarded appropriately'. However, even though this norm may be universally endorsed and the employer willing to comply, a conflict-free bonus setting is not guaranteed. Social norms generally leave a fairly large room for interpretation (Hechter and Opp, 2001), and the employer and employee will typically disagree in predictable ways about the employee's performance and what size of bonus is 'appropriate'. In their discussion, they might evoke arguments of fairness that favor their own narrow self-interest, and appeal to relevant facts in a selective, self-serving manner (Messick and Sentis, 1979; Babcock et al., 1995). This shows that the notion of fairness can be distorted and perceived through self-interested eyes. In this paper we demonstrate that self-serving biases can have an even stronger effect: they create incentives to strategically seek and avoid information in normative choice situations.

The example suggests that self-serving biases in norm compliance are based on different mechanisms: they can make use of uncertainty about either the behavioral rule to be applied ('what *exactly* am I supposed to do?') or about the state we are in ('what *exactly* is the case?'). We call the former *normative* uncertainty and the latter *factual* uncertainty. Konow (2000) develops a model of self-serving biases due to the former and experimentally tests for these biases by studying the strategic manipulation of beliefs about the applicable norm. Related experiments are conducted by Bicchieri and Chavez (2013), Bicchieri and Mercier (2013), and Rodriguez-Lara and Moreno-Garrido (2012).

Dana, Weber and Kuang (2007, henceforth DWK) find evidence for the relevance of factual uncertainty, demonstrating that uncertainty over the negative externalities of one's behavior can open 'moral wiggle room' (see also Krupka and Weber, 2013; Grossman, 2010; Van der Weele, 2012; Conrads and Irlenbusch, 2013; Larson and Capra, 2009; Matthey and Regner, 2011). More specifically, they show that uncertainty over the receiver's payoff considerably increases 'selfish' choices. Strikingly, the uncertainty can be resolved virtually costlessly, but most dictators with selfish behavior decide to remain ignorant, suggesting that these dictators 'have an illusory preference for fairness' and 'dislike appearing unfair' (DWK, p. 67). DWK's results, in line with considerable evidence on the context-specificity of giving-behavior (see Bardsley, 2007)[1], have important ramifications, both for

[1]Most notably, Cherry et al. (2002) find dictator-giving to all but vanish when wealth is earned

the widespread use of dictator-game giving as a measure of 'pro-social preferences' (advocated, for instance, in Camerer and Fehr, 2004, with a notable critique by Levitt and List, 2007) and, more fundamentally, for explaining 'pro-social' behavior by directly incorporating preferences over final allocation of wealth into the utility function (e.g., Bolton and Ockenfels, 2000; Fehr and Schmidt, 1999).

Largely missing from this debate, however, is a parsimonious theory to explain how precisely the 'moral wiggle room' opens. DWK's interpretation is that uncertainty about the consequences of one's action is a 'veil' that allows people to make selfish choices while maintaining the illusion that one is a non-selfish type. For this 'self-deception' (Benabou and Tirole, 2011, p. 824) to work, the signal of a selfish choice must be processed in different ways under transparent and intransparent conditions. With lack of transparency, beliefs about one's moral convictions, identity or self-image must be updated in a skewed, i.e., strategically non-Bayesian way, while transparent conditions mitigate such biases. For this theory to account for DWK's data, it also has to explain why the majority of DWK's dictators do in fact *not* 'wiggle' and instead behave consistently with the assumption of 'pro-social' preferences by revealing the game's payoff-structure and choosing the 'fair' option. Either the 'fair informed' dictators are genuine non-selfish types or they are less receptive to using the lack of transparency as a 'veil'.

We propose a different approach: we locate the source of 'moral wiggle room' not in belief manipulation, but in the way normative prescriptions respond to signals about the state of the world. Furthermore, our model does not rely on heterogeneity in 'pro-social' preferences to account for the co-presence of 'strategic ignorance' and full information acquisition in normative choice situations. We show that dictators, holding Bayesian beliefs about the state of the world, can use factual uncertainty in a self-serving manner if individuals interpret the demands of a norm in a way that opens the door to strategic behavior: they are 'subjective norm compliers'; that is, they take normative demands as a function of their beliefs. Subjective norm compliers will strategically avoid as well as *seek* information if the mapping from beliefs to prescriptions is 'coarse'. This means that the normative obligations sometimes stay put despite an update of beliefs that makes the normatively more demanding state more probable. By contrast, agents will

and anonymity guaranteed.

acquire full information if they are 'objective norm compliers', who derive their normative demands from what is factually the case.

Our model of norm compliance under uncertainty offers a novel take on DWK's results and provides the first attempt to explain both strategic ignorance and full-information acquisition within one parsimonious framework. It also implies new testable predictions. In particular, our model suggests that ignorance itself is not a desirable state (as assumed by DWK, Bicchieri, 2006, 128–9, and others). Rather, individuals will only stay ignorant if resolving uncertainty will, in expected terms, increase normative obligations. By contrast, individuals will actively *seek* information if, in expected terms, normative obligations will decrease. Therefore, our framework goes beyond the notion of 'strategic ignorance' with its exclusive emphasis on information avoidance.

To fully develop the implications of objective and subjective compliance for the incentives to resolve or maintain factual uncertainty, we introduce a new setup that departs from DWK's experimental paradigm. In section I, we motivate the different notions of norm compliance and the formal model that follows in section II. In section III, we present applications and relate the model to the existing literature on strategic ignorance. We then test the model's core predictions with a laboratory experiment in section IV. In the final section, we summarize this paper's theoretical and empirical contributions and put them into perspective.

I. Objective and subjective norm compliance

A social norm tells us what we ought to do if we find ourselves in a certain situation; more technically, a social norm provides a mapping from a state to a behavioral rule, i.e., a prescribed or proscribed act (see Bicchieri, 2006, ch. 1). In short, social norms take the form: 'if X obtains, I ought to do ϕ'. The condition 'if X obtains' for which the norm prescribes a certain behavior ('I ought to do ϕ') can be interpreted in different ways. On the one hand, the clause can be substituted by 'if X is the case'. This *objective* interpretation leads to prescriptions that are conditional only on the state of the world and are thus entirely independent from the agent's beliefs. With this formulation, the norm prescribes actions no matter what the agent believes about the situation. Consequently, an *objective norm* *complier* strives to perform the action the state of the world demands. On the other hand, the clause can be substituted by 'if I believe that X'. This *subjective* interpretation leads to prescriptions that are entirely contingent on the beliefs of the agent. If the norm is understood in the subjective sense, it is only prescriptive if the agent has the specified belief. Thus, a *subjective norm-complier* strives to perform the action his beliefs demand.

This distinction is echoed in ethical theory and epistemology.[2] For example, Zimmerman (2008) distinguishes between the objective and subjective view of moral obligations. According to the objective view, one's moral obligations are determined only by the relevant facts, not by what one knows about these facts. The objective view therefore entails that ignorance or incomplete information does not change one's moral obligations at all (though most proponents of this view would say that wrongdoing due to exculpatory ignorance may be blameless). According to the subjective view, by contrast, moral obligations are a function of one's knowledge. This entails that incomplete or false knowledge of the facts changes one's moral obligations.[3]

Depending on one's subjective or objective understanding of norms, one can be subject to different forms of psychological costs when violating a norm. Following Festinger (1957), and, more recently, Konow (2000), these psychological costs from non-compliance can be called *cognitive dissonance*. The dissonance arises because the agent experiences an unpleasant tension between what she ought to do, and what she actually does. In Konow's model, the experienced dissonance is traded off against utility from violating the norm, leading to more selfish behavior than prescribed by the norm. The agents perform norm transgressions as long as they result in more payoff utility than the disutility caused by these transgressions.

Following the objective norm-compliance interpretation, agents experience cognitive dissonance if the conjunction of the state of the world and the norm implies

[2]In this context, the core question in ethical theory is whether moral obligations (or moral blame) do depend on knowledge. In epistemology (or, more precisely, in the ethics of belief), the question is which obligations we have to gather evidence. See Ross (1939, ch. VII), and more recently Jackson (1991) for foundational debates in the former, and Kornblith (1983), Feldman (2000), and Chignell (2013) for the latter.

[3]Zimmerman finds neither view convincing and instead defends the prospective view, which (very roughly) expresses moral obligations as determined by expected values, given one's available evidence.

(or might imply) prescriptions they (possibly) violate. Following the subjective norm compliance interpretation, individuals experience cognitive dissonance if the conjunction of their beliefs and the norm implies prescriptions they don't comply with.

Returning to our example of the end-of-contract bonus, suppose the employer does not know how well the employee actually performed. Let's assume that the employee either did perform well and deserves a bonus payment $B > 0$ (according to the social norm shared by the employer), or did not perform well and does not deserve any bonus. If the employer is an objective norm complier, she will suffer from dissonance, arising from the possibility of violating a norm. This is because under uncertainty about the performance of the employee, an objective norm complier faces a dilemma: any amount of bonus payment causes cognitive dissonance because it is possible that the bonus does not match what is fair in the *actual* (but unknown) state. For example, giving a 'compromise' bonus of $B/2$ leads to dissonance in expected terms: the bonus is either too high in case the employee performed poorly, or too low in case the employee performed well. By contrast, if the employer is a subjective norm complier, she experiences dissonance caused by the difference between her belief-based prescription and her selfish actions. For instance, she might think that $B/2$ is the fair amount to give in case her epistemic state (her belief about the state of the world) is one of uncertainty, and therefore does not suffer dissonance when giving $B/2$, in stark contrast to the objective norm complier.

Under certainty, objective and subjective compliance are behaviorally equivalent because the epistemic state matches the state of the world. However, the two types of compliance come apart under uncertainty. For a subjective norm complier, uncertainty is just another possible epistemic state to which a norm applies. Doing what the epistemic state requires suffices to meet one's obligations. By contrast, an objective norm complier will suffer from dissonance under uncertainty because she cannot (at the same time) comply with what the norm prescribes for two different states. In other words: an objective norm complier wants to 'get it right' and suffers dissonance because she (potentially) fails to do so. Given this difference, objective and subjective compliers have different incentives to resolve uncertainty. If our employer is an objective norm complier, she is better off with more information because this increases her chance to match her action (giving a bonus) to the state (employee's performance). If the employer is a subjective norm-complier, by contrast, she does not have any general incentive to gather more information; instead, she will strategically choose the type of information that leads, in expected terms, to less demanding prescriptions.

Lowering the normative demand by strategic information choice is possible if the relevant norm is 'coarse', in the sense that it does not always respond to changes in degrees of belief. To illustrate one simple case, let the norm only distinguish between three types of beliefs (which we call epistemic states) with the corresponding decreasing levels of normative demand: H (employee performed well; high demand), U (uncertainty over the employee's performance; medium demand) and L (employee performed poorly; low demand). A subjective norm complier will prefer to be in state L and try to avoid being in state H. In a state of uncertainty, a subjective norm complier will actively seek signals on which he has a prior belief that they offer reaching state L without risking attaining state H. He will avoid signals that he believes may lead to state H without offering the opportunity to reach state L.

Before formalizing this choice of signals, an illustration can elucidate which sort of information the two types of agents would gather in different ways. Suppose the employee (from our earlier example) has worked in a customer service function, and the employer, unable to observe the employee's performance directly, has to ask clients to learn whether the employee performed well or poorly. To find out, she can contact two different, honest clients for references: one known to be a friend and one known to be a foe of the employee. The friend likes the employee, but tends to have a poor response rate to such reference requests. If the employer asks him and the employee's performance was good, the friend sometimes replies with a positive report and sometimes fails to respond. If there is nothing positive to report, the friend will never respond. The foe does not like the employee and has an equally low response rate. If the employer asks him and the employee's performance was poor, he sometimes replies with a negative report and sometimes fails to respond. The foe never responds if the employee performance was positive.

In other words, the employer has the option to choose one or two types of probabilistic signals about the employee's performance. Each signal can only lead to certainty for one of the two possible states of the world, as displayed in Table 1.

		employee's performance	
		high	**low**
reference from	**friend**	'high' or no response	no response
	foe	no response	'low' or no response

Table 1: Friend and foe reports on employee's performance.

An objective norm complier, who wants to get it right, will choose both types of signals because this increases her chance to pay out the correct bonus. By contrast, a subjective norm complier might follow a norm that prescribes the same behavior for any outcome in which the employer remains uncertain about the performance. Under the assumption of such a coarse-grained norm, a subjective complier will only ask the foe, as it offers a risk-free prospect to lower the normative demands on him: He will either learn of the employee's low performance and give no bonus or he will remain uncertain with unchanged normative demands.

Similar strategic incentives to strategically acquire information can also apply in many other settings. Many people endorse norms prohibiting animal cruelty; but they then buy cheap dairy products, ignoring information about cruel production methods, while lapping up news about the failures of organic farming, thereby seeking justification for the consumption of the cheapest products. In the same vein, people may endorse a norm against harming others. At the same time, they cause excessive greenhouse gas emissions, ignoring information about the adverse impacts of climate change, while seeking out reports that emphasize the uncertainties in climate research, thereby promoting their belief that greenhouse gas emissions are not harmful. This behavior can be rationalized within our model by assuming coarseness in the mapping from beliefs to prescriptions and the availability of signals on which a subjective norm complier has prior beliefs on whether he may attain an attractive or an unattractive epistemic state after acquiring the signals.

II. Formal model

In order to introduce a formal distinction between objective and subjective compliance, we model a dictator game, enriched by social norms of equity, such that the receivers are more or less deserving. The dictator is initially uncertain about the deservingness of the receiver, but can acquire signals to eliminate this uncertainty. We will embed this setting into a simplified version of Konow's (2000) model, and then extend this baseline model to fit it to our setup.

The dictator has the amount $\overline{y}$ to distribute, such that he gives y to himself and $x = \bar{y} - y$ to the receiver ($0 \leq y \leq \bar{y}$). The payoff utility derived from this decision is $v(y)$, with the usual assumption of positive, but decreasing marginal utility of money, such that $v'(y) > 0$ and $v''(y) < 0$. If dictators were maximizing utility from monetary payoff only, their obvious choice would be to set $y = \bar{y}$. But besides monetary payoff, dictators also care about limiting the difference between what they think is normatively required and what they actually give. The greater the difference, the higher are the non-compliance costs. Konow thinks about these costs in terms of *cognitive dissonance*, but it is equally plausible to think about image concerns or about norm violation costs more generally.

For simplicity, we assume a dichotomous state space $\Omega = \{L, H\}$, where L and H can conveniently be interpreted as 'low' and 'high', indicating the deservingness of the receiver.[4] We call the actual state ω. For an objective norm complier, the amount one is normatively permitted to keep (the 'fair point') is a function of the actual state ω; this action is characterized by keeping ϕ_ω. For a subjective norm complier, by contrast, the normatively required action is a function of an epistemic state, characterized by the (Bayesian) probability p that state L obtains (and $1-p$ that state H obtains). This distinction between objective and subjective definitions of what is normatively required, expressed by ϕ_ω and ϕ_p respectively, is the crucial extension of Konow's model. To simplify the model, we assume, in contrast to Konow, that the ϕs for all (epistemic or factual) states are fixed exogenous variables for the dictator.[5] In the setting we consider, we assume the

[4] A generalization to other state space partitions should be straightforward.

[5] Konow (2000) also considers self-serving biases in the belief formation about what is fair. As our research question focuses on strategic acquisition of information on the receiver's entitlement once the dictator has formed his belief on what he ought to do under certainty,

ϕs to be determined by shared empirical and normative expectations of a relevant reference group.[6]

The dissonance cost experienced by the dictator is a function of the difference $\Delta = |y - \phi|$ between the normatively required amount to keep and what she actually keeps, and the dissonance cost function $f(y, \phi) = f(\Delta)$ determines the experienced disutility. As in Konow (2000), f is a twice differentiable, strictly convex function, such that $f(0) = 0$ (that is, if $y = \phi$), $f'(\Delta) > 0$ for $\Delta \neq 0$, and $f''(\Delta) > 0$.

The dictator's decision problem is to trade off the utility from keeping more money against the disutility from cognitive dissonance created by deviating from the perceived fair distribution:

$$\max_y \mathrm{E}\left[u(y, \phi)\right] \equiv \mathrm{E}[v(y) - f(y, \phi)] \text{ subject to } 0 \leq y \leq \overline{y},\ 0 \leq \phi \leq \overline{y}.$$

The behavior of a dictator is therefore determined by fair point ϕ and by the relative value of money and cognitive-dissonance avoidance.

We first model maximization under certainty, in which factual and epistemic states match. Let ϕ_L and ϕ_H denote the fair points given the respective state, assuming that $\bar{y} > \phi_L > \phi_H > 0$ (that is, it is fair to keep more in state L than in state H, but it is never fair to keep everything). Let $\hat{u}(\phi) = \max_y(v(y) - f(y, \phi))$ be the maximum utility achievable as a function of ϕ. This function is increasing and concave in ϕ (the proof is provided in the appendix). Since we have assumed that $\phi_L > \phi_H$, it follows that $\hat{u}(\phi_L) > \hat{u}(\phi_H)$.

When deciding on an allocation in state H, the dictator reaches the maximum utility $\hat{u}(\phi_H)$, with $y_H^* = \underset{y_H}{\operatorname{argmax}}(u(y_H, \phi_H))$. Similarly, when deciding on an allocation in state L, the maximum utility is $\hat{u}(\phi_L)$, with $y_L^* = \underset{y_L}{\operatorname{argmax}}(u(y_L, \phi_L))$. To ensure an interior solution, we assume that $v'(\bar{y}) < \frac{\partial f(\bar{y}, \phi_L)}{\partial y}$, which implies that

we do not model how individuals reach their normative beliefs. We therefore take them as given, even though we concur that this is another important form of self-serving bias. For simplicity, we assume that both types of players have the same normative beliefs (e.g. because of the salience of applicable norms so that the cost of manipulating normative beliefs are prohibitive). However, our model could readily be extended to account for differences in normative beliefs between types by introducing a belief-manipulation process as in Konow (2000).

[6] For details see Bicchieri (2006, ch. 1).

$y_L^* < \bar{y}$ and $y_L^* > y_H^*$.

The novel aspect of our model is the treatment of uncertainty. For an objective norm complier, cognitive dissonance depends on the state and can only be $f(y, \phi_H)$ or $f(y, \phi_L)$. Thus, under uncertainty, an objective complier cannot make sure to choose the morally appropriate action that the (unknown) state requires. Therefore, the *maximum expected* utility under uncertainty is achieved when keeping the amount $y_U^* = \underset{y_U}{\operatorname{argmax}}(v(y_U) - (1-p)f(y_U, \phi_H) - pf(y_U, \phi_L))$.

For a subjective norm complier, the normatively required action is a function of his epistemic state. A norm implies a function that maps a state onto a required action, in our case the amount one is permitted to keep. For the subjective norm complier, this function can take different forms. One simple form only distinguishes between three epistemic states: knowing that the receiver is deserving, knowing that the receiver is undeserving, or being uncertain about the receiver's deservingness. In such a case, the fair amount ϕ is determined by a step function:

$$\phi_p = \begin{cases} \phi_L & \text{if } p \geq 1-\epsilon \\ \phi_U & \text{if } \epsilon < p < 1-\epsilon \\ \phi_H & \text{if } p \leq \epsilon. \end{cases} \tag{COARSE}$$

We assume that $\phi_L > \phi_U > \phi_H$. The parameter ϵ ($\frac{1}{2} > \epsilon \geq 0$) represents the margin of tolerance for 'near certainty'.

We believe that social norms typically only provide such a coarse-grained mapping from states to prescriptions. For a social norm to exist, individuals must be able to reliably distinguish between compliant and non-compliant agents in order to form behavioral expectations and sanction transgressions. Since degrees of beliefs are not observable in detail, it is unlikely that social norms take them as argument with any great precision. This is mirrored in our everyday language regarding normative choices, in which we rarely refer to degrees of beliefs (let alone a Bayesian updating of normative prescriptions). The maximum utility under uncertainty is $\hat{u}(\phi_p)$, with $y_p^* = \underset{y}{\operatorname{argmax}}(u(y, \phi_p))$.

Before making their allocation decision, dictators can optionally acquire costless signals to learn about the state of the world. There are two different signals available, represented by random variables S_L and S_H, and for each the dictator can choose whether she wants to receive them. When obtaining signal S_L, the dictator has a chance to learn that the state is L. More precisely, the conditional probability $s \in (0,1)$ of learning that the state is L, given that the state is indeed L, is $\mathrm{s} = \Pr(S_L = L|\omega = L)$. Similarly, when obtaining signal S_H, we assume the same value s for the conditional probability of learning that the state is H, given that the state is indeed H: $s = \Pr(S_H = H|\omega = H)$. In all other cases the dictator receives a 'null' signal, which means the dictator remains uncertain about the state. After a null signal, dictators perform a Bayesian update on the probability p that state L obtains. If the dictator receives only signal $S_L = 0$, then she updates such that $p' = (1-s)p/\left((1-s)p + (1-p)\right)$. Similarly, if the dictator receives $S_H = 0$, then she updates such that $p' = p/(p + (1-s)(1-p))$. We assume throughout that receiving a null signal never removes uncertainty, such that $\epsilon < p' < 1 - \epsilon$. Finally, if a dictator acquires both signals, but both are null signals, no update is necessary, as the two signals cancel each other out.

We can now state how objective and subjective norm compliers will acquire the signals on offer.

Proposition 1. *Under uncertainty, objective compliers will acquire both signal S_L and signal S_H.*

Both types of signals increase the dictator's chance to reach his utility maxima given the respective states. The dictator has an expected utility gain of $sp(\hat{u}(\phi_L) - u(y^*_U, \phi_L))$ by acquiring S_L and of $s(1-p)(\hat{u}(\phi_H) - u(y^*_U, \phi_H))$ by acquiring S_H. Intuitively, by acquiring S_L (S_H) he gains the additional utility from keeping y^*_L (y^*_H) instead of keeping y^*_U provided that the state is indeed L (H) weighted with the probability that the signal is L (H). The proof is provided in the appendix.

Proposition 2. *Under uncertainty, subjective compliers who follow a coarse-grained norm (COARSE) will acquire signal S_L, but not signal S_H.*

For a sketch proof, recall that the maximum utility is increasing in the fair amount to keep: $\hat{u}(\phi_L) > \hat{u}(\phi_U) > \hat{u}(\phi_H)$. COARSE, together with the assumption that a null signal never removes uncertainty, ensures that any update from p to p' after receiving a null signal does not change ϕ_p. It is immediately obvious that obtaining signal S_L is beneficial because there is no down-side risk, but a potential gain: it only increases the probability of receiving the highest utility $\hat{u}(\phi_L)$. And it is equally obvious that obtaining signal S_H is never beneficial because there is a down-side risk, but no potential gain: it only increases the probability of receiving the lowest utility $\hat{u}(\phi_H)$. A formal proof is provided in the appendix.[7]

In our model all dictators hold unbiased (Bayesian) beliefs and all types of dictators would prefer to be in a world in which they are paired with a low performer and would use this fact to give little. However, an objective norm complier, who only cares about the state of the world, cannot change the state and is therefore always better off with more information. A subjective norm complier with coarse-grained norms, by contrast, who takes prescriptions as a function of his beliefs, has 'moral wiggle room' (DWK).

III. Model Applications and Extensions

A. Climate Change

The model is flexible in its application to different situations of subjective norm compliance. In order to do so, think of the fairness points ϕ_H, ϕ_L, and ϕ_U as posing a high demand (a costly action), a low demand (an inexpensive action) and a moderate demand (a mildly costly action), respectively. One example is a social norm saying that the amount of greenhouse-gas emissions an individual is allowed to emit is linked to the belief about whether greenhouse-gas emissions cause severe externalities on future generations. For simplicity, assume there are just two possible states of nature with respect to global warming: state H, in which climate change is anthropogenic with severe consequences for future generations, and state L, in which the change in global temperature is either outside human control or will be easily dealt with in the future. Many people endorse a social norm that requires to choose a lifestyle according to the belief about the state,

[7] All that is really necessary to allow for selective information acquisition is some element of coarseness in the norm's response to new evidence, i.e., a range of probabilities from which the same prescriptions follows. In this case, modified signals $S_L^{'}$ and $S_H^{'}$ exist such that a subjective complier will acquire $S_L^{'}$ but reject $S_H^{'}$ when being offered both.

which may have the step-wise form from above:

$$\phi_p = \begin{cases} \phi_L & \text{if } p \geq 1-\epsilon \\ \phi_U & \text{if } \epsilon < p < 1-\epsilon \\ \phi_H & \text{if } p \leq \epsilon. \end{cases}$$

As before, p is the belief that L is the case. With $p \geq 1-\epsilon$, agents are reasonably certain to be in state L (such that ϕ_L applies) and with $p \leq \epsilon$ to be in state H (such that ϕ_H applies). For degrees of beliefs between the two threshold levels, the agent is uncertain about the state such that action ϕ_U (a moderately carbon-conscious lifestyle) is prescribed.

If an agent's prior belief is between the two threshold levels $\epsilon < p < 1-\epsilon$ and there is a signal on offer that will either change the belief to $p \geq 1-\epsilon$ or leave it in the boundaries of uncertainty $\epsilon < p < 1-\epsilon$, then the agent has an incentive to acquire this signal (and avoid other signals that might bring about $p \leq \epsilon$). As the 'green lifestyle' (ϕ_H) requires drastic changes in consumption patterns, agents would rather be in state L, in which driving gas-guzzling cars and flying long distances are permitted. Such subjective compliers will not read articles in a Greenpeace magazine (or the IPCC reports for that matter), but may well pay active attention to 'climate change deniers'. By consuming information from the latter, the agents expect to either find convincing evidence that there is not much to worry about one's personal lifestyle ($S_L = L$) or to learn nothing substantially new at all ($S_L = 0$). The 'moral wiggle room' of proposition 2 opens because the social norm is coarse-grained such that receiving a null-signal does not cause a change in what is normatively required.[8]

B. Binary fairness points: an alternative explanation of strategic ignorance

DWK (2007) provide experimental evidence that agents often prefer to remain ignorant about the negative consequences of their selfish acts. We will show that our model offers a parsimonious explanation of their results.

[8]Formally, that is the case if $\epsilon < p' = (1-s)p/((1-s)p + (1-p))$.

The DWK subjects play a binary version of the dictator game. Dictators can choose between actions A and B to determine the payoffs for them and their receiver. In the baseline treatment, A results in distribution (6,1), B in (5,5). Here, 74% of dictators choose the fair and efficient option B. By contrast, in the hidden information treatment, the outcomes are assigned in two different ways with equal probability: either A causes (6,1) and B (5,5), or A causes (6,5) and B (5,1). By default, it is unclear whether A or B hurts the receiver, but, importantly, the dictator can optionally resolve this uncertainty costlessly by clicking on a button. Almost half of the dictators, however, deliberately remain ignorant and 86% of those choose the payoff-dominant action A, even though there is a chance of $\frac{1}{2}$ to impose a severe negative externality on the receiver.

We call the state in which A causes a negative externality H, and the state in which it does not L, with p as the Bayesian belief that L is the case. DWK's setup can be modeled with a social norm that has only two fairness points:

$$\text{prescription} = \begin{cases} \text{do B} & \text{if } p < \mu \\ \text{do A} & \text{if } p \geq \mu \end{cases} \tag{BINARY}$$

Such a social norm allows to choose the payoff-dominant action as long as the probability that L being the case is at least μ. Because the states are known to be equiprobable, p is $\frac{1}{2}$ under uncertainty. If the threshold level μ ($1 \geq \mu \geq 0$) is not larger than $\frac{1}{2}$, such a binary social norm can explain the seeming discrepancy in the behavior of the dictators in the baseline and hidden information treatment. Many dictators behave consistently with such a binary social norm and choose action B under certainty about the state being H ($p = 0$). Under uncertainty, by contrast, subjective compliers decide to remain ignorant and enjoy the monetary benefits of choosing A. Revealing the game's payoff structure is akin to acquiring a signal that will reveal the state of nature being either H ($p = 0$) or L ($p = 1$). Therefore, by clicking on the button that is being offered, the dictator can only lose: either he will have to choose B or he stays put where he would be under uncertainty anyway – being permitted to choose A.

DWK do not provide a formal model for the behavior they observe, but they suggest that dictators have an 'illusory preference for fairness' and that many of them only want to appear fair, either to themselves or to others. More generally, DWK view ignorance as a particularly desirable state for dictators, as their selfish behavior can be hidden (from others or themselves). We are less convinced that the subjects' behavior is motivated by the desire to hide selfishness. In fact, in our model, ignorance is not used to hide at all – in the DWK setup it simply turns out to be an attractive epistemic state for subjective norm compliers because it is undemanding. Our explanation of strategic information acquisition also does not rely on deceiving oneself or others about one's selfish preferences; rather, signal choices can be modeled as a deliberate and rational process. In addition, our model offers an interesting interpretation for the sizable fraction of dictators in DWK who choose to reveal: they can be modeled as *objective* norm compliers who, unlike the subjective compliers, 'want to get it right' and are therefore better off with more information.

C. Social norms with continuous fairness points

While there are good reasons to take the mapping from beliefs to prescription to be coarse-grained, the model also allows for different mappings. On the other end of the spectrum from coarse to continuous, it could be assumed that the social norm is perfectly sensitive to p, stating a different prescription for each epistemic state. For instance, the fair amount could be a weighted average of the prescriptions under certainty, such that

$$\phi_p = p\phi_L + (1-p)\phi_H. \tag{CONT}$$

Such a norm is directly proportional to the Bayesian belief.[9]

With CONT, the predictions for information acquisition are dramatically different:

Proposition 3. *Under uncertainty, subjective compliers who follow a continuous norm CONT will not acquire any signals.*

Intuitively, signals provide fair lotteries over different levels of ϕ_p. However,

[9]More precisely, CONT minimizes the expected distance between ϕ_p and ϕ_ω, the ϕ for the true (but unknown) state of the world.

because of the decreasing marginal value of money the dictators are risk averse ($\hat{u}$ is concave), therefore all such lotteries have a lower expected utility than the utility of the status quo. A proof is provided in the appendix.

Note that the difference between propositions 2 and 3 lies in the way dictators do or do not update what is normatively required when uncertainty remains after acquiring a signal. When receiving a null signal after acquiring S_L, the fair point of a dictator following CONT will increase such that he will feel forced to give more to the receiver. In expected terms, this increase in the fair point entirely offsets the possible decrease of the fair point if the dictators learns that the state is L. Based on the assumption that marginal utility from keeping money is decreasing, a dictator following CONT will therefore gladly decline the lotteries offered by either S_L or S_H. By contrast, a dictator following COARSE does not distinguish the normative demand arising from different levels of probability under uncertainty, and acquiring S_L either leads to the desired state of certainty (and therefore lowers his normative demands), or to no change at all – an attractive proposition.

IV. Experimental test

Our new experimental setup is designed to test whether individuals strategically *seek* (as well as avoid) normatively relevant information. All studies known to us have only offered a signal similar to S_H in the terminology of our model, but never a signal similar to S_L. This has led to the belief that 'wiggle room' is intimately linked to the option of remaining uncertain or 'strategically ignorant'. In marked contrast, our model suggests that getting some information, but not all information, is often the best way to 'wiggle'.

The theoretical underpinning is provided by proposition 2, which is a generalization of claims pertaining to strategic ignorance. To test proposition 2, our experimental design explicitly implements a coarse-grained norm, distinguishing it from all previous experimental work in this area. The design also uses the crucial elements of the model in section II: a dictator game embedded in social norms of equity such that the receivers are either deserving or undeserving; clearly specified fairness points ϕ_H, ϕ_L and ϕ_U; and, crucially, an opportunity for dictators to acquire signals S_H and S_L, which may reduce the dictator's uncertainty about the deservingness of the receiver.

The null hypothesis H0 states that dictators will acquire S_L and S_H equally often. H0 can be derived in the form of full information acquisition from theories that assume genuine preferences for fair outcomes; it is also captured by proposition 1 of our model, based on the assumption of objective norm compliance. Note that H0 is also consistent with 'strategic ignorance'. In contrast to our setup, DWK and subsequent studies have only ever offered a signal akin to S_H. The novel part of our experimental test is the alternative hypothesis H1, derived from proposition 2, based on the assumption of subjective norm compliance. It states that dictators will seek information strategically by acquiring S_L but not S_H.

Both hypotheses, as the corresponding propositions, rely on the assumption that dictators do in fact have fairness points $\phi_H < \phi_L$. As this cannot be taken for granted for all subjects, we use a 'within-subject' design, in which we first observe the dictators' compliance with norms and later give them the—unannounced—opportunity to acquire signals.

A. Experimental design

We now explain the stages of the experiment in greater detail (a more expansive account is provided in appendix E). First, to establish or reinforce a social norm of appropriate giving as a function of desert, our subjects are informed about the modal normative beliefs in a comparable group regarding the appropriate giving behavior. In particular, we tell our subjects that most participants in an earlier survey session would expect people to give 10 out of 20 Euros to (deserving) high performers and 5 Euros to (undeserving) low performers. We thereby emphasize that deserving and undeserving receivers ought to be treated differently.

Second, we create the types by playing a competitive knowledge quiz. All subjects answer knowledge questions taken from 'Who Wants to be a Millionaire' under time pressure. The best 75% performers are declared 'high performers', the lowest 25% 'low performers'. All subjects are informed that doing well in the quiz (i.e., being a high performer) makes it more likely to (i) be a dictator in a subsequent dictator game; and (ii) to win a 'bonus' of 20 Euro that is available for distribution between the dictator and a receiver later on. We then assign dictator and receiver roles such that all dictators are high performers, while receivers are, in equal shares, high and low performers.

Third, before the dictator-receiver pairs are formed, the dictator game is played with a strategy method. More precisely, all dictators submit a strategy of how much to give:

(i) in case they learn they are paired with a low performer; and

(ii) in case they learn they are paired with a high performer.

When entering the strategy, the dictators do not know whether or under which circumstances information about the type of their receiver might become available to them, but they are told that their strategy choice is binding. Before choosing the strategy, we inform the dictators that if they remain uncertain about the type of their receiver, the mean of the two stated amounts will be transferred to the receiver. This setup is crucial to induce a coarse-grained norm. With the amount for uncertainty fixed externally, the dictators cannot change their giving continuously as a function of their belief about the receiver type.

Fourth, after entering the strategy information, the dictators are paired with equal probability with either a high performer or a low performer, but they do not learn by default which type of receiver they are paired with. The dictators now have an unannounced opportunity to acquire information that may inform them about the type of their receiver. To make this optional information uptake intuitively plausible, we tell the dictators that the information about the (lack of) deservingness of their receiver is contained in exactly one of four envelopes symbolically displayed on screen. If the receiver is a high performer, the information is in one of two envelopes called 'gold envelopes'. If the the receiver is a low performer, the information is in one of two envelopes called 'silver envelopes'. The subjects can open up to one envelope of each type. More formally, the signals available are S_L and S_H, as described above. That implies four different sets of signals can be chosen: $\{\}$, $\{S_L\}$, $\{S_H\}$, or $\{S_L, S_H\}$. A dictator wishing to obtain as much information as possible will open one envelope each, a dictator who only wants to increase the chance of learning that the receiver is a low performer will only open a silver envelope, and so on. The prior probability for the types is $p = \frac{1}{2}$ and the probability of resolving uncertainty when choosing the 'correct' signal is $s = \frac{1}{2}$.

Finally, the dictator game is implemented. Whether a dictator-receiver pair gets a bonus of 20 Euro for distribution depends on the type of the receiver: the bonus is always provided if the receiver is a high performer, but only with probability $\frac{1}{2}$ if the receiver is a low performer. This fact, which the subjects were informed about at the start of the session, underscores the distinction between deserving and undeserving receivers: only being paired with a high performer increases the chance to win a bonus. At the end of the treatment, the bonus (if available) is distributed according to the strategy of the dictator and the information the dictator obtained about the receiver. All parameters and stages in the experiment apart from the information acquisition are common knowledge among the subjects.

B. Procedures

Our subjects were recruited with the online recruitment system ORSEE by Greiner (2004) from the student subject-pool of the Cologne Laboratory for Economic Research at the University of Cologne (CLER). Subjects had not previously participated in dictator-game or normative-choice experiments. However, all subjects had some previous experience with laboratory experiments. We ran one survey session with 26 participants to elicit normative evaluations, which we then used in three subsequent main sessions with 32 participants each, resulting in 48 independent observation of dictator behavior. Subjects took part in only one session and assumed only one role. General instructions about the experiment were provided on paper (see appendix F). The summary part of the instructions was also read aloud to the subjects with two PowerPoint slides facilitating understanding. All subsequent interactions took place at computer terminals in cubicles, controlled with z-Tree (Fischbacher, 2007). Anonymity was guaranteed by ensuring that subjects were randomly matched and by prohibiting communication between the subjects during the experiment. Average payments in the experiment were about 15 Euros, and sessions lasted on average 90 minutes.

C. Results

Before proceeding to the signal acquisition choices, the main variable of interest, we first look at the dictators' allocation strategies. The creation of a wedge in entitlements, in line with the model's assumptions, is successful: dictators would give[10], under certainty, significantly more to a high performer than to a low performer ($p<0.001$, Wilcoxon signed-rank test).[11] As can be seen from Figure 1, no dictator would give more to a low performer than to a high performer. Note that the size of the bubbles in Figure 1 indicates the relative frequency of the coordinate, each data-point corresponding to the transfer strategy of one dictator. On average, dictators would give a substantial amount to both types of receivers. The mean contribution to low performers is about 3.7 and to high performers 6.6.[12]

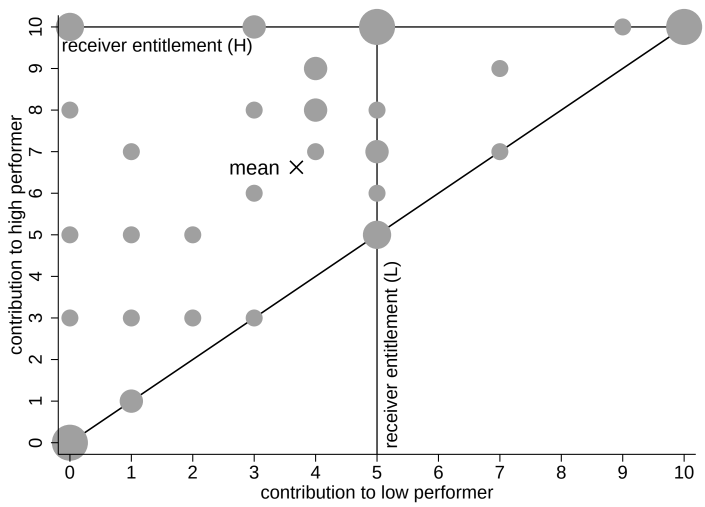


Figure 1: Dictator transfer strategies.

We distinguish between dictators with 'differentiated' and 'undifferentiated' giving strategies. The former give more when they learn that they are paired with a high performer than a low performer, the latter give the same amount. Figure 2 depicts the information acquisition choices of the dictators with differentiated strategies in light gray and of the dictators with undifferentiated strategies in dark

[10]Here we refer to and analyze the dictator's giving strategy. Recall that the actual amount transferred to the receiver is also based on the dictator's signal choices and the bonus draw.

[11]All statistical tests are two-sided.

[12]Note that the significantly ($p<0.001$, Wilcoxon signed-rank test) closer alignment of dictator-giving with normative prescriptions in state L than in state H is also predicted by our model. On average, dictators fall short of what is normatively required by only 1.3 in state L but by 3.4 in state H.

gray. The signal choice of one self-reportedly confused subject[13] is excluded, which leaves us with 30 dictators with differentiated and 17 with undifferentiated strategies. The latter dictators are apparently not receptive to the norm of desert we tried to make salient, either because they refuse to consider any normative considerations (as a 'homo economicus'[14] would), or because they follow a different norm, for instance a norm prescribing an egalitarian distribution.[15] We asked our subjects in a post-experimental questionnaire about the motives of their giving decision. Not all subjects stated clear reasons, but the answers nevertheless shed some light on the difference between differentiating and undifferentiating dictators. Specifically, most of the 11 dictators who mentioned considerations of equality and selfishness did not differentiate.[16] By contrast, the 21 subjects who mentioned either entitlement/desert or norms as motives (while neither mentioning equality nor selfishness) differentiated much more strongly in the giving decision (average difference 4.7 Euros). This confirms that subjects motivated by egalitarian values or selfishness tended to reject the norm we instilled, while others were receptive to it.

As our model is based on the assumption of a wedge in entitlements, we first focus on the dictators with differentiated strategies. For this group of dictators, S_L is the modal choice of signals (43.3%), in line with H1. Acquiring both signals accounts for 30% of types of signal choices. Based on descriptive statistics, subjective compliance therefore organizes the data better than objective compliance or notions of 'genuine fairness preferences'. The null hypothesis of equally frequent choices of S_L and S_H is rejected at a significance level of p=0.029 (Wilcoxon signed-rank test).

The rejection of H0 is driven by the marked difference in selective information acquisition: There are three times as many dictators who only chose S_L than dictators who only chose S_H. As acquiring both signals is the second most frequent choice, this type of behavior cannot be dismissed easily. At first sight, this seems

[13]The subject stated in the post-experimental questionnaire of having mixed up her or his choices.

[14]Within our framework, this type of agent can either be modeled by assuming zero cost of cognitive dissonance or by setting their personal fair points $\phi_H = \phi_L = 20$.

[15]An egalitarian norm can be modeled by setting $\phi_H = \phi_L = 10$

[16]6 subjects mentioned considerations of equality, 4 selfishness, and 1 both equality and selfishness. Among these 11 subjects, 9 had completely non-differentiated giving strategies, the other 2 gave only 2 Euros more in case they learn they are paired with a high performer.

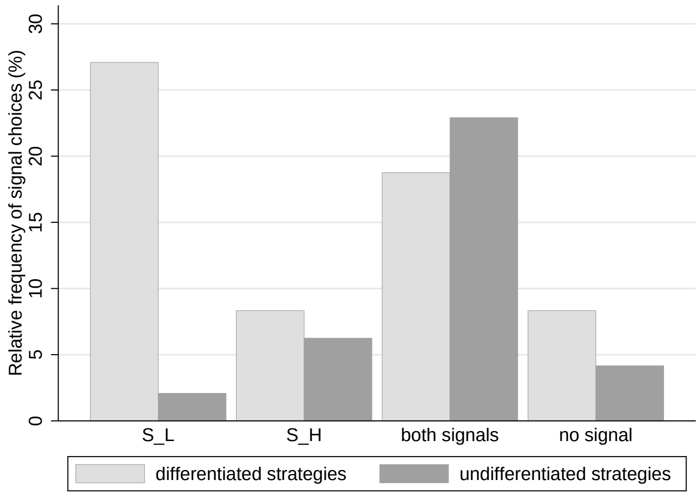


Figure 2: Distribution of information acquisition choices for subjects with differentiated and undifferentiated strategies.

to indicate the presence of objective compliers among the dictators. However, Figure 2 also clearly shows that acquiring both signals becomes the overwhelming choice for dictators with undifferentiated strategies. While this group of dictators is not a randomly selected control group[17], their behavior nevertheless suggests that getting as much information as possible is the default choice when no economic incentives are at stake. The motivation behind this can be compliance with an epistemic norm to acquire as much information as possible or, plainly, curiosity.[18] The marked and statistically significant (p=0.008, Mann-Whitney U test) jump in the choice of S_L when comparing dictators with and without economic stakes makes the evidence for subjective compliance in line with the alternative hypothesis 1 even stronger: strategic information acquisition is virtually non-existent for the

[17] Let us remind the reader, however, that the group of non-differentiating dictators is very heterogeneous and consists of pure 'egoists' as well as 'egalitarians'.

[18] There is some evidence, however, that the choice of acquiring both signals follows different processes when comparing those with and without economic incentives: decision times are, on average, considerably and weakly significantly longer for the former than the latter (25.9 vs 14.6 seconds, p=0.073, Mann-Whitney U test). This suggests that acquiring both signals is the result of a deliberative process for the group of dictators with differentiated strategies; curiosity alone might not be a good explanation for this group's choice of signals. Future research may try to further distinguish objective compliance from other motivations to acquire maximum information.

latter, but makes up the largest share of information choices for the former.

V. Summary and discussion

Depending on one's objective or subjective interpretation of norms, one can be subject to different forms of psychological costs when violating a norm. In line with Festinger (1957) and Konow (2000), we model these costs as *cognitive dissonance* that arises when acts do not match what the norm requires. Following the subjective interpretation, individuals experience cognitive dissonance if their beliefs and the norm together imply prescriptions they violate. When an agent is a subjective complier, she may strategically choose the sort of information that renders selfish actions morally appropriate. By contrast, an objective norm complier is always better off with more information about the state of the world, as this improves his chance to choose the morally appropriate action. These implications are similar in spirit to Rabin's (1995) distinction between moral preferences and moral constraints.[19]

Our model is widely applicable and can, for instance, explain strategic ignorance in dictator games, as found by Dana, Weber and Kuang (2007–DWK) and others. DWK interpret their results as evidence for an 'illusory preference for fairness'. They view ignorance as a desirable state that comes with lower normative demands or allows to hide (from others or themselves) their selfishness. In our model, subjective compliers do not strive for ignorance as such; instead they decide with their signal choice whether and which lotteries to play over their beliefs about the state of the world. The signals offered by DWK (and others) just weren't very attractive as normative demands could only increase. In addition, our model shows that preferences for fairness are not a necessary condition for unbiased, full information acquisition. Perhaps most surprisingly (though anticipated by Rabin 1995), we do not need to assume non-Bayesian belief updating. Our model locates the source of the 'moral wiggle room' in the strategic use of coarse-grained norms

[19]However, in contrast to Rabin (1995), in our model normative prescriptions always enter the utility function in the same way and all types of agents' distributional preferences are 'selfish'; yet, if agents are objective compliers, they will try their best to gather the type of information that allows them to perform the morally appropriate action.

(for example, if there is only one prescription for uncertainty), and not in biased beliefs (as suggested by Benabou and Tirole, 2011).

A new experimental test provides convincing evidence for strategic information acquisition in line with subjective norm compliance. By giving our subjects the chance to selectively avoid and acquire information, we find empirical support for our hypothesis that subjects engage in self-serving information acquisition (not just avoidance) to reduce norm compliance costs.

Our model and the supportive experimental results suggest that future research on 'moral wiggle room' should move away from focusing on 'ignorance' only. Furthermore, the recent focus on self-deception, fascinating as the results may be, might have diverted attention from a much simpler explanation: the imperfect incentives created by social norms, which some smart individuals are bound to exploit.

Acknowledgements

We would like to thank James Konow and Dirk Sliwka for extensive written comments. We are also grateful to Roberto Weber for an illuminating discussion. This paper was presented at the ESA European Conference 2012 in Cologne, the ESA 2013 World Meetings in Zurich, the EEA-ESEM 2013 in Gothenburg, and many smaller workshops; we owe thanks to all our audiences.

References

Babcock, L., Loewenstein, G., Issacharoff, S., and Camerer, C. (1995). Biased judgments of fairness in bargaining. *The American Economic Review*, 85(5):1337–1343.

Bardsley, N. (2007). Dictator game giving: altruism or artefact? *Experimental Economics*, 11(2):122–133.

Benabou, R. and Tirole, J. (2011). Identity, morals, and taboos: Beliefs as assets. *The Quarterly Journal of Economics*, 126(2):805–855.

Bicchieri, C. (2006). *The Grammar of Society: The Nature and Dynamics of Social Norms*. Cambridge University Press, Cambridge et al.

Bicchieri, C. and Chavez, A. K. (2013). Norm manipulation, norm evasion: Experimental evidence. *Economics and Philosophy*, 29(2):175–198.

Bicchieri, C. and Mercier, H. (2013). Self-serving biases and public justifications in trust games. *Synthese*, 190(5):909–922.

Bolton, G. and Ockenfels, A. (2000). ERC: A theory of equity, reciprocity, and competition. *American Economic Review*, 90(1):166–193.

Camerer, C. F. and Fehr, E. (2004). Measuring social norms and preferences using experimental games: A guide for social scientists. In Henrich, J., Boyd, R., Bowles, S., Camerer, C., Fehr, E., and Gintis, H., editors, *Foundations of Human Sociality - Experimental and Ethnographic Evidence from 15 Small-Scale Societies*, pages 55–95. Oxford University Press, New York.

Cherry, T., Frykblom, P., and Shogren, J. (2002). Hardnose the dictator. *The American Economic Review*, 92(4):1218–1221.

Chignell, A. (2013). The ethics of belief. In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy*. Spring 2013 edition. Available at: `plato.stanford.edu/archives/spr2013/entries/ethics-belief`.

Conrads, J. and Irlenbusch, B. (2013). Strategic ignorance in ultimatum bargaining. *Journal of Economic Behavior & Organization*, 92:104–115.

Dana, J., Weber, R. A., and Kuang, J. X. (2007). Exploiting moral wiggle room: experiments demonstrating an illusory preference for fairness. *Economic Theory*, 33(1):67–80.

Fehr, E. and Schmidt, K. M. (1999). A Theory of Fairness, Competition, and Cooperation. *Quarterly Journal of Economics*, 114(3):817–868.

Feldman, R. (2000). The ethics of belief. *Philosophy and Phenomenological Research*, 60(3):667–695.

Festinger, L. (1957). *A Theory of Cognitive Dissonance*. Stanford University Press, Palo Alto.

Fischbacher, U. (2007). z-tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics*, 10(2):171–178.

Greiner, B. (2004). An online recruitment system for economic experiments. In Kremer, K. and Macho, V., editors, *Forschung und wissenschaftliches Rechnen 2003*, number 63 in GWDG Bericht, pages 79–93. Ges. für Wiss. Datenverarbeitung, Göttingen.

Grossman, Z. (2010). Strategic ignorance and the robustness of social preferences. Available at: `http://escholarship.org/uc/item/60b93868`.

Hechter, M. and Opp, K.-D. (2001). Introduction. In Hechter, M. and Opp, K.-D., editors, *Social Norms*, pages xi–xx. Russell Sage, New York.

Jackson, F. (1991). Decision-theoretic consequentialism and the nearest and dearest objection. *Ethics*, 101(3):461–482.

Konow, J. (2000). Fair shares: Accountability and cognitive dissonance in allocation decisions. *The American Economic Review*, 90(4):1072–1091.

Kornblith, H. (1983). Justified belief and epistemically responsible action. *The Philosophical Review*, 92(1):33–48.

Krupka, E. and Weber, R. (2013). Identifying social norms using coordination games: Why does dictator game sharing vary? *Journal of the European Economic Association*, 11(3):495–524.

Larson, T. and Capra, C. (2009). Exploiting moral wiggle room: Illusory preference for fairness? a comment. *Judgment and Decision Making*, 4(6):467–474.

Levitt, S. D. and List, J. A. (2007). What Do Laboratory Experiments Measuring Social Preferences Reveal About the Real World ? *Journal of Economic Perspectives*, 21(2):153–174.

Matthey, A. and Regner, T. (2011). Do I Really Want to Know? A Cognitive Dissonance-Based Explanation of Other-Regarding Behavior. *Games*, 2(1):114–135.

Messick, D. M. and Sentis, K. P. (1979). Fairness and preference. *Journal of Experimental Social Psychology*, 15(4):418–434.

Rabin, M. (1995). Moral preferences, moral constraints, and self-serving biases. Available at: `http://escholarship.org/uc/item/97r6t5vf`.

Rodriguez-Lara, I. and Moreno-Garrido, L. (2012). Self-interest and fairness: self-serving choices of justice principles. *Experimental Economics*, 15:158–175.

Ross, W. D. (1939). *Foundations of Ethics*. Clarendon Press, Oxford.

Van der Weele, J. J. (2012). When ignorance is innocence: On information avoidance in moral dilemmas. *SSRN*. Available at: `http://ssrn.com/abstract=1844702`.

Varian, H. E. (1992). *Microeconomic Analysis*. W. W. Norton, New York, 3rd edition.

Zimmerman, M. (2008). *Living with uncertainty: the moral significance of ignorance*. Cambridge University Press Cambridge, Cambridge.

For online publication: Appendix

A. Proof that $\hat{u}$ is increasing and concave

To show that value function $\hat{u}'(\phi) > 0$, note that when applying the envelope theorem we find $\hat{u}'(\phi) = \frac{\partial u(y,\phi)}{\partial \phi}\Big|_{y=y^*} > 0$, because any increase in ϕ will decrease dissonance for keeping more and thus increase utility.

To see that $\hat{u}$ is concave, consider what happens if we increase ϕ to $\phi + \delta$ (with $\delta > 0$). The increase shifts $f(y - \phi)$ to the right by δ. If (against our assumptions) the slope of $v(y)$ were a constant $c = v'(y)$ (and $v''(y) = 0$), then $\hat{u}$ would increase by δc which would imply that $\hat{u}'(\phi)$ is constant and $\hat{u}''(\phi) = 0$. However, since $v''(y) < 0$, the increase of ϕ to $\phi + \delta$ will lead to smaller increases of $\hat{u}$ the further the dissonance function $f(y - \phi)$ is shifted to the right, so that $\hat{u}''(\phi) < 0$.

B. Proof of proposition 1

To show that objective compliers will acquire both signals on offer, we need to compare the expected utilities Eu_L, Eu_H, Eu_{LH} and Eu_0 of receiving signal S_L, S_H, both signals, or none:

$$
\begin{aligned}
Eu_L &= psu(y_L^*, \phi_L) + (1 - ps)v(y_U^*) - p(1 - s)f(y_U^*, \phi_L) - (1 - p)f(y_U^*, \phi_H) \\
Eu_H &= (1 - p)su(y_H^*, \phi_H) + (1 - s + ps)v(y_U^*) - (1 - p)(1 - s)f(y_U^*, \phi_H) - pf(y_U^*, \phi_L) \\
Eu_{LH} &= (1 - p)su(y_H^*, \phi_H) + psu(y_L^*, \phi_L) + (1 - s)v(y_U^*) - p(1 - s)f(y_U^*, \phi_L) \\
&\quad -(1 - p)(1 - s)f(y_U^*, \phi_H) \\
Eu_0 &= v(y_U^*) - pf(y_U^*, \phi_L) - (1 - p)f(y_U^*, \phi_H)
\end{aligned}
$$

We show that taking both signals is better than taking no signals by rearranging:

$$
\begin{aligned}
Eu_{LH} - Eu_0 &= (1-p)su(y_H^*, \phi_H) + psu(y_L^*, \phi_L) + (1-s)v(y_U^*) - p(1-s)f(y_U^*, \phi_L) \\
&\quad -(1-p)(1-s)f(y_U^*, \phi_H) - v(y_U^*) + pf(y_U^*, \phi_L) + (1-p)f(y_U^*, \phi_H) \\
&= (1-p)su(y_H^*, \phi_H) + psu(y_L^*, \phi_L) - sv(y_U^*) + psf(y_U^*, \phi_L) + s(1-p)f(y_U^*, \phi_H) \\
&= (1-p)s\left[u(y_H^*, \phi_H) - (v(y_U^*) - f(y_U^*, \phi_H))\right] + ps\left[(u(y_L^*, \phi_L) - (v(y_U^*) - f(y_U^*, \phi_L))\right] \\
&= \underbrace{(1-p)s\left[u(y_H^*, \phi_H) - u(y_U^*, \phi_H)\right]}_{>0} + \underbrace{ps\left[(u(y_L^*, \phi_L) - u(y_U^*, \phi_L)\right]}_{>0} > 0
\end{aligned}
$$

Both summands are positive since given ϕ_H (respectively: ϕ_L) utility is maximal at $\hat{u}(\phi_H) = u(y_H^*, \phi_H)$ (respectively: $\hat{u}(\phi_L) = u(y_L^*, \phi_L)$. Similarly, taking both signals is better than taking only S_H or S_L:

$$
\begin{aligned}
Eu_{LH} - Eu_H &= (1-p)su(y_H^*, \phi_H) + psu(y_L^*, \phi_L) + (1-s)v(y_U^*) - p(1-s)f(y_U^*, \phi_L) \\
&\quad -(1-p)(1-s)f(y_U^*, \phi_H) - (1-p)su(y_H^*, \phi_H) - (1-s+ps)v(y_U^*) \\
&\quad +(1-p)(1-s)f(y_U^*, \phi_H) + pf(y_U^*, \phi_L) \\
&= ps\left[u(y_L^*, \phi_L) - v(y_U^*) + f(y_U^*, \phi_L)\right] \\
&= ps\left[\underbrace{v(y_L^*) - v(y_U^*)}_{>0} + \underbrace{f(y_U^*, \phi_L) - f(y_L^*, \phi_L)}_{>0}\right] > 0
\end{aligned}
$$

$$
\begin{aligned}
Eu_{LH} - Eu_L &= (1-p)su(y_H^*, \phi_H) + psu(y_L^*, \phi_L) + (1-s)v(y_U^*) - p(1-s)f(y_U^*, \phi_L) \\
&\quad -(1-p)(1-s)f(y_U^*, \phi_H) - psu(y_L^*, \phi_L) - (1-ps)v(y_U^*) \\
&\quad +p(1-s)f(y_U^*, \phi_L) + (1-p)f(y_U^*, \phi_H) \\
&= (1-p)su(y_H^*, \phi_H) + (p-1)sv(y_U^*) + (1-p)sf(y_U^*, \phi_H) \\
&= (1-p)sv(y_H^*) - (1-p)sf(y_H^*, \phi_H) + (p-1)sv(y_U^*) + (1-p)sf(y_U^*, \phi_H) \\
&= (1-p)s\left[v(y_H^*) - f(y_H^*, \phi_H) - (v(y_U^*) - f(y_U^*, \phi_H))\right] \\
&= (1-p)s\left[u(y_H^*, \phi_H) - u(y_U^*, \phi_H)\right] > 0
\end{aligned}
$$

Therefore taking both signals is the most preferred choice.

C. Proof of Proposition 2

As in the previous proof, we compute the expected utilities of acquiring signals S_L, S_H, both signals, or none. Note that, by assumption, receiving a null signal always yields $\phi_p = \phi_U$. This simplifies the expected utilities of the respective signal acquisitions:

$$
\begin{aligned}
Eu_L &= ps\hat{u}(\phi_L) + (1-ps)\hat{u}(\phi_U) \\
Eu_H &= (1-p)s\hat{u}(\phi_H) + (1-s+ps)\hat{u}(\phi_U) \\
Eu_{LH} &= ps\hat{u}(\phi_L) + (1-p)s\hat{u}(\phi_H) + (1-s)\hat{u}(\phi_U) \\
Eu_0 &= \hat{u}(\phi_U)
\end{aligned}
$$

Since $\phi_L > \phi_U > \phi_H$ it follows that $\hat{u}(\phi_L) > \hat{u}(\phi_U) > \hat{u}(\phi_H)$. It is now obvious that $Eu_L > Eu_H$, $Eu_L > Eu_{LH}$ and $Eu_L > Eu_0$. Therefore acquiring only S_L is the best choice.

D. Proof of Proposition 3

When no signal is acquired, ϕ_p remains unchanged at $p\phi_L + (1-p)\phi_H$ between ϕ_L and ϕ_H. Acquiring signal S_L either leads to ϕ_L with probability ps or to an increase in ϕ_p after the Bayesian update to $p' = (1-s)p/\left((1-s)p + (1-p)\right)$. Acquiring signal S_H either leads to ϕ_H with probability $(1-p)s$ or to a decrease in ϕ_p after the Bayesian update to $p' = p/(p + (1-s)(1-p))$. When acquiring both signals, ϕ_p can go to ϕ_L, ϕ_H, or stay the same (in case of two null signals). As can be checked easily, the expected value $E\phi_p$ for all these (fair) lotteries is just the initial $\phi_p = p\phi_L + (1-p)\phi_H$. However, since $\hat{u}$ is a strictly concave function, $E\hat{u}(\phi_p) < \hat{u}(E\phi_p)$ because of Jensen's Inequality.[20] Therefore, the expected utility of acquiring a signal is always lower than the utility obtained when not acquiring a signal, and thus not acquiring a signal is always preferred.

E. Details of experimental design

A. Creation of social norms about receiver-entitlements

In order to test for subjective and objective norm compliance, we create (or make salient) a set of social norms of entitlement whose violation would create cognitive dissonance. We take several measures to achieve this:

[20] Let X be a nondegenerate random variable and $f(X)$ be a strictly concave function of this random variable. Then $Ef(X) < f(EX)$ (see e.g. Varian, 1992, 182).

First, the receivers' entitlements were manipulated by linking the performance in the quiz to the chance of a dictator-receiver pair winning the bonus of 20 Euros. Subjects are informed that a low-performing receiver does not contribute to the pair's chance of winning the bonus, whereas a high-performing receiver contributes as much as the dictator to the pair's winning chances. In the experiment, this was represented by high-performers contributing a 'winner lot', but low-performers only contributing a 'blank lot' to each pair's bonus draw. The existence of the 20 Euro fund would be determined by a draw of one lot from the pair playing, with a winner lot providing funds, a blank lot no funds. A pair of two high-performing players would always have 2 winner lots, therefore always draw a winner lot and consequently always have 20 Euros to distribute. A pair of one high-performing and one low-performing player would win funds with probability 0.5.

Second, to turn this wedge in entitlements into a set of social norms on dictator-giving, we aim at creating social norms, i.e., shared beliefs of what others expect one ought to do (Bicchieri, 2006). To that effect, the subjects in the main study are informed about the results of an earlier survey regarding the normatively appropriate behavior in the experiment. We reported, truthfully, that the mode of respondents in the earlier survey thought that a dictator should give 10 (out of 20) Euros to a high performer, and that the mode of respondents thought a dictator should give 5 (out of 20) Euro to a low performer. More precisely, in this survey, 18 of 26 subjects judged 10 to be the appropriate payment to a high entitlement receiver, while 12 of 26 though that a payment of 5 was appropriate for a receiver with low entitlement (and 5 was the mode for that question). As these numbers suggest, the variance in the survey regarding the latter question was significantly higher. This is unsurprising because a dictator playing against a receiver with high entitlement is in a clear situation of symmetry with the receiver: they are both in the same performance class, and have equally contributed to the funds for distribution, strongly suggesting an equal distribution. By contrast, it is less obvious how much a dictator (who always has a high entitlement) should give a receiver with low entitlement. Reassuringly, 18 of 26 respondents in the preliminary study stated that a lower amount for a low performer is appropriate, in line with our expectations.

Third, all subjects are made aware that each receiver will be informed about the matched dictator's transfer decisions in order to make receiver expectations salient.

Fourth, to assess and further strengthen the relevance of the social norms, we asked all subjects before performing the quiz how they valued, on a scale from 1 to 4, the correctness and personal relevance of the announced normative expectations elicited in the prior norm induction session. Precisely, subjects were asked to rate for both modes of normative expectations to what extent they consider the 26 students' opinion in the previous survey a) to be 'right' and b) to be 'important' for them personally.

Based on the model's assumption of ϕ as exogenous parameters set by social norms, the fair points are estimated as $\phi_H = 10$ and $\phi_L = 5$.

This fourfold treatment makes the mentioned fair points salient, communicates a normative expectation as to what one should do, and, since all subjects are informed, makes it *ceteris paribus* more likely that other dictators will also comply, creating or reinforcing a social norm of equity. Compliance can plausibly be conceived in a subjective or an objective sense, as the wording did not refer to the epistemic state of the dictator.

Subsequently, all subjects played a competitive quiz consisting of 15 questions in the style of the well-known TV-show 'Who Wants to be a Millionaire' under time pressure. Correct answers were rewarded with positive points, incorrect or missing answers with point deductions. Points increased with difficulty. The performers in the top three quartiles were assigned the name 'Gold Quiz-Players' (high entitlement), the last quartile 'Silver Quiz-Players' (low entitlement). In each session, all 16 dictators were drawn from the 24 Gold Quiz-Players, while the receivers were constituted by the remaining 8 Gold and 8 Silver Quiz-Players (see assignment as Screen 1 in Appendix G).

B. Dictators' transfer strategy

The dictators' giving strategy was elicited in order to have an empirical estimation of the individual responsiveness to the entitlement norm. At that stage dictators know that they are a high performer themselves, but they do not know which type of receiver they will face. Dictators bindingly state two amounts: the amount they give to the receiver if they learn the receiver is a high performer (x_H), and the amount they give if they learn the receiver is a low performer (x_L). Dictators do not directly choose the amount to be transferred if they do not learn which type the receiver is. Instead, this amount (x_U) is calculated as $(x_H + x_L)/2$, of which dictators are explicitly made aware of (see Screen 2 in Appendix G). We refrained from letting the dictators choose x_U in order to create a coarse-grained normative system and test proposition 2. Thus, we inform the subjects in the written instructions that giving $(x_H + x_L)/2$ under uncertainty is appropriate because this is equi-distant from x_H and x_L. This has the added benefit that, in absolute terms, the monetary consequences of acquiring either signal are the same. When entering the strategy, the dictators do not know under which circumstances the information about their receiver might become available. The written instructions informed the subjects that more information on this would become available onscreen. This seems to have successfully inhibited questions from subjects on this delicate element as no subject asked the experimenter about the conditions under which the dictator would know which type of receiver she would be matched with.

C. Information acquisition and payoffs

The optional opportunity to acquire information about the receiver is the crucial design element to test for objective and subjective norm compliance. After each dictator has been matched with a receiver, they have the options (i) to receive no signal, (ii) signal S_L, (iii) signal S_H or (iv) both signals. The prior probability that the state is L is $p = \frac{1}{2}$ and the probability of removing uncertainty when choosing the 'correct' signal is $s = \frac{1}{2}$. The four options are implemented by presenting dictators with four envelopes on their computer screens (see Screens 3–5 in Appendix G). We call two of these envelopes 'gold envelopes', the other two 'silver envelopes'. Exactly one of these four envelopes contains (represented electronically) the lot of the receiver (which in turn reveals the performance of the receiver). If the receiver has a winner-lot, it will always be in one of the two gold-envelopes, while a blank lot will always be in one of the two silver-envelopes. The dictator can now choose to open one gold-envelope; open one silver-envelope; open one gold- and one silver-envelope; or open no envelope. We subsequently (electronically) ‘open’ the chosen envelopes and show the results. If the dictators find the receiver lot, they have certainty about the receiver’s type. If they do not find the lot, uncertainty about the type remains. Opening a ‘silver’ envelope means receiving S_L and opening a ‘gold’ envelope receiving S_H.

As the dictator’s contribution strategy depends on what the dictator learns about the type of the receiver, the information received determines which amount will be transferred if the bonus is won in the final stage. Dictators are informed about the monetary consequences of the information acquisition. In fact, dictators have to run through a series of test question in order to minimize errors on their part. Note that receiving only S_L (only S_H) never leads to certainty about the receiver being a high performer (low performer); x_H (x_L) will therefore never be transferred. If the bonus is won, the dictator receives y_H if he has found a ‘winner lot’ (and therefore has learned that the receiver is a high performer), y_L if he has found a ‘blank lot’ (and has therefore learned that the receiver is a low performer) and y_U if he has not found the receiver’s lot (and therefore has not learned whether the receiver is a low or a high performer).[21]

F. Written Instructions for All Participants

[21] We deliberately left open whether the receivers would be informed about the signal-choice of the dictators (in fact, receivers are not informed), as we did not want the dictators to wonder about information acquisition norms or create an experimenter demand effect towards acquiring or not acquiring signals.

Experiment Instructions (translated)

Thank you for agreeing to participate in this decision-making experiment. Please read this description of the experiment carefully. For the entire duration of the experiment any communication with other experiment participants is prohibited. Please turn off your cell phones now. It is a mandatory requirement for participation in this experiment to comply with these rules.

If some points remain unclear, please read the experiment instructions again. For any remaining questions, please raise your hand. We will come to your desk and answer your questions in person. All important details of the experiment are also shown on screen. In addition, test questions help to ensure that all participants understand the experiment correctly.

You can earn money in this experiment. You will always receive a compensation of 5 Euros for participation in today's experiment. For the completion of questionnaires at the end of the experiment you will receive an additional 2 Euros. How to earn any further money will be explained in these instructions. Your data and decisions are anonymous. Your answers and decisions cannot be linked to your identity and no person-identifying data will be saved.

Part 1: Quiz

All participants take part in a knowledge quiz.

After the completion of the quiz, participants are divided into two groups:

- "Gold Quiz-Players": the 75% of all participants with the highest relative quiz performance
- "Silver Quiz-Players": the 25% with the lowest relative quiz performance
- You will not be informed of your individual performance in the quiz.
- Gold Quiz-Players get a **winner lot** for their performance. Silver Quiz-Players get a **blank lot**. These lots are used in a 20-Euros-draw, which is explained below.
-

Part 2: 20-Euros-draw

- Two players each form a participant-pair and take part in the 20-Euros-draw. A pair always consists of an allocator and a recipient.
- Which receiver and which allocator form a participant-pair is determined at random.
- Both the allocator and the recipient bring their lot from the quiz to the draw.
- All allocators are recruited from the Gold Quiz-Players and therefore always bring a winner lot to the draw.
- Half of the recipients are recruited from the Gold Quiz-Players and half from the Silver Quiz-Players. So half of the recipients bring a winner lot and half bring a blank lot to the draw.
- Neither the recipient nor the allocator knows if the recipient brings a winner lot or a blank lot to the draw. However, depending on the further course of the experiment, this information might be revealed to you. More information on this will be available on the screen.
- One of the two lots provided by the participant-pair will be drawn.
- If a winner lot is drawn: the allocator has to distribute the 20 Euros between himself and the recipient.
- If a blank lot is drawn: the 20 Euros are not won.

Task for the Allocator:

- The allocator decides on the allocation of the 20 Euros before he gets matched with a recipient. At this point it is still unknown which type of lot the recipient will bring into the lottery and if the 20 Euros will be won.

- The allocator decides on the allocation **in advance and bindingly**. **The distribution depends on the information the participant-pair will receive about the lot of the recipient:**
    - The allocator determines in advance the amount that will be transferred to the recipient in case the pair learns that the recipient is a **Gold Quiz-Player** and brings a **winner lot** into the lottery. This amount is called **GOLD-transfer**.
    - The allocator determines in advance the amount that will be transferred to the recipient in case the pair learns that the recipient is a **Silver Quiz-Player** and brings a **blank lot** into the lottery. This amount is called **SILVER-transfer**.
- These amounts determine:
    - **GOLD/SILVER-transfer**: the amount that will be transferred to the recipient if the pair **does not learn** which lot the recipient brings to the draw.
        - **GOLD/SILVER-transfer = average of the GOLD-transfer and the SILVER-transfer.** This amount is of equal distance to the GOLD-transfer and the SILVER-transfer. This amount is chosen because it is either possible that the recipients brings a winner lot, or that he brings a blank lot.
- In case of winning the 20 Euros:
    - <u>The recipient gets</u>, depending on the three cases, <u>either the GOLD-transfer, or the SILVER-transfer, or the GOLD/SILVER-transfer.</u>
    - <u>The allocator gets 20 Euros minus the corresponding transfer. So either 20 minus the GOLD-transfer, or minus the SILVER-transfer, or minus the GOLD/SILVER-transfer.</u>

Information for the recipient at the end of the experiment:

- The amounts of the GOLD-transfer, SILVER-transfer, and GOLD/SILVER-transfer, which were determined by the allocator;
- If the 20 Euros have been won;
- Which kind of information the allocator got regarding the recipient’s lot;
- In case of winning the 20 Euros, whether the GOLD-transfer, the SILVER-transfer, or the GOLD/SILVER-transfer was transferred to him.

Chronological sequence:

1. Quiz
2. Players with the highest relative performance become Gold Quiz-Player (with a winner lot) and players with the relatively lowest quiz performance become Silver Quiz-Player (with a blank lot).
3. Division into allocators and recipients. Allocators are Gold Quiz-Player. Recipients are, distributed randomly, half Gold Quiz-Players and half Silver Quiz-Players.
4. Allocators decide on GOLD-transfer and SILVER-transfer. From these amounts the GOLD-SILVER-transfer results (The average of GOLD-transfer and SILVER-transfer).
5. Each allocator is matched with a recipient.
6. The participant-pair might or might not get the information about the type of the recipient’s lot.
7. The draw decides whether the pair wins 20 Euros or not. Exactly one of the two lots of the players is picked at random.
8. In case of winning: Distribution corresponding to allocator's transfer decisions.
9. Allocators and recipients get information about the results of the experiment and their own payoff.

This sequence is represented graphically below:

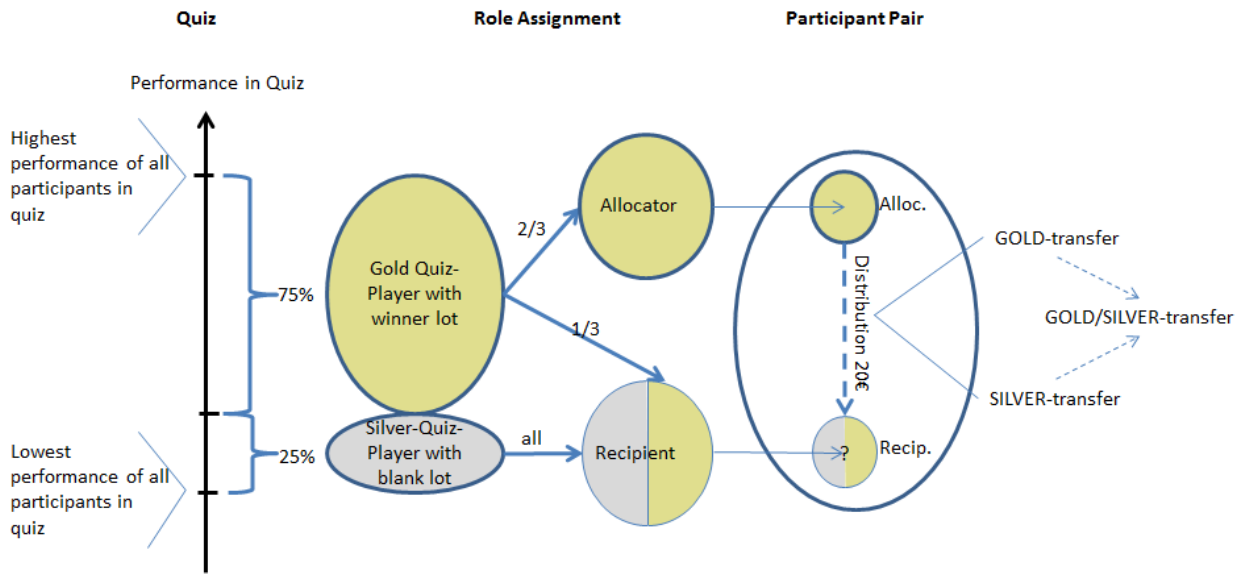


Figure 1

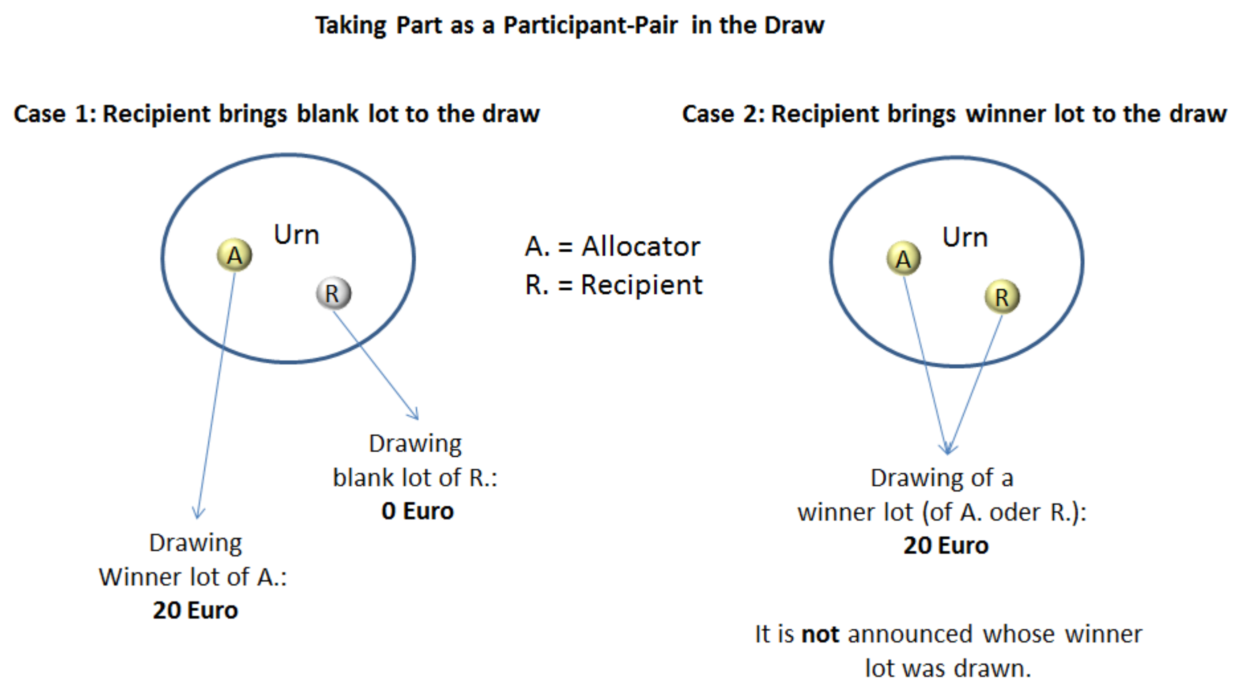
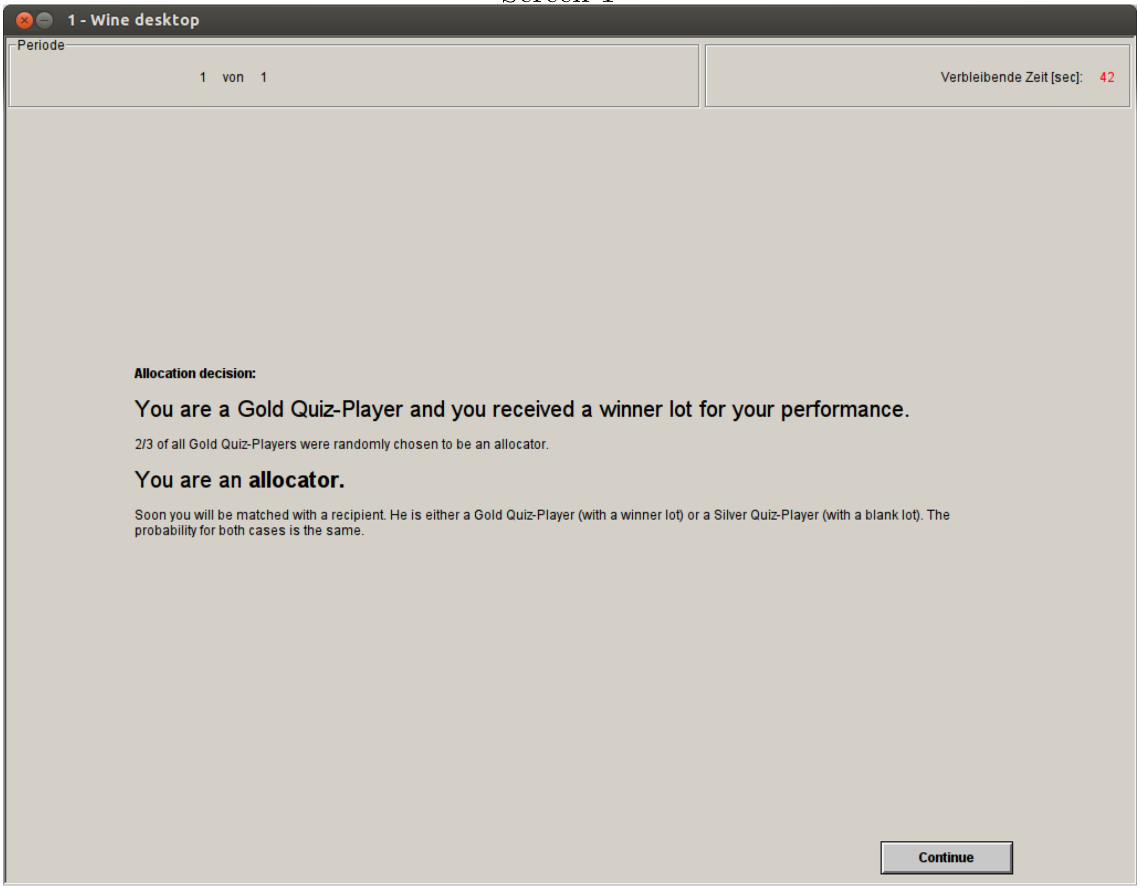


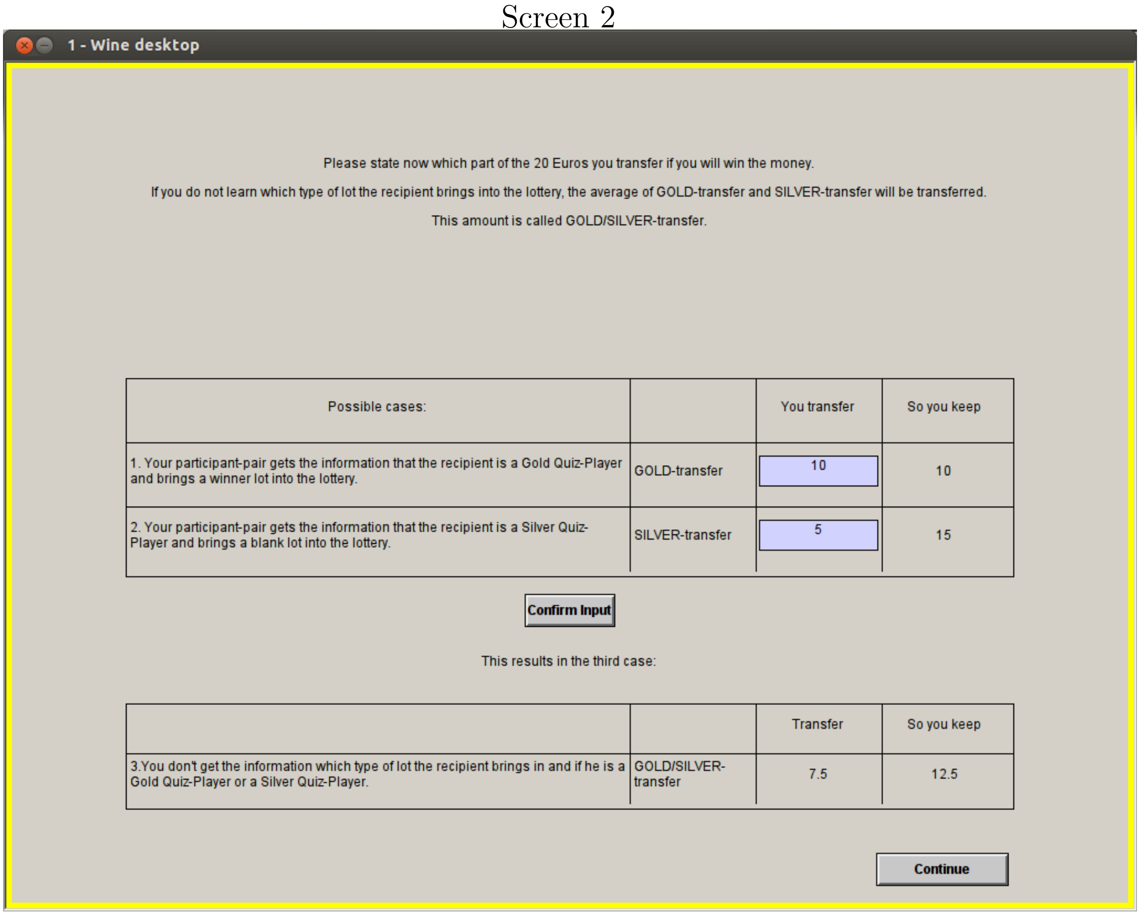
Figure 2

G. Important Treatment Screens (Translated)

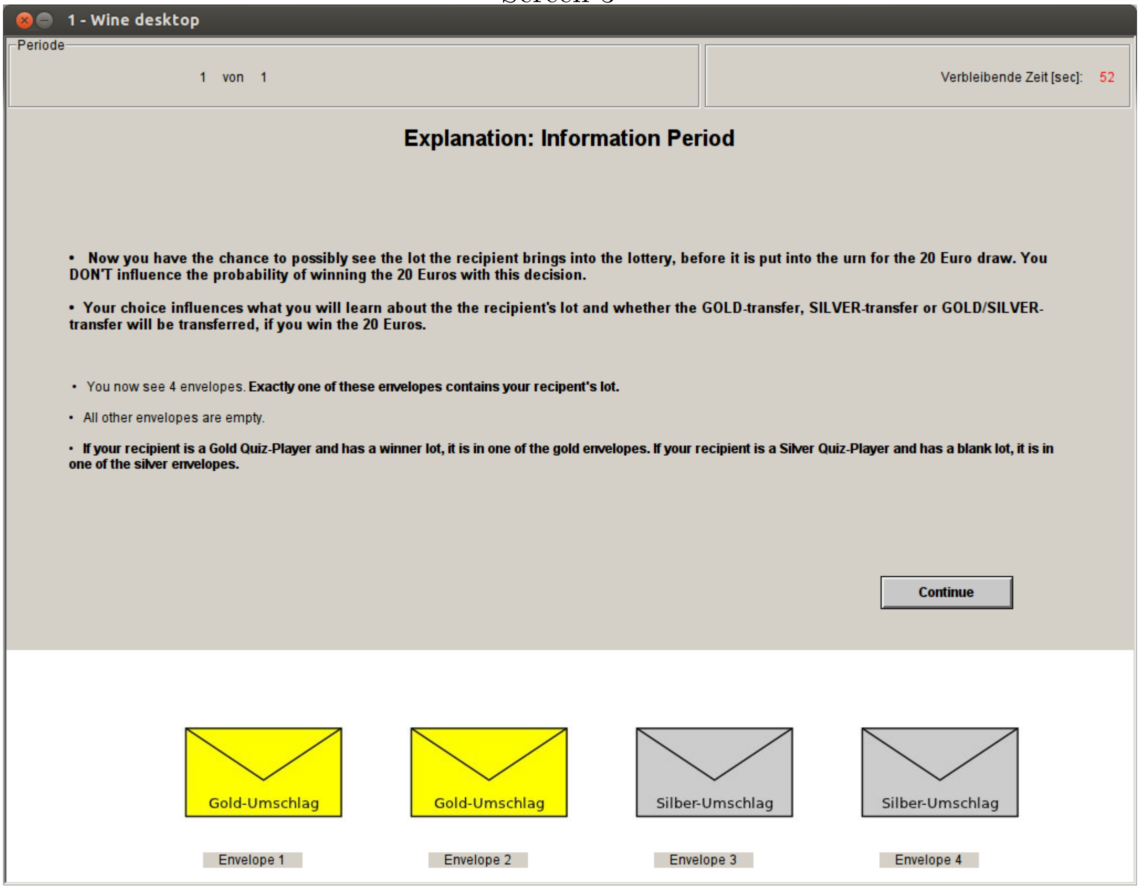
Screen 1



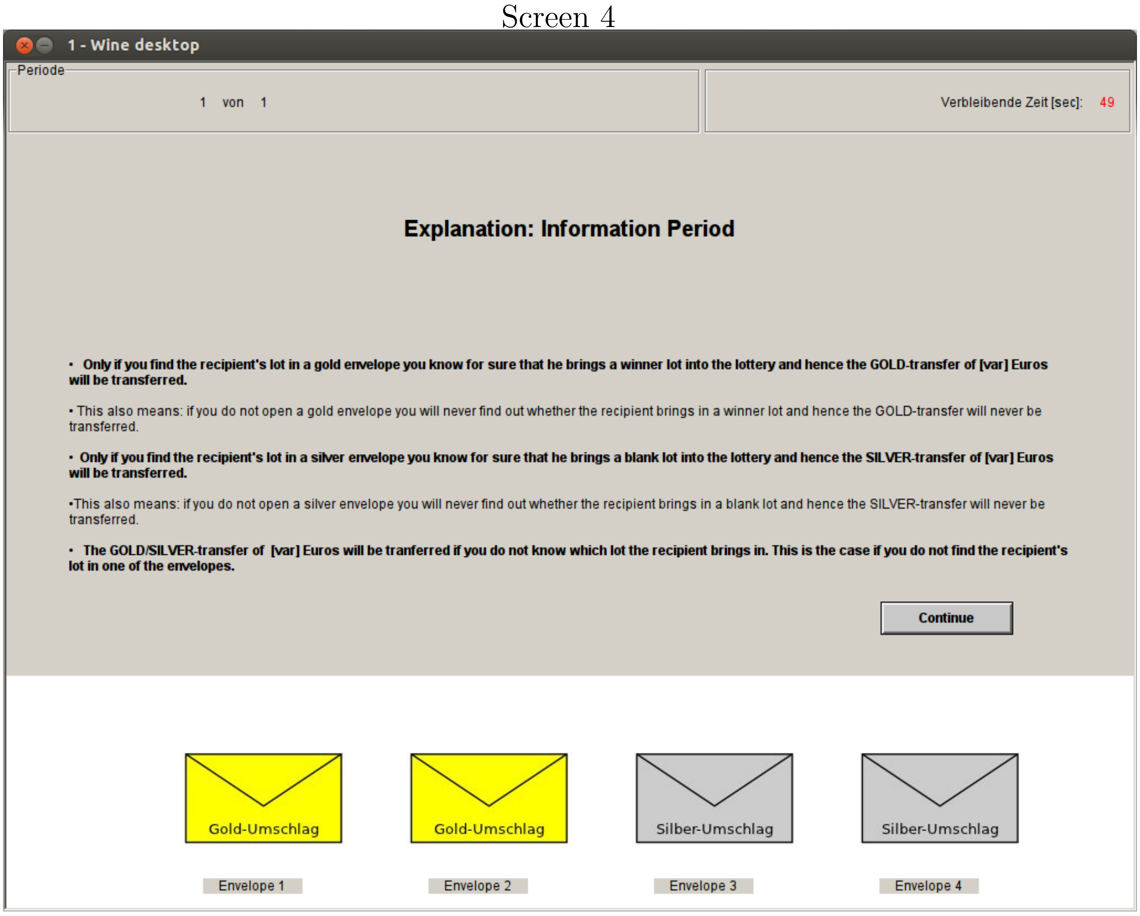
Screen 2



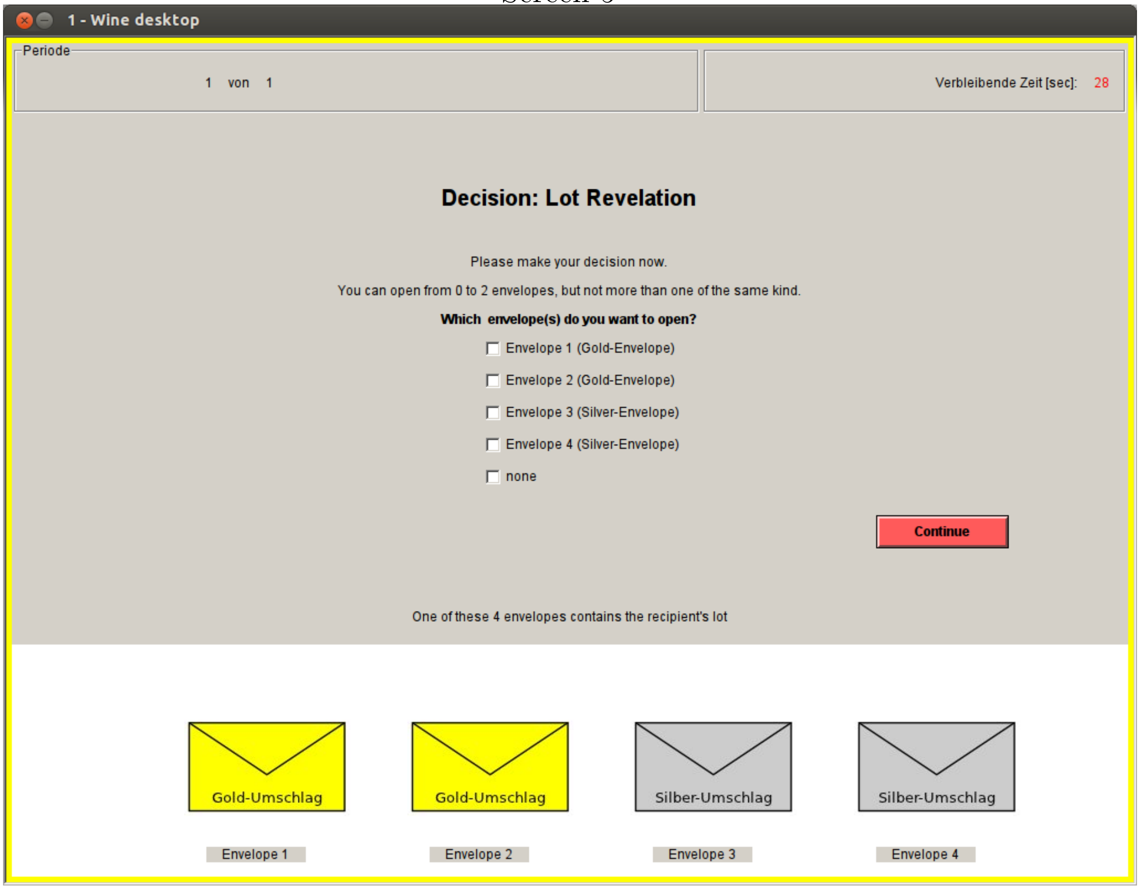
Screen 3



Screen 4



Screen 5


1 - Wine desktop
Periode
1 von 1
Verbleibende Zeit [sec]: 28
Decision: Lot Revelation
Please make your decision now.
You can open from 0 to 2 envelopes, but not more than one of the same kind.
Which envelope(s) do you want to open?
Envelope 1 (Gold-Envelope)
Envelope 2 (Gold-Envelope)
Envelope 3 (Silver-Envelope)
Envelope 4 (Silver-Envelope)
none
Continue
One of these 4 envelopes contains the recipient's lot
Gold-Umschlag
Gold-Umschlag
Silber-Umschlag
Silber-Umschlag
Envelope 1
Envelope 2
Envelope 3
Envelope 4