Subjective and Objective Compliance: Choosing What to Know in Normative Choice Situations

Preliminary — Comments Welcome

Kai Spiekermann*, Arne Weiss†

September 19, 2013

Abstract

We propose a cognitive-dissonance model of norm compliance to identify conditions for strategic information acquisition. The model explains such strategic choices by distinguishing 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 epistemic state. The former seek as much information as possible; the latter acquire only information that lowers, in expected terms, normative demands. The source of such 'moral wiggle room' is not belief manipulation, but the way normative prescriptions respond to signals about the state of the world. 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 about the social norm that applies to this decision: 'if an employee performed well, he ought to be

*London School of Economics, Department of Government, Connaught House, CON.517, London WC2A 2AE

†Corresponding author, Staatswissenschaftliches Seminar für Experimentelle Wirtschafts- und Verhaltensforschung, Aachener Straße 217, D-50931 Cologne, Germany. Email: `arne.weiss@uni-koeln.de`

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. First, social norms typically leave a fairly large room for interpretation (Hechter and Opp, 2001); the employer and employee will typically disagree in predictable ways. Second, they might disagree about the facts pertaining to the employee's performance. In their discussion, they will most likely evoke arguments of fairness that would favor their own narrow self-interest, and they would appeal to relevant facts in a selective, self-serving manner (Messick and Sentis, 1979; Babcock et al., 1995). In this paper we show 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 (2012), Bicchieri and Mercier (2012), and Rodriguez-Lara and Moreno-Garrido (2012).

Dana, Weber and Kuang (2007, henceforth DWK) find evidence for the relevance of factual uncertainty, demonstrating that 'strategic ignorance' over the negative externalities of one's behavior can open 'moral wiggle room' (see also Krupka and Weber, 2008; Grossman, 2010; Van der Weele, 2012). They infer that many dictators do not have a preference for the fair outcome but feel compelled to bring it about under transparent conditions. Strikingly, such dictators show selfish behaviour even though they can resolve the uncertainty they face virtually costlessly. In fact, most dictators who choose 'selfishly' decide to remain ignorant, suggesting that these dictators 'have an illusionary preference for fairness' and 'dislike appearing unfair' (DWK, p. 67). DWK's results have important ramifications, both for the widespread use of dictator-game giving as a measure of 'pro-social preferences' 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 an 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 intransparency, 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 intransparency 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 way if two conditions are met. First, those individuals interpret the demands of a norm in a way that opens the door to strategic behaviour: they are 'subjective norm compliers', that is, they take normative demands as a function of their beliefs. Second, the mapping from beliefs to prescriptions must be 'coarse' such 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 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 subjective and objective compliance for the incentives to resolve or maintain factual uncertainty, we introduce a new set-up 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 relate the model to 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 entirely independent from the epistemic state of the agent. 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 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 addressed 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.[1] 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 exculpable 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.[2]

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

[1] In ethical theory, the core question is whether moral obligations (or moral blame) do depend on knowledge, for epistemology 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) and Feldman (2000) for the latter.

[2] 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.

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, individuals experience cognitive dissonance if the conjunction of the state of the world and the norm implies (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 employer to determine an 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 employer's own norms), or did not perform well and does not deserve any bonus. If the employer is an objective norm complier, she will suffer from expected dissonance based on her Bayesian beliefs, 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 state-conditional cognitive dissonance because it is possible that the bonus does not match the employer's norm of what is fair in the *actual* (but unknown) state. For example, giving a 'compromise' bonus of $B/2$ leads to expected dissonance: the bonus is either too high in case the employee performed poorly, or too low in case the employer 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 is one of uncertainty, and therefore does not suffer dissonance when giving $B/2$, in stark contrast to the objective norm complier.

Under certainty, subjective and objective 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 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 information if it 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 the simplest 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.

An objective norm complier, who wants to get it right, will choose both types of signals

		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.

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 riskless 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 and give B/2 (as he would without aquiring signals).

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 states after acquiring the signals. In the example of farming, a subjective norm complier might follow a norm that forbids buying conventional dairy products if there is sufficient evidence that alternative farming methods would improve animal welfare. He will then not read articles in a Greenpeace magazine on dairy farming, since he has sufficient prior knowledge that its content can only change his beliefs towards an unattractive state (banning the consumption of cheap products). Since his norm permits buying conventional farming products if the benefits of the 'responsible' alternatives are questionable, he may instead pay active attention to statements from the conventional farming association.

II. Formal model

In order to introduce a formal distinction between subjective and objective 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 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.[3] 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 natural) states are fixed exogenous variables for the dictator.[4] In the setting we consider, we assume the ϕs to be determined by social norms in the form of shared expectations

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

[4] 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, 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).

and normative beliefs of a relevant reference group.[5]

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:

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

The behavior of a dictator is therefore determined by her 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). When deciding on an allocation in state H, the dictator reaches the maximum utility u_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 u_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 $y_L^* < \bar{y}$ and $y_L^* > y_H^*$.

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 $u_L^* > u_H^*$.

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 to a required action, in our case the amount one is permitted to keep. For the subjective norm complier, this

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

function can take different forms. The most 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 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 Bayesian updating).

The model, however, also allows for different mappings from epistemic states to prescriptions. On the other end of the spectrum, it could be assumed 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.[6]

But regardless of how precisely subjective compliers determine their fair amounts ϕ_p to give under uncertainty, the maximum utility under uncertainty is $u_p^* = v(y_p^*) - f(y_p^*, \phi_p)$, with $y_p^* = \underset{y}{\text{argmax}}(u(y, \phi_p))$.

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 $\text{s} = \Pr(S_L = L | \omega = L)$. Similarly, when ob-

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

taining 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: $s = \Pr(S_H = H|\omega = H)$. However, the dictator receives a 'null' signal with probability $1-s$ for each signal acquired, in case of which 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.

A. Propositions

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

Proposition 1. *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 $sp(u_L^* - u(y_U^*, \phi_L))$ by acquiring S_L and of $s(1-p)(u_H^* - u(y_U^*, \phi_H))$ 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. *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 $\hat{u}(\phi_L) > \hat{u}(\phi_U) > \hat{u}(\phi_H)$. 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 $\hat{u}(\phi_H)$.[7]

[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.

Proposition 3. *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, 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.

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. Theoretical contribution and related literature

DWK (2007) provide experimental evidence that agents often prefer to remain ignorant about the negative consequences of their selfish acts, using the 'moral wiggle room' that uncertainty provides. We will show that our model offers a parsimonious explanation of DWK's results and that it generalizes to settings beyond strategic ignorance.

DWK 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). 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). A dominates B in terms of dictator payoff, but it comes with a certain (respectively: possible) negative externality for the receiver in the baseline (hidden information) treatment. 74% of dictators choose the fair and efficient option B in the baseline treatment. In the hidden information treatment it is by default unclear which of the two actions hurts the receiver, but, importantly, the dictator can resolve this uncertainty costlessly by clicking on a button. Almost half of the these dictators, however, deliberately remain ignorant and 85% of those choose the payoff-dominant action A, even though there is a chance of $\frac{1}{2}$ that this imposes a severe negative externality on the receiver.

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. They also offer the competing explanation that prescriptions for fair behavior might be stronger in transparent environments than in intransparent environments. More generally, DWK and follow-up research (Larson and Capra, 2009; Matthey and Regner, 2011; Van der Weele, 2012) view ignorance as a particularly desirable state for dictators, as their selfish behavior can be hidden (from others or themselves).

We agree with DWK that ignorance might come with less demanding prescriptions than certainty. In light of our model, the DWK subjects that remain uncertain and choose selfishly are subjective compliers, following a norm that does not make strong prescriptions for uncertainty. However, we are less convinced that their behavior is motivated by the desire to hide selfishness or by an illusionary preference for fairness. In fact, if our model is correct, 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 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 sizeable fraction of dictators in DWK who choose to reveal: they can be modelled as *objective* norm compliers who, unlike the subjective compliers, 'want to get it right' and are therefore better off with more 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. However, the next section is devoted to an experiment showing that dictators can also actively seek information that justifies relatively selfish choices and ignore information demanding more equitable choices in a environment with clearly specified social norms. Sometimes getting some

information, but not all information, is the best way to 'wiggle'.

IV. Experimental test

Our new experimental setup is designed to test whether individuals strategically *seek* (as well as avoid) normatively relevant information. The theoretical underpinning is provided by proposition 2, which is a generalization of claims pertaining to strategic ignorance, as studied by DWK and others. In order to test proposition 2, our experimental design implements a coarse-grained norm. It 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 **??**). 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 behaviour. 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 continously as a function of their belief about the receiver type. Since a continuous reaction is impossible, a coarse-grained norm is induced or reinforced.

Fourth, after entering the strategy information, the dictators are paired with equal probability with either a high performer or a low performer receiver, but they do not learn by default which type 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 (un)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 underserving 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 experience with laboratory experiments (between 1 and 6 previous experiments at CLER). 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. Subjects took part in only one session and assumed only one role. General instructions about the experiment were provided on paper (see appendix **??**). 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 in the main sessions would give[8], under certainty, significantly more to a high performer than to a low performer ($p<0.001$, Wilcoxon signed-rank test).[9] As can be seen from Figure 1 (note that the size of the bubbles in Figure 1 indicates the relative frequency of the coordinate), no dictator gives more to a low performer than to a high performer. On average, dictators give a substantial amount to both types of receivers. The mean of contribution to low performers is about 3.7 and to high performers 6.6.

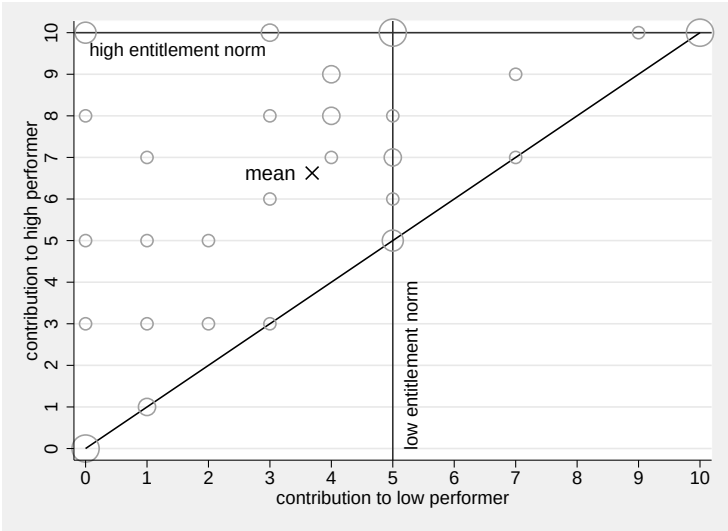

Figure 1: Dictator transfer strategies.

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 gray. The signal choice of one self-reportedly confused subject[10] is excluded, which leaves us with 30 dictators with differentiated and 17 with undifferentiated strategies. The latter dictators are apparently

[8] Here we refer to and analyze 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.

[9] All statistical tests are two-sided.

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

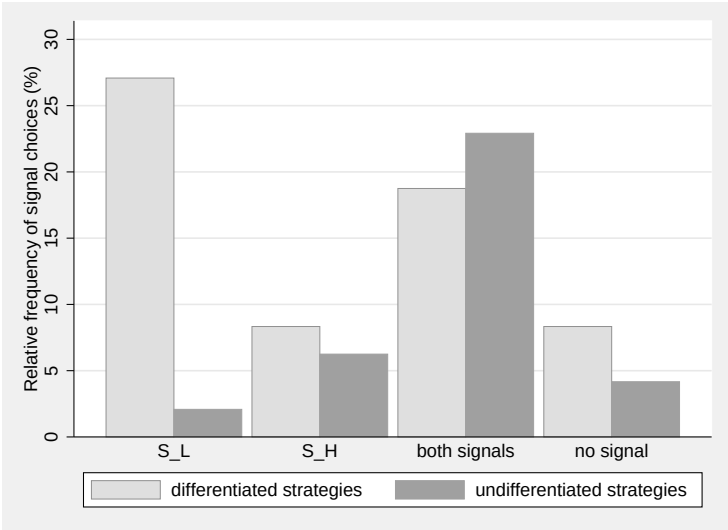

Figure 2: Distribution of information acquisition choices for subjects with differentiated and undifferentiated strategies.

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'[11] would), or because they follow a different norm, for instance a norm prescribing an egalitarian distribution.[12] We asked our subjects in a post-experimental questionaire about the motives of their giving decision. Not all subjects stated clear reasons, but 6 subjects mentioned considerations of equality, 4 selfishness, and 1 both equality and selfishness. Among these 11 subjects, 9 had completely undifferentiated giving strategies, the other 2 gave only 2 Euros more in case they learn they are paired with a high performer. By contrast, 21 subjects mentioned either entitlement/desert or norms as motives (while neither mentioning equality nor selfishness) and those subjects 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

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

[12] An egalitarian norm can be modelled by setting $\phi_H = \phi_L = 10$

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 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[13], 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.[14] This 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 latter, but makes up the largest share of information choices for the former.

V. Summary and discussion

Depending on one's subjective or objective 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 more morally appropriate, provided that norms are

[13] 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'.

[14] 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.

coarse-grained; for example, if there is only one prescription for all levels of uncertainty. 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. In Rabin's model, agents either genuinely care about the consequences of their acts (preference-agents) or treat norms as constraints they would rather circumvent (rule-agents). The distinction becomes behaviorally relevant when agents have the opportunity to shape their beliefs. Preference-agents are typically better off with more information, whereas rule-agents shape their beliefs strategically in order to sidestep norms that prohibit selfish acts. In contrast to Rabin (1995), in our model norms 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.

DWK interpret their results on strategic ignorance 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 epistemic states. The lotteries for information offered by DWK (and others) just weren't very attractive. 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, as our model locates the source of the 'moral wiggle room' in the strategic use of coarse-grained norms, and not in biased beliefs.

One caveat applies: if there were powerful epistemic norms prescribing to acquire all relevant information, subjective and objective norm compliers would show similar behavior, even though their underlying reasoning would be quite different. However, epistemic norms are likely weaker than 'direct' norms on behavior: experimental evidence by Krupka and Weber (2008) shows that ignorance about the negative consequences of one's actions, even when the ignorance is easily avoidable, serves as an acceptable excuse for harming others, while results by Conrads and Irlenbusch (2011) suggest that even deliberate ignorance is not being punished.

To sum up: based on the notion of subjective norm compliance, our model yields novel predictions that cannot be explained by other theories, and it provides an alternative

explanation for DWK's results. As DWK's design is not a suitable test for our new model, we conducted a new experimental test for the prediction of strategic information acquisition in normative choice situations. We find convincing evidence for strategic information acquisition in line with subjective norm compliance.

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.

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. (2012). Norm manipulation, norm evasion: Experimental evidence. *SSRN*.

Bicchieri, C. and Mercier, H. (2012). Self-serving biases and public justifications in trust games. *Synthese*.

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

Conrads, J. and Irlenbusch, B. (2011). Strategic ignorance in bargaining. *IZA Discussion Papers*.

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.

Grossman, Z. (2010). Strategic ignorance and the robustness of social preferences.

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. (2008). Identifying social norms using coordination games: Why does dictator game sharing vary? *IZA Discussion Paper No. 3860.*

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

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.

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.*

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