Global Inspection Games (GIG) in the laboratory

Sanchez Villalba, Miguel

August 2017

Global Inspection Games (GIG) in the laboratory*

Miguel Sanchez Villalba

August 9, 2017

Abstract

Sanchez Villalba (2015) claims inspection games can be modelled as global games when agents face common shocks. For the tax evasion game –his leading example– he prescribes that the tax agency should audit each individual taxpayer with a probability that is a non-decreasing function of every other taxpayer's declarations ("relative auditing strategy").

This paper uses experimental data to test the predictions of the model and finds supporting evidence for the hypothesis that the relative auditing strategy is superior to the alternative "cut-off" one.

It also finds that data fit the qualitative predictions of the global game model, regarding both participants' decisions and the experiment's comparative statics.

JEL Classification: C91, C7, D8, D9, H26

Keywords: Global Games, Experimental Economics, Tax Evasion, Rationality, Information, Beliefs

*I thank Frank Cowell, Georg Weizsacker, Silvia Martínez Gorricho and seminar participants at the London School of Economics for helpful comments and discussions. I gratefully acknowledge financial support from the Spanish Ministry of Science and FEDER funds under project SEJ 2007-62656 and from the Spanish Ministry of Economics and Competition under project ECO2012-34928.

1 Introduction

Common income shocks that affect rather homogeneous agents in similar ways are well documented: the fact that airlines' sales plummeted after 9/11, chicken breeders faced low demand after the avian flu outbreak, and emergent markets have difficulties attracting investors every time the U. S. Federal Reserve increases interest rates are just a few examples one can bring forward. Furthermore, they show that often these shocks are the main source of income variability, greatly overshadowing the significance of idiosyncratice shocks.

Hence, it is not surprising that a tax agency that ignores common shocks will choose a clearly suboptimal auditing strategy. But this is exactly what happens if they follow the most popular policy prescribed by the literature: the "cut-off rule" (Reinganum and Wilde (1985)). It states that the agency should not audit any firm that declares above a certain fixed cut-off income level, and should audit with a sufficiently high probability those who declare below it. Combined with common income shocks, this policy leads to systematic mistargeting: the agency audits "too much" in bad years and "too little" in good ones.

In this scenario, Sanchez Villalba (2015) finds that the agency's optimal policy (named "relative auditing strategy" by the author) consists in auditing every firm with a probability that is a non-decreasing function of every other taxpayer's declarations. This is because other firms' declarations give the agency information about the realisation of the shock and so the probability of a given taxpayer being an evader is (weakly) higher the higher are her fellow taxpayers' declarations.

The purpose of this paper is therefore to test Sanchez Villalba (2015)'s model (henceforth, GIG model, or "Global Inspection Game" model). This is a relevant task because it will help determining which of the alternative rules (relative or cut-off) is superior to the other and, indirectly, whether the data is consistent with the modelling of tax evasion as a global game and its associated predictions.

However, real-world data on tax evasion is not readily available. Those who engage in tax evasion are not willing to indicate it for obvious reasons, but also tax agencies are reluctant to provide data because of the confidentiality of tax returns: even if the datapoints are not labelled, in many cases it is quite easy to identify which individual firm they belong to, thus revealing sensitive information that could affect the company negatively.

For this reason, the current paper will use the second-best available dataset, namely, the one collected in a computerized experiment in which participants interacted with each other in situations that resembled the scenario described by the GIG model. This methodology has the obvious disadvantage of making difficult the extrapolation of results from the sample to the population, but it gives the experimenter a greater control over the variables under study and in the case of tax evasion it is, as mentioned before, the only available one anyway.

The econometric analysis finds that the agency is better off when using the relative rule than when using the cut-off one, and so that the key prediction of the GIG model is strongly supported. It also supports the hypothesis that people make decisions (qualitatively) consistent with higher-order beliefs (which play an important role in ensuring the uniqueness of the global game equilibrium) and that the comparative statics follow the ones predicted by the global game technique.

To the best of my knowledge, nobody tested empirically (using either real-world or experimental data) the predictions of a GIG-like model, but plenty of laboratory experiments were framed as/based on tax compliance problems. The closest reference is Alm and McKee (2004), which analyses tax evasion as a coordination game. In contrast, the present analysis considers it as a global game, which requires not only the strategic uncertainty generated by the coordination game but also the "fundamental uncertainty" created by the incompleteness of information regarding the payoff functions. Tests of the global game technique seem to support it in terms of predictive power (Cabrales et al. (2007)) and/or comparative statics (Heinemann et al. (2004)), but are less supportive of the participants' use of "higher-order beliefs" when making decisions. The latter result is also the conclusion of other studies, like Stahl and Wilson (1994) and Bosch-Domenech et al. (2002).

The rest of the paper is organized as follows. Section 2 summarizes Sanchez Villalba (2015)'s theoretical model and its predictions. Section 3 explains the experimental design and the testable hypotheses. Section 4 presents the results and finally section 5 concludes.

2 Tax Evasion as a Global Game

The global game methodology (Carlsson and van Damme (1993), Morris and Shin (2002)) is a mechanism that, thanks to the existence of some uncertainty about the payoff functions of the players, selects one of the multiple equilibria of a coordination game.

Sanchez Villalba (2015) claims that, in the presence of common income shocks, tax evasion can be modelled as a global game because the agency's optimal policy generates a coordination game and taxpayers' imperfect information about the agency's "type" creates the uncertainty about payoffs.

Drawing on the fact that most tax agencies worldwide partition the population of taxpayers into categories where members share some non-manipulable characteristics, he analyses the agency's problem within each one of them. The high degree of homogeneity within a category implies that the idiosyncratic shocks will be small compared to the common ones, and so, for all practical purposes, one can assume that every member has the same income y: it is high ($y = 1$) in "good" years (which occur with probability γ) and low ($y = 0$) in "bad" ones (with probability $1 - \gamma$).

The situation is thus modelled as a one-shot game, with the following timing: in the first stage all actors learn their private information; in the second one taxpayers submit their declarations and; in the third stage the agency (after observing all declarations) undertakes audits (if any). An agency's private information is its "type", parameterized by λ and interpreted as the effective budget the agency has for undertaking audits. In turn, the private information of a taxpayer i consists of her income $y_i \in \{0,1\}$ and her signal $s_i := \lambda + \varepsilon_i$, where ε_i is a white noise error term. This signal embodies all the information about the agency's type available to the taxpayer (news, previous experience, conversations with colleagues/friends, etc.). All actors (taxpayers and agency) know every parameter of the game and their own private information. They also know the probability distributions of other actors' private information, but not their realizations.[1]

Every taxpayer has to decide how much income to declare, $d_i \in \{0,1\}$, in order to maximize her expected utility. The optimal declaration strategy follows the standard literature except for the fact that, since the exact probability of detection a_i is unknown to the taxpayer, her declaration will be a (weakly) increasing function of her **belief** about a_i.

The agency chooses a_i in order to minimize the expected losses associated with making targeting errors, subject to the effective budget constraint determined by its type λ. Targeting errors can take two forms: *zeal* errors (Z) occur when resources are wasted on auditing compliant taxpayers; *negligence* errors (N) take place when evaders are not caught and the corresponding fines are not collected.[2] The agency minimizes a "loss function" that aggregates errors into one metric and can be written as $L = \mu N + (1-\mu) Z$, where μ is the loss associated with letting an evader get away with her evasion.

Sanchez Villalba (2015) found that the agency's optimal auditing policy regarding taxpayer i, a_i, is (weakly) increasing in the agency's type, λ, and the declarations of every other taxpayer in the category, d_j, $j \neq i$. The last result is especially important because it generates a *negative externality* between taxpayers: the higher the declaration of a taxpayer j, the higher the probability that another taxpayer i ($i \neq j$) is audited and the lower the latter's expected utility. Together with the optimal declaration strategy, this creates the *strategic complementarities* between taxpayers' declarations that constitute the defining feature of a **coordination game**. Specifically, the higher the declaration of taxpayer j, the higher the incentives of taxpayer i to comply as well.

The associated problems of multiplicity of equilibria are, however, side-stepped because of the taxpayers' uncertainty about the realized agency's type, λ, and the heterogeneous beliefs

[1] Except in the case of income y, of course, because it is assumed that everyone in a category has exactly the same level of income. Adding some income heterogeneity avoids this "perfect observability" issue but does not provide any new insight or affect the predictions, so for simplicity this avenue is not pursued.

[2] Formally, if $\mathbf{1}$ is an indicator function that takes the value 1 if the taxpayer is audited and 0 if she is not, then a zeal error ($Z = 1$) occurs when $\mathbf{1}(1-(1-d)y) = 1$; a negligence one ($N = 1$) when $(1-\mathbf{1})(1-d)y = 1$. Implicit in the latter formula is the assumption that it is always profitable for the agency to audit a known evader (i.e., in such cases the fine is greater than the cost of the audit). The alternative possibility implies the uninteresting solution where nobody is audited, even known evaders.

about a_i they derive from their disparate private signals, $E(a_i \mid s_i)$. This "*fundamental uncertainty*", plus the "*strategic uncertainty*" generated by the coordination game, create the conditions for modelling tax evasion as a **global game**. This leads, through a process akin to the "iterated deletion of strictly dominated strategies" (IDSDS) method, to a unique equilibrium: in each iteration, signals provide information about what other taxpayers will *not* do, and in the end it ensures that only one strategy survives, namely, one where taxpayers with low signals (and hence low beliefs about being discovered) evade, while those with high signals comply. Furthermore, equilibria with full, partial and zero evasion can arise, depending on the value of the parameters.

The key prediction of the GIG model is that an agency that implements this "relative" auditing policy will do (weakly) better than if it implemented the standard "cut-off" one, *ceteris paribus*. Testing this hypothesis is the main purpose of the present study, though the experimental dataset is rich enough as to allow for the investigation of others that will also be analysed, like the use of higher-order beliefs or the comparative statics generated by changes in the parameters of the problem.

3 Experiment design

The experiment took place at the ELSE computer laboratory of the University College London (United Kingdom).[3] 76 people took part in four treatments (labelled *GC*, *GE*, *LC* and *LE* for reasons to be explained later in this section), each involving a 60-to-90-minute long session.[4] They were not allowed to communicate for the entirety of the session and could not see other people's screens.

Each session consisted of 6 sections, namely, instructions, short quiz, trial rounds, experimental rounds, questionnaire and payment. The instructions were read aloud by the instructor and, in order to ensure their correct understanding, the participants were asked to complete a "short quiz" (shown in appendix A; correct answers and the rationale for them were provided by the instructor after a few minutes). For the same reasons, participants then played two "trial" (practice) rounds whose outcomes did not affect their earnings. After each of these first three stages the instructor answered subjects' questions in private. Twenty independent experimental rounds were then played, and after that, subjects completed a questionnaire with information regarding personal data and the decision-making

[3] The pool of participants was recruited by ELSE from their database of about 1,000 people (most of them UCL students). Two hundred of them were chosen randomly and invited to take part and the first 100 who accepted the offer were allocated to sessions according to their time preferences. Five "reserve" people were invited to each session and 7 of them had to be turned down because the target number (20 per session) was reached or because the treatment required an even number of participants (treatments GC and GE). Each one of them was paid the £5 show-up fee before being dismissed. No person was allowed to participate in more than one session.

[4] Participants were lined up outside the lab according to their arrival time. At the designated time they entered and freely chose where to sit.

process they followed. Finally, each participant was paid an amount of money consisting of a fixed show-up fee (£5) and a variable component equal to the earnings accumulated over the 20 experimental rounds.[5] Table 1 shows the exchange rate used to translate experimental currency into money, as well as other payment-related summary statistics.[6]

Treatment	Participants	£ per 1000 points	Min/Avg/Max Payment		
GC	18	0.50	10.80	11.52	11.80
GE	18	0.90	7.40	9.30	9.80
LC	20	0.50	11.60	11.65	11.80
LE	20	0.90	9.80	11.20	11.60
All	76		7.40	10.95	11.80

Note: £ per 1000 points is the exchange rate at which 1000 "experimental points" where transformed into pounds.

Table 1: Treatments. Participants and Money.

Each experimental round consisted of two stages: the "Choice" one, where participants had to make a decision that would affect their payoffs, and the "Feedback" one, where they got information about the round outcome.

		Column player	
		Y	Z
Row player	Y	$x(Y, Y, q)$	$x(Y, Z, q)$
	Z	$x(Z, Y, q)$	$x(Z, Z, q)$

Note: Only Row player's payoffs (x) are shown. Payoff's components are Row player's action, Column player's action and the realisation of the random variable q. Column player's payoffs are symmetrical.

Table 2: Stage game.

In the "choice" stage a one-shot game was played where the subjects had to choose one of two possible actions (Y or Z) interpreted as *Evasion* and *Compliance*, respectively (the game's normal form for the 2-person case is shown in table 2). In the experiment we focus on the case in which all taxpayers/players have high income ($y = 1$). The reasons for this are that introducing the possibility of low income periods will not add to our knowledge (trivially, if $y = 0$ everyone declares truthfully) and that all interesting hypotheses to test are related to the high-income scenario (not to mention the extra cost and time that running this expanded experiment will demand). Thus, in the experiment choosing Y (Z) corresponds to declaring low (high) income: $d_i = 0$ ($d_i = 1$) in the terminology of section 2.

A participant i's payoff is a function of her own decision, $d_i \in \mathcal{D} := \{Y, Z\}$, the decisions of

[5] In other experimental studies (Heinemann et al. (2009) among them) participants were paid according to the result of one randomly-chosen round. The rationale for this is that it avoids hedging, something that is not a problem here: the maximum payment a person can receive in any given round is £0.50 or £0.90 (depending on the treatment), with expected values in the £0.30-£0.35 range.

[6] In order to minimize delays and computational hassle, every person's payment was rounded *up* to the closest multiple of £0.20. Participants were not told about this arrangement until *after* they completed their questionnaires in order to avoid strategic play with respect to this peripheral matter.

the other $n-1$ people in her category, $\mathbf{d}_{-i} := (d_1, ..., d_{i-1}, d_{i+1}, ..., d_n)$, $\mathbf{d}_{-i} \in \mathcal{D}^{n-1}$, and the realisation of a random variable, $q \in \mathcal{Q} := \{A, B, C\}$. Formally,

$$x_i := x\left(d_i, \mathbf{d}_{-i}, q\right) \tag{1}$$

Different choices have different effects on payoffs, and so, while the payoff of choosing Y (evasion) can vary, that of option Z (compliance) is a known, fixed quantity. Formally, for every $\mathbf{d}_{-i}, \mathbf{d}'_{-i} \in \mathcal{D}^{n-1}$, $q, q' \in \mathcal{Q}$,

$$x(Z) := x\left(Z, \mathbf{d}_{-i}, q\right) = x\left(Z, \mathbf{d}'_{-i}, q'\right) \tag{2}$$

The random variable q can take values A, B and C with probabilities $\Pr(A) = 0.20$, $\Pr(B) = 0.60$ and $\Pr(C) = 0.20$, respectively. It represents the different possible "types" of agency regarding evasion (A : soft, B : medium, C : tough) and corresponds to the "λ" mentioned in section 2. It affects evasion payoffs negatively: the tougher the agency, the more likely the evader will be audited and the lower her expected payoff.[7] Formally, for every $\mathbf{d}_{-i} \in \mathcal{D}^{n-1}$,

$$x\left(Y, \mathbf{d}_{-i}, A\right) > x\left(Y, \mathbf{d}_{-i}, B\right) > x\left(Y, \mathbf{d}_{-i}, C\right) \tag{3}$$

At the time of making a decision participants do not know the value of q, but each one of them gets a private signal $s_i \in \mathcal{S} := \{a, b, c\}$ (called "hint" in the experiment) that is related to the realized value of q as shown in table 3 (and in the Instructions sample in appendix A). The instructions highlighted the fact that different people could get different hints but q was the same for everyone. No other probabilities were provided explicitly, though the instructions did supply the information required for their computation, namely, the prior probability distribution of q, $\Pr(q)$, and the conditional one, $\Pr(q|s_i)$.[8]

If hint =...	...then q =...	...with probability $\Pr(q\|s_i)$ =...
a	A	1.000
	A	0.125
b	B	0.750
	C	0.125
c	C	1.000

Table 3: Hints and q.

[7] In the theoretical model $x(Y, \mathbf{d}_{-i}, q)$ corresponds to the *expected* payoff taxpayer i gets if she evades: with some probability she is caught and pays a fine (low payoff) and with the remaining probability she gets away with her evasion (high payoff).

In the experiment, however, audits are not undertaken and therefore payoffs are fixed. This is so because the experimental setup is already quite demanding for subjects as to increase the level of complexity by introducing uncertainty and, furthermore, doing so is not expected to provide any significant insight beyond the ones obtained with this simpler, neater setup.

Of course, subjects do face uncertainty regarding the choices made by other players ($\mathbf{d}_{-i}$) and the type of agency they face (q), as suggested by the theoretical model.

[8] A "Choice stage" screenshot (labelled "Choice screen" in the experiment) can be seen in the instructions sample in appendix A. The programme used was z-Tree (Fischbacher (2007)).

The participant's submission of her decision (Y or Z) ended the "Choice" stage and gave way to the "Feedback" one, in which the person was informed about the realized value of q, the signal she received, her choice and her payoff for the round.[9] At no stage was a subject given any information about the signals or choices of any other participant, since usually taxpayers have little information about what other taxpayers know or how much income they declare.

By clicking on the "Continue" button, participants exited the "Feedback" stage and moved on to the next round (if any was left). Rounds were identical to each other in terms of their structure (Choice and Feedback stages) and rules (payoff computations, prior and conditional probability distribution of q), but may have differed in the *realized* values of the random variables (q and $\mathbf{s}$). Participants were told explicitly about this and informed that each round was independent from every other one.

3.1 Treatments

The experiment's treatments were defined according to the policy used (relative v cut-off, or "global" (G) v "lottery" (L)) and the predicted optimal strategy of the participants (which for this experiment, as will be shown later, reduces to determining the optimal choice when hint b is received: to evade E (corresponding to choosing Y) or to comply C (corresponding to choosing X).[10] This way the experimental setup can be visualized as in table 4.

		Participant's optimal strategy if hint $= b$	
		Comply (C)	Evade (E)
Auditing rule	Relative (G)	GC	GE
	Cut-off (L)	LC	LE

Table 4: Treatments.

The difference between Global and Lottery treatments is related to the effect of other subjects' choices on the payoffs of individual participants. In the Lottery treatments the rule implemented by the agency is of the cut-off type, and so what other people do does not

[9] A "Feedback stage" screenshot (labelled "Results screen" in the experiment) can be seen in the instructions sample in appendix A.

[10] Tax evasion has often been compared to a gamble in which the taxpayer "wins" (i.e., gets away with evasion) with probability w, and "loses" (i.e., is caught and has to pay a fine on top of the unpaid taxes) with probability $1 - w$. The cut-off rule is equivalent to a standard lottery (and hence the name of the treatment) because it *fixes* the chances of winning (say $w = 1 - p$) and losing ($1 - w = p$). Evasion can therefore be seen as equivalent to buying $(1 - p)\,N$ out of a total pool of N raffle tickets, each one of them equally likely to be the winner.

In the Global treatments, on the other hand, those probabilities are *not* fixed, because they are affected by what other people do. In particular, since other people's compliance has a negative impact on my payoff, the fact that other people comply is equivalent to having the total number of tickets increased to, say, $N' > N$, so that my probability of winning w' (in spite of my holding the same number of tickets as before, $(1 - p)\,N$) is now comparatively lower: $w' = \frac{(1-p)N}{N'} < \frac{(1-p)N}{N} = w$.

affect player i's payoff. Formally, for every $q \in \mathcal{Q}$,[11]

$$\mathrm{x}(Y, Y, q) = \mathrm{x}(Y, Z, q) \quad \text{if treatment} \in \{LC, LE\} \tag{4}$$

In Global treatments, on the other hand, the auditing policy followed is the relative one, implying that other people's declarations do have an impact on player i's payoff via the probability of detection. Formally, for every $q \in \mathcal{Q}$,

$$\mathrm{x}(Y, Y, q) > \mathrm{x}(Y, Z, q) \quad \text{if treatment} \in \{GC, GE\} \tag{5}$$

It is worth mentioning here that the Lottery treatment can be interpreted as a special (limit) case of the Global one in which the effect of other people's decisions on a certain participant's payoff is arbitrarily small. Consequently, and without loss of generality, henceforth the analysis will be restricted to the Global case, with the occasional reference to the Lottery one provided only when relevant.

For the experiment, participants in the Global treatments were divided in 9 groups of 2 people each, the matching protocol being random (equi-probable) within rounds and independent across them.[12] The experimental setup reproduced the three typical scenarios described by the global game literature:

† The two **extreme** cases in which the "fundamentals" are "so bad"/"so good" that there exists a strictly dominant strategy. In the experiment the fundamental is the agency's "toughness", q, and so strict dominance requires that everyone should evade when the agency is very soft ($q = A$) and that everyone should comply when it is very tough ($q = C$). Formally, for every $d' \in \mathcal{D}$,

$$x(Y, d', A) > x(Z) \tag{6}$$

$$x(Y, d', C) < x(Z) \tag{7}$$

† The **intermediate** one in which the "fundamentals" are neither "so bad" nor "so good". In this case a coordination game is created and, consequently, no strategy dominates all others: which one is optimal depends on what other people do. In the experiment, this corresponds to the scenario in which the agency's type is "medium" ($q = B$): if the other person in my group evades, it is optimal for me to evade as well; if the other person complies, I am better off complying too.[13] Formally,

$$x(Y, Y, B) > x(Z) > x(Y, Z, B) \tag{8}$$

[11] I restrict my attention to the 2-person case, which will be the relevant one throughout the paper. The extension to the n-person case is straightforward.

[12] For the rest of the paper, the variables corresponding to the two members of a group will be denoted by lowercase letters (e.g., signal s, decision d, etc.) and by primed lowercase letters (e.g., signal s', decision d', etc.), respectively.

[13] Clearly, this does not apply to the Lottery case.

Turning now to the other dimension that defines treatments, the difference between the Evasion and Compliance ones is due to their different predictions regarding what a participant's optimal strategy should be if signal b is observed. Thus, distinguishing E from C treatments demands the solving of the taxpayer problem, namely, choosing between Evasion (Y) and Compliance (Z) using all the information available (s) in order to maximize expected utility. In this setup, therefore, a taxpayer's strategy $\boldsymbol{\sigma}$ is a vector of decisions, one for each possible signal $s \in \mathcal{S}$. Formally, $\boldsymbol{\sigma} := (\sigma(a), \sigma(b), \sigma(c))$, where $\sigma : \mathcal{S} \rightarrow \mathcal{D}$ is a function that maps signals into decisions.[14] Therefore, finding the solution requires comparing the (certain) utility of compliance, $u(Z)$, and the expected utility of evasion:

$$Eu(Y, \mathbf{k}'(\mathbf{s}')|s) := \sum_{q \in \mathcal{Q}} \Pr(q|s) \sum_{s' \in \mathcal{S}} \Pr(s'|q) \{k'(s')\, u(Y, Y, q) + [1 - k'(s')]\, u(Y, Z, q)\} \tag{9}$$

where $u(Y, d', q) := u(x(Y, d', q))$ is the utility I derive from receiving payoff $x(Y, d', q)$; $s' \in \mathcal{S}$ and $d' \in \mathcal{D}$ are respectively the signal and decision of the other member of my group; $\Pr(s'|q) \in [0, 1]$ is the conditional probability of the other member getting signal s' given that the agency's type is q; and $\mathbf{k}'(\mathbf{s}') := (k'(a), k'(b), k'(c))$, such that $k'(s')$ and $1 - k'(s')$ are my beliefs regarding what the other member of my group would do if she received signal s': if I expect her to choose Y then $k'(s') = 1$ (and $1 - k'(s') = 0$), if I expect her to choose Z then $k'(s') = 0$ (and $1 - k'(s') = 1$).

This comparison depends crucially on the beliefs a player holds about the actions to be followed by the other member of her group, $\mathbf{k}'(\mathbf{s}')$, and, thus, on the ability and sophistication of the subjects at forming them, a matter that is directly related to the concepts of common knowledge and higher-order beliefs (HOBs, Carlsson and van Damme (1993)). HOBs refer to the levels of reasoning involved in reaching a conclusion and are neatly connected to the (game theoretical) method of Iterated Deletion of Strictly Dominated Strategies (IDSDS): with each iteration, the order of beliefs increases one level. Furthermore, HOBs are the key factor behind the uniqueness of the global game equilibrium: in the first iteration, $t = 1$, my private signal gives me information about the set of strategies (out of the original set, Σ^0) that are strictly dominated by others and will therefore *never* be played. In the second iteration, $t = 2$, the set of those strategies that survived the previous round of deletions is the new feasible set, Σ^1. Via an analogous mechanism, a new group of strictly dominated strategies will be discarded and after that a new iteration $t = 3$ with feasible set Σ^2 will begin. The theory of global games proves that in the limit, after an arbitrarily large number of iterations, the feasible set Σ^∞ has only one element, $\boldsymbol{\sigma}^*$. In other words, the equilibrium is unique.

In the experiment, only 2 iterations are needed to find the unique solution to the taxpayer problem.[15] Thus we can classify players based on the number of iterations used (1 or 2):

[14] Actually, it maps signals into *probability distributions* over decisions, if one allows for mixed strategies. However, this possibility was explicitly ruled out here because its inclusion would not have provided any extra, significant insight as to justify the complexity-associated problems it would have entailed.

[15] This does not apply to Lottery treatments for the obvious reason that in those cases, by definition, a

Definition 1 A player who uses only 1 iteration is defined as "Rudimentary."

Definition 2 A player who uses 2 iterations is defined as "Sophisticated."

In other words, both types of players understand the game-theoretical concept of *dominant/dominated strategy*, but differ in the scope of their understanding: while Rudimentary players only recognize what is evident, Sophisticated ones go one step further and build up on what Rudimentary players do. The following two propositions state how they rank the available strategies.[16]

Proposition 1 (Rudimentary Dominance (RD)) *According to Rudimentary players:*

1. *if $s = a$ (signal is low), Evasion strictly dominates Compliance;*
2. *if $s = b$ (signal is medium), no strategy strictly dominates the other; and*
3. *if $s = c$ (signal is high), Compliance strictly dominates Evasion.*

Proposition 2 (Sophisticated Dominance (SD)) *According to Sophisticated players:*

1. *if $s = a$ (signal is low), Evasion strictly dominates Compliance;*
2. *if $s = b$ (signal is medium), then:*
   (a) *in* E *treatments, Evasion strictly dominates Compliance; and*
   (b) *in* C *treatments, Compliance strictly dominates Evasion; and*
3. *if $s = c$ (signal is high), Compliance strictly dominates Evasion.*

The rationale for taking into account both scenarios when $s = b$ in proposition 2 reflects, above all, the lack of theoretical predictions or stylized facts about what strategy we should expect to be played in that case.

The optimal strategy of a player is therefore:

Hypothesis 1 (Optimal Strategy (OS)) *According to the global game technique, the optimal strategy of a player is as follows:*

1. *If signal is soft ($s = a$) then evade ($d = Y$);*
2. *if signal is medium ($s = b$) then:*

taxpayer's payoff does not depend on other people's choices or the taxpayer's beliefs about them.

[16] The derivation of these two results is shown in appendix B.

(a) *in* E *treatments, evade* ($d = Y$)*; and*

(b) *in* C *treatments, comply* ($d = Z$)*; and*

3. *if signal is tough* ($s = c$) *then comply* ($d = Z$)*.*

If choices satisfy all three parts of the hypothesis, then one can say they are "consistent with the SD predictions" and label the player as "Sophisticated". If they only satisfy the parts 1 and 3, they are "consistent with the RD predictions" and the player can be labelled as "Rudimentary".

3.2 Selection of payoffs

The key hypothesis to test is the following one:

Hypothesis 2 (Superiority of Relative Auditing Strategy (SRAS)) *For a given level of enforcement, Global treatments generate less (expected) targeting errors than Lottery ones for all possible types of agency,* $q \in \mathcal{Q}$*.*

The payoffs of the four treatments (shown on table 5) were chosen to make the satisfaction of hypothesis 2 as difficult as possible. This way, if the data supports the global game predictions in these most demanding conditions, then the theory could be expected to be an even better predictor in more favorable environments.

Person 1's choice	Person 2's choice	Type of agency	GC	GE	LC	LE
Y	Y	A	1,000	1,000	715	1,000
Y	Y	B	655	145	655	145
Y	Y	C	579	6	579	1
Y	Z	A	658	156	715	1,000
Y	Z	B	651	135	655	145
Y	Z	C	0	0	579	1
Z	{Y,Z}	{A,B,C}	654	140	654	140

Note: Only payoffs of Person 1 are shown. Those of Person 2 are symmetric.

Table 5: Payoffs. All treatments.

It is worth noting at this point that the global game technique selects one of the equilibria of a coordination game, an equilibrium that coincides (for 2×2 games like the ones used here) with the one selected by the "risk dominance" criterion (Harsanyi and Selten (1988)). Intuitively, the latter chooses the equilibrium which, if abandoned, inflicts the highest costs on the players. In the experiment, the risk-dominant equilibrium depends on the treatment: it is (Y, Y) in the GE treatment and (Z, Z) in the GC one. These are, not surprisingly, the choices that proposition 1 predicted to be optimal in those treatments, thus confirming that both the global games theory and the risk-dominance criterion select the same equilibrium.

There is, however, an important competitor for the risk-dominance/global game criterion: the payoff-dominance criterion. It simply states that if all equilibria can be Pareto-ranked, players will coordinate on the dominant one. In the experiment, the payoff-dominant equilibrium is always (Y, Y), regardless of the treatment.

Thus, the payoff-dominance and risk-dominance criteria select the same equilibrium in the GE treatment but different ones in the GC one. The fact that the criteria reinforce each other in GE but compete against each other in GC suggests the following hypothesis:

Hypothesis 3 (Relative Frequency (RF)) *The frequency of choices that are consistent with the global game/risk-dominance predictions is (weakly) higher in* GE *than in* GC*.*

Finally, it is important to mention here that risk aversion could dramatically alter the predictions of the model, and this may be especially important since evidence indicates that attempts to induce risk-preferences seem not to work (Selten et al. (1999)). The solution implemented in the experiment was to choose parameters such that all constraints will be satisfied for a large range of risk preferences. In particular, in E-treatments parameters are robust for degrees of relative risk aversion as high as 0.4 (about 60% of the population, according to Holt and Laury (2002)). In C-treatments, they are robust for values as low as 0 (about 80% of the population, according to the same study). Also, it is acknowledged in the experimental literature that when playing complex games people often avoids the complications of utility maximisation and instead simply maximize payoffs, which implies that risk preferences should not be an important issue here.

4 Results

A total of 1,520 observations were collected in the experiment, and table 6 shows the breakdown by treatment. It also shows summary statistics of the key variables needed for testing the hypotheses of the previous section:

Sophisticated Dominance measures the coincidence between the data and the global game theoretical predictions about the subjects' choices (SD=1 if data fits predictions and 0 otherwise). Its name reflects the fact that those predictions are based on the concept of sophisticated dominance (proposition 2).

Errors quantifies the number of tageting errors (per observation/datapoint) made by the agency (ERR=1 if an error was made, 0 otherwise).

Note that Sophisticated Dominance is never lower than 50% and Errors never above 35%.

Treatment	Observations	Sophisticated Dominance (SD)		Errors (ERR)	
		Mean	St. Dev.	Mean	St. Dev.
GC	360	0.7722	0.4200	0.1522	0.2252
GE	360	0.8639	0.3434	0.2028	0.3034
LC	400	0.5450	0.4986	0.3473	0.3303
LE	400	0.9300	0.2555	0.3243	0.3726
All	1,520	0.7757	0.4173	0.2608	0.3248

Note: SD=1 if subject's choice coincides with global game's prediction, 0 otherwise. ERR=1 if agency made an error, 0 otherwise.

Table 6: Summary Statistics. Dominance and Errors.

For hypothesis testing, it would be useful to aggregate data in two different ways, depending on the information available to the relevant actor. Thus, for hypotheses related to the decisions of the taxpayers (OS and RF), data are aggregated by signal (columns 3-5 in table 7). For those related to actions of the agency (SRAS), on the other hand, the aggregation is done according to the type of agency (columns 6-8 in the same table).

Treatment	Observations	Signal (s)			Agency's type (q)		
		a	b	c	A	B	C
GC	360	7	295	58	18	234	108
GE	360	29	292	39	54	234	72
LC	400	29	330	41	60	260	80
LE	400	51	337	12	100	280	20
All	1,520	116	1,254	150	232	1,008	280

Note: Interpretation of s/q: a/A: "soft"; b/B: "medium"; c/C: "tough".

Table 7: Number of observations, aggregated by signal and type of agency.

For the analysis, data from all subjects for all periods were pooled. This is justified because there is little variability in behavior after the first few rounds of each treatment:[17] many people choose exactly the same option every time they receive a given signal. This lack of variability over time is not a bad thing in itself (the theory actually predicts such rigidity), but it precludes the possibility of using other econometric techniques (*e.g.*, panel data).

4.1 *Optimal Strategy* and *Relative Frequency* hypotheses

The set of variables that is going to be used for testing is described in table 8.

[17] Except in the *GE* one, that requires 10 rounds to become stable. This, however, does not usually have an impact on results, and when it does, it will be mentioned in the text.

Variable	Role	Type	Description
SD	Dependent	Dummy	1 if choice coincides with prediction, 0 otherwise
Ds	Dependent	Dummy	Idem SD, but for $s \in \mathcal{S}$ given
RD	Dependent	Dummy	Idem SD, but for $s \in \{a, c\}$
AD	Dependent	Dummy	Idem SD, but for $s = b$
g	Explanatory	Dummy	1 if G treatment, 0 otherwise
e	Explanatory	Dummy	1 if E treatment, 0 otherwise
ge	Explanatory	Dummy	Interaction term: 1 if GE treatment, 0 otherwise
a	Explanatory	Dummy	1 if $s = a$, 0 otherwise
b	Explanatory	Dummy	1 if $s = b$, 0 otherwise
c	Explanatory	Dummy	1 if $s = c$, 0 otherwise

Note: "Predictions" as defined in hypothesis 1.

Table 8: Variables of the model. Dominance.

Dep. Var.→	Da	Db	Dc	RD	AD	SD
a				1.0205		0.7091
				[0]		[0]
b						0.5030
						[0]
c				0.9855		0.7671
				[0]		[0]
g	0.0000	0.2803	-0.0345	-0.0201	0.2803	0.2227
	[0.082]	[0]	[0.158]	[0.35]	[0]	[0]
e	-0.0196	0.4714	0.0000	-0.0297	0.4714	0.3928
	[0.323]	[0]	[0.706]	[0.072]	[0]	[0]
ge	0.0196	-0.3543	-0.0681	-0.0095	-0.3543	-0.2998
	[0.323]	[0]	[0.217]	[0.804]	[0]	[0]
cons	1.0000	0.4485	1.0000		0.4485	
	[.]	[0]	[0]		[0]	
Obs	116	1,254	150	266	1,254	1,520
LC	1.0000	0.4485	1.0000	1.0000	0.4485	.5450
LE	0.9804	0.9199	1.0000	0.9841	0.9199	.9300
GC	1.0000	0.7288	0.9655	0.9692	0.7288	.7722
GE	1.0000	0.8459	0.8974	0.9412	0.8459	.8639

Note: Top panel: Probability that estimate =0 is shown in brackets below estimate. Bottom panel displays observed average values of the dependent variable.

Table 9: Estimation. Dominance. Overall and by signal.

Ds measures Dominance when only observations with a given signal s are considered. RD means Rudimentary Dominance and considers only observations when signals are soft (a) or tough (c). AD measures "Advanced Dominance" and only takes into account observations with medium signals (hence, it is identical to Db).[18] The unit of observation is the individual

[18] Thus, loosely speaking, we can say that Sophisticated Dominance is the sum of Rudimentary and Advanced Dominance: Rudimentary people do only one iteration (see appendix B) and, consequently, follow the optimal strategy (hypothesis 1) *only* when they receive soft or tough signals (parts 1 and 3 of the hypothesis). In turn, Sophisticated people do two iterations meaning that, *on top of* following parts 1 and 3, the *also* follow part 2 of the hypothesis (when the signal is medium). This second, incremental iteration is thus directly connected to the concept of Advanced Dominance as defined in the text.

player and the model used is

$$SD = \beta_1 g + \beta_2 e + \beta_3 ge + \gamma_1 a + \gamma_2 b + \gamma_3 c + \varepsilon \tag{10}$$

(analogous ones are used for the other dependent variables considered). The estimates are shown in table 9.

Dep. Var.→	Da	Db	Dc	RD	AD	SD
LC		X			X	X
LE		X			X	X
GC		X			X	X
GE		X	X	X	X	X
LC=GC		GC			GC	GC
LE=GE	GE	LE	LE		LE	LE
LC=LE	LC	LE			LE	LE
GC=GE		GE			GE	GE

Note: Top panel: Empty if data fits prediction in hypothesis OS; "X" otherwise. Bottom panel: Empty if no statistically-significant difference, treatment with higher dominance otherwise.

Table 10: Dominance tests. Predictions and inter-treatment comparisons.

Table 10 shows the results of the tests in a schematic way.[19] The first panel tests the OS hypothesis (see note below the table for interpretation of symbols). The null hypothesis is that data are consistent with the predictions of the Global Games technique (hypothesis 1),[20] a hypothesis that is supported in the cases of soft and tough signals ($s = a$ and $s = c$) and that implies that people are, at least, Rudimentary.[21] When the signal is medium, however, the Global Game predictions are rejected for all treatments and, therefore, the OS hypothesis is quantitatively rejected as well (*i.e.*, those aspects related to part 2 of the hypothesis). Qualitatively, however, the results do support the predictions, as can be seen in figures 1 and 2, where the observed strategies resemble the shape of the predicted ones (except for LC).[22] Having in mind the discreteness of the model (which amplifies divergences) and that the parameters were chosen to make the test as difficult to pass as possible for the Global Games theory, the result is still encouraging.

Result 1 (Qualitative Sophistication (QS)) *People are, at least, Rudimentary: they act as predicted by the* OS *hypothesis when signals are soft or tough. The hypothesis that they make decisions in a way consistent with the second part of the* OS *hypothesis is rejected in quantitative terms (and so is the* OS *hypothesis, consequently) but supported in qualitative terms.*

The bottom panel of table 10 compares the levels of Dominance of the different treatments.

[19] The tests are shown in table 17 in appendix E.

[20] The predicted value is 1 for all cases, which means that all observations should match predictions.

[21] The null hypothesis is rejected in the GE case because of an outlier. If discarded, the hypothesis cannot be rejected.

[22] In the figures, 1 corresponds to Evasion (choice Y) and 0 to Compliance (choice Z).

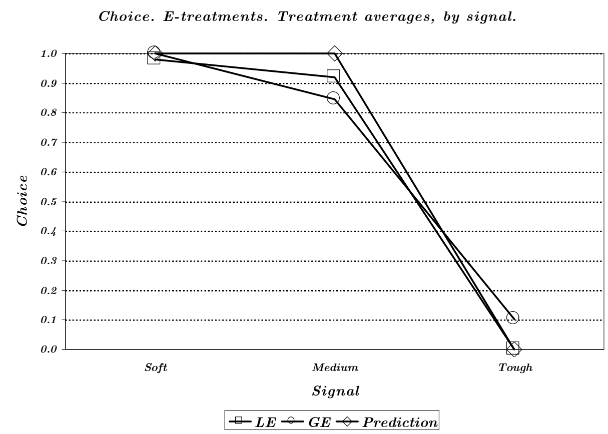

Figure 1: Observed and Predicted choices. E-treatments.

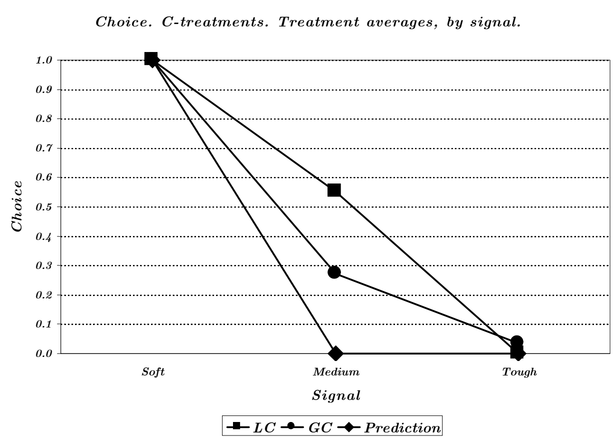

Figure 2: Observed and Predicted choices. C-treatments.

The null hypothesis for the first two lines is that Dominance is the same in Global and Lottery treatments, *i.e.*, when the Relative Auditing Strategy (RAS) and the Cut-Off Rule (COR) are used, respectively. The table shows that the hypothesis is supported for RD but not for AD and SD. On the other hand, the theory cannot explain why AD and SD are higher for *Global* in the C-treatments but higher for *Lottery* in the E-treatments. It is worth noting, though, that the difference between GE and LE vanishes when only the last 10 periods of both treatments are considered (see figure 4). So we get that:

Result 2 (G v L Dominance (GLD)) *Global treatments foster more Advanced and Sophisticated Dominance than Lottery ones. There is no difference between the two treatments in terms of Rudimentary Dominance.*

For the last two lines, the null hypothesis is that Dominance is the same in Evasion and Compliance treatments. Once again, it is satisfied for RD but not for AD and SD. But now AD and SE are higher in E treatments than in C ones, regardless of the auditing rule used (RAS or COR). This can be explained –for the GC v GE case, last line of the panel– by the coincidence of the risk- and payoff-dominant equilibria in the GE treatment and the discrepancy between them in the GC one.[23] This is thus consistent with the RF hypothesis:

Result 3 (C *v* E Dominance (CED)) *Evasion treatments foster more Advanced and Sophisticated Dominance than Compliance ones. There is no difference between the two treatments in terms of Rudimentary Dominance. Thus, the* RF *hypothesis cannot be rejected.*

These results can also be visualized in figures 3 and 4. The first one confirms that RD is strongly supported by data and that different treatments do not affect it. The second one focuses on choices when the signal is medium and attests that AD and SD predictions are quantitatively rejected, though they are qualitatively supported in all treatments but LC. It also shows that treatments can be ranked as determined by the tests, namely, (from higher to lower Sophisticated Dominance), LE, GE, GC and LC.[24]

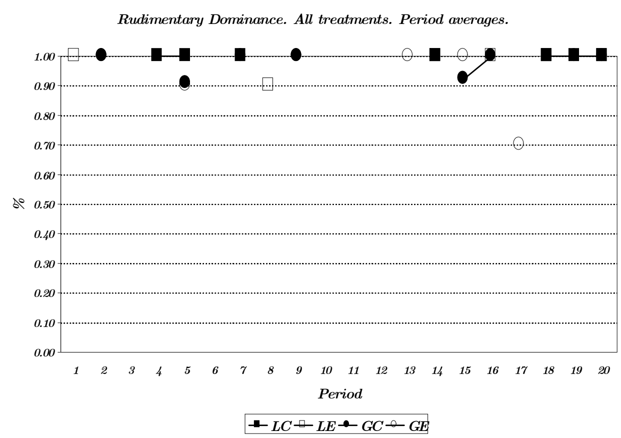

Figure 3: Rudimentary Dominance. All treatments. Period averages

4.2 *Superiority of Relative Auditing Strategy* hypothesis

The key prediction of the GIG model is that a tax agency would be advised to use the relative auditing strategy (RAS) and to discard the cut-off rule (COR). Following Sanchez Villalba (2015), this means that –for given enforcement costs– the agency would make less targeting

[23] The theory is unable to explain the difference between E and C in the Lottery treatments.

[24] Restricting attention to the last 10 periods so that the learning process in GE converges, the difference between GE and LE.

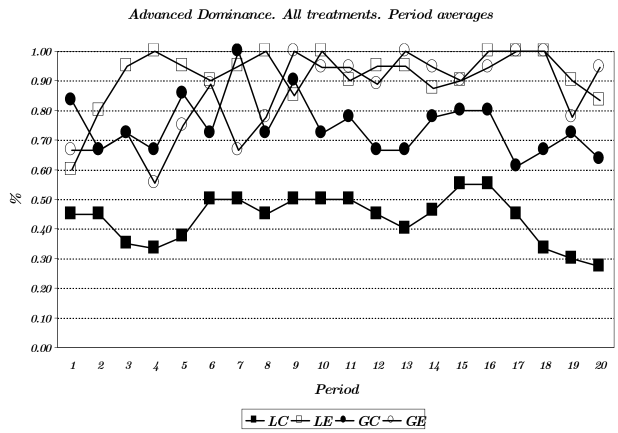

Figure 4: Advanced Dominance. All treatments. Period averages.

errors if implementing the RAS than if using the COR. These targeting errors are the *Zeal* and *Negligence* ones defined in section 2 (see especially footnote 2), though –for the reasons explained on page 6– the analysis will focus on the *Negligence* errors only.

The unit of observation is the 2-person group of players in the G-treatments and it is the individual player in the L-treatments. Thus, in order to be able to compare them, the G-treatment errors were normalised and expressed in *per capita* terms. The model to be estimated is thus:

$$ERR = \beta_1 g + \beta_2 e + \beta_3 ge + \gamma_1 A + \gamma_2 B + \gamma_3 C + \varepsilon \tag{11}$$

where the variables are defined as in table 11.

Variable	Role	Type	Description
ERR	Dependent	Ordinal	$\begin{cases} \text{In } LC, LE\text{: 1 if an error was made, 0 otherwise} \\ \text{In } GC, GE\text{: 1 if 2 errors, } \frac{1}{2} \text{ if 1 error, 0 otherwise} \end{cases}$
ERRq	Dependent	Ordinal	Idem ERR, but for $q \in \mathcal{Q}$ given
g	Explanatory	Dummy	1 if G treatment, 0 otherwise
e	Explanatory	Dummy	1 if E treatment, 0 otherwise
ge	Explanatory	Dummy	Interaction term: 1 if GE treatment, 0 otherwise
A	Explanatory	Dummy	1 if $q = A$, 0 otherwise
B	Explanatory	Dummy	1 if $q = B$, 0 otherwise
C	Explanatory	Dummy	1 if $q = C$, 0 otherwise

Note: ERR measures negligence errors *per capita* in a 2-person group in treatments GC, GE and individual negligence erros in treatments LC, LE.

Table 11: Variables of the model. Errors.

The estimates can be seen in table 12.

Dep. Var.→	ERRA	ERRB	ERRC	ERR
A				0.8610
				[0]
B				0.2886
				[0]
C				0.1526
				[0]
g	0.0059	-0.1610	-0.1847	-0.1242
	[0.956]	[0]	[0]	[0]
e	0.3718	-0.2130	-0.1950	-0.1007
	[0]	[0]	[0]	[0]
ge	-0.0967	0.1466	0.1856	0.0804
	[0.423]	[0]	[0]	[0]
cons	0.5482	0.3477	0.1954	
	[0]	[0]	[0]	
Obs	232	1,008	280	1,520
LC	0.5482	0.3477	0.1954	0.3473
LE	0.9200	0.1346	0.0004	0.3243
GC	0.5541	0.1866	0.0107	0.1522
GE	0.8293	0.1203	0.0013	0.2028

Note: Top panel: Probability that estimate =0 is shown in brackets below estimate. Bottom panel displays observed average values of the dependent variable.

Table 12: Estimation. Errors. Overall and by type of agency.

In a fashion similar to the one used in section 4.1, several tests are shown in a schematic form in table 13 (the values of the tests can be found in table 18 in appendix E).

Dep. Var.:	ERRA	ERRB	ERRC	ERR
LC	+	+	+	+
LE	-	-		
GC	+	+		+
GE	-	-		-
LC=GC		GC	GC	GC
LE=GE		GE	LE	GE
LC=LE	LC	LE	LE	LE
GC=GE	GC	GE		

Note: Top panel: Empty if data fits predictions; "+" if observed errors are higher than predicted; "-" otherwise. Bottom panel: Empty if no statistically-significant difference, treatment with less errors otherwise.

Table 13: Errors tests. Predictions and inter-treatment comparisons.

The top panel tests the accuracy of predictions and shows that the data do not fit them. In particular, errors are usually higher than predicted in C treatments but lower than predicted in E ones. This is consistent with the Dominance results, which indicate that "too many" people evade when they should comply (C treatments) and comply when they should evade (E treatments). The main conclusion, thus, is basically the same as the one found for

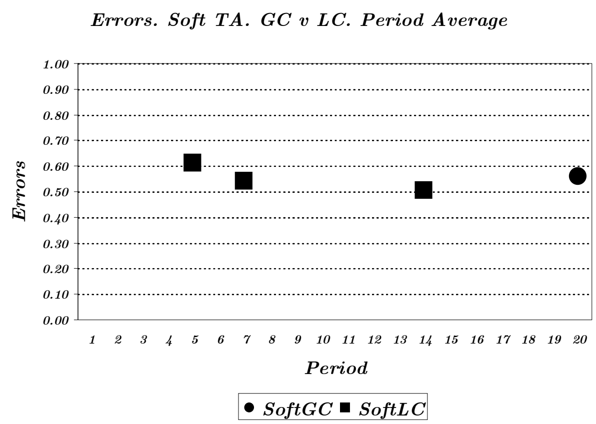

Figure 5: Errors. Soft agency. GC v LC.

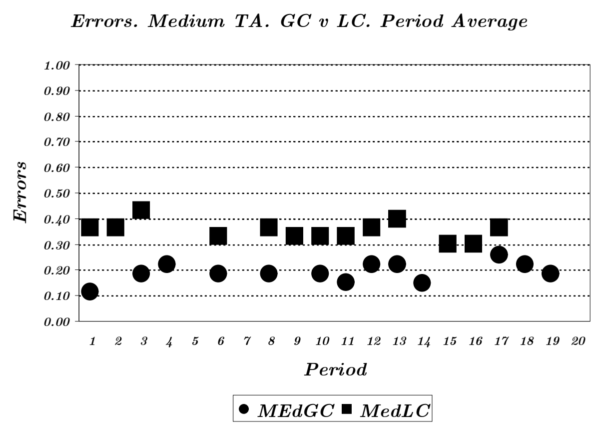

Figure 6: Errors. Medium agency. GC v LC.

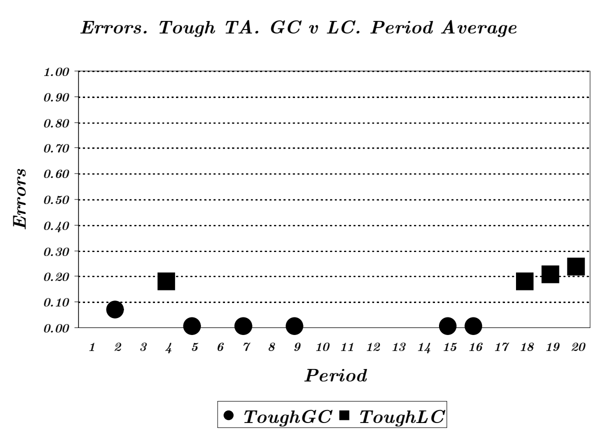

Figure 7: Errors. Tough agency. GC v LC.

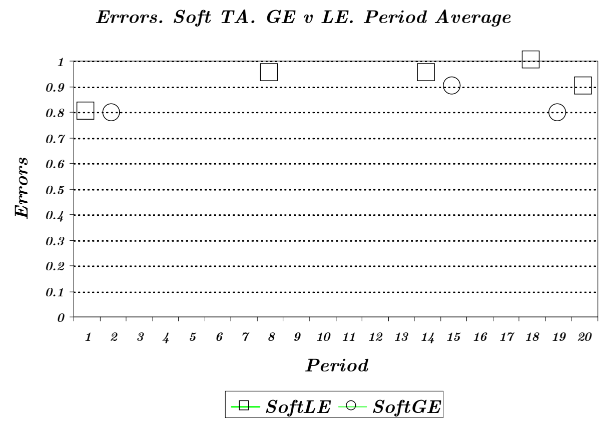

Figure 8: Errors. Soft agency. GE v LE.

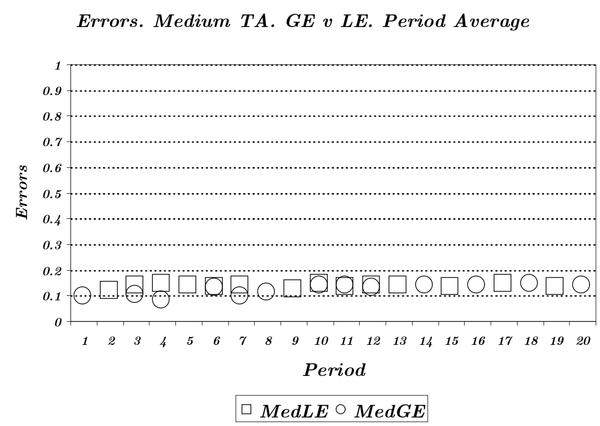

Figure 9: Errors. Medium agency. GE v LE.

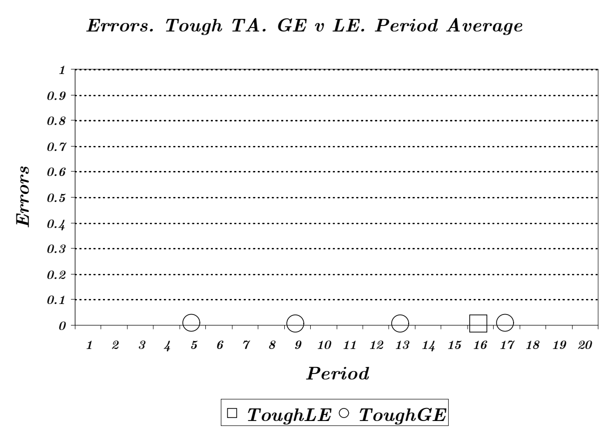

Figure 10: Errors. Tough agency. GE v LE.

Dominance in Result 1, and subject to the same qualifications.

The first two lines of the bottom panel are the important ones: they show the tests for the *SRAS* hypothesis. Given the minimum variability in the extreme cases (when the agency is too soft, $q = A$, or too tough, $q = C$), the relevant tests are those for the medium one, and this one shows clearly that the *Global* treatments lead to less errors per capita than the *Lottery* ones. In other words, the *SRAS* hypothesis is strongly supported.

Result 4 (Superiority of the Relative Auditing Strategy (SRAS)) *From the agency's perspective, the Relative Auditing Strategy (RAS) is better than the Cut-Off Rule (COR).*

The last two lines test whether there are significant differences between E and C treatments and show (again focusing on the medium case) that the first lead to less errors than the second. Again, this can be linked to the Dominance analysis, where E treatments show a higher degree of coincidence with predictions than C ones. This means, in other words, than in the latter many people evaded when they should have complied, and the higher number of associated errors thus explains the present result.

Finally, it is important to notice that all these findings are also supported graphically, as shown in figures 5 to 10. It can be clearly seen there that G treatments (*i.e.*, those in which the Relative Auditing Strategy is implemented) lead to (weakly) less errors than L ones (those in which the Cut-Off Rule is used). The figures also show that errors are a decreasing function of the agency's "toughness", which is consistent with the comparative statics predicted by the Global Games theory.

Result 5 (Effect of agency's type (EAT)) *Errors decrease with the agency's "toughness".*

4.3 Characteristics and Decisions

The analysis can be deepened by using the information collected in the questionnaire run after the experimental rounds. The relevant variables are shown in table 14.

Variable	Role	Type	Description
AD	Dependent	Dummy	1 if data fits proposition SD (part 2), 0 otherwise
g	Explanatory	Dummy	1 if G treatment, 0 otherwise
e	Explanatory	Dummy	1 if E treatment, 0 otherwise
ge	Explanatory	Dummy	Interaction term: 1 if GE treatment, 0 otherwise
gender	Explanatory	Dummy	1 if female, 0 otherwise
age	Explanatory	Natural	
study	Explanatory	Ordinal	0 : no study, 1 : non-economics, 2 : economics
# exp	Explanatory	Ordinal	0 : none, 1 : 1 to 4, 2 : 5+ experiments
math	Explanatory	Ordinal	0 : none, 1 : basic, 2 : advanced knowledge
prob	Explanatory	Ordinal	0 : none, 1 : basic, 2 : advanced knowledge
game	Explanatory	Ordinal	0 : none, 1 : basic, 2 : advanced knowledge

Note: "Study" refers to "area of study". "Math"/"Prob"/"Game" refer to knowledge of mathematics, probability theory and game theory, respectively.

Table 14: Questionnaire variables. Dominance.

The analysis will be restricted to that of AD. The reasons for this are two: first, the previous section proved that RD is satisfied almost perfectly for the whole sample of participants, regardless of their individual characteristics; and second, AD is the main source of SD variability, since in most observations the signal is medium (see table 7).

The question we want to address is: what (if any) are the personal characteristics that drive players' choices?[25] In order to answer it, the variables defined in table 14 were used to estimate the following model (the unit of observation is the individual player):

$$\begin{aligned} AD = \alpha + \beta_1 g + \beta_2 e + \beta_3 ge + \\ + \gamma_1 gender + \gamma_2 age + \gamma_3 \# \exp + \gamma_4 math + \gamma_5 prob + \gamma_6 game + \varepsilon \end{aligned} \tag{12}$$

The results (shown in table 15) indicate that estimates are robust to the specification of the model (last three columns)[26] and that most of the times there is not much difference between treatments or between individual treatments and the whole sample. The analysis finds that being male, young, not-knowledgeable at maths and not-knowledgeable at game theory makes a subject more likely to make decisions that coincide with the predictions of the Global Games theory. There is no rationale for the gender effect (which, apart from the whole sample, is significant only in the LE treatment), though it is important to note that a similar result is found by Heinemann et al. (2009). The age effect may seem to reflect that most subjects are university students, but actually it is driven by a few older outliers: if the analysis restricts its attention to "up-to-25-year-olds" (1,050 observations), age becomes non-significant. A similar story can be told about mathematics: it becomes insignificant

[25] To complement this enquiry, subjects were classified into categories according to the strategies that they followed in the experiment. The analysis is presented in appendix C of the appendix. The "Chance Maximizers" category is particularly important, as it is postulated as the main factor that could explain why treatment LC yields results significantly different from the ones predicted by the Global Games theory (together with the risk-dominance/payoff-dominance equilibrium).

[26] For this very reason, only OLS estimates are shown throughout the whole paper.

when the “young” sample is used (thus eliminating the puzzling result that the estimate’s sign was negative). Area of study is not significant and, surprisingly, neither are knowledge of probability theory or participation in other experiments (though Heinemann et al. (2004) find the same result regarding experience[27]).

	OLS					Probit	Logit
	GC	GE	LC	LE	All	All	All
g					0.2914	0.8573	1.3932
					[0]	[0]	[0]
e					0.4895	1.7349	3.0951
					[0]	[0]	[0]
ge					-0.3894	-1.4149	2.4897
					[0]	[0]	[0]
gender	-0.0306	0.0078	0.0967	-0.0738	-0.0616	-0.2761	0.4752
	[0.745]	[0.853]	[0.31]	[0.005]	[0.011]	[0.003]	[0.004]
age	-0.0385	-0.0282	-0.0251	0.0039	-0.0078	-0.0304	0.0540
	[0]	[0]	[0]	[0]	[0]	[0]	[0]
study	0.1916	0.0349	-0.5397	-0.0149	-0.0306	-0.1446	0.2348
	[0.004]	[0.514]	[0]	[0.787]	[0.369]	[0.312]	[0.382]
#exp	0.0006	0.1402	-0.0677	-0.0160	-0.0059	-0.0416	0.0729
	[0.988]	[0]	[0.161]	[0.575]	[0.738]	[0.513]	[0.507]
maths	-0.5119	0.0418	0.1660	-0.0788	-0.0993	-0.3749	0.6982
	[0]	[0.432]	[0.196]	[0.099]	[0.002]	[0.001]	[0.002]
prob	-0.0454	-0.0060	0.0033	0.1204	0.0047	0.0219	0.0314
	[0.635]	[0.893]	[0.963]	[0.007]	[0.866]	[0.834]	[0.867]
game	0.3409	0.1840	0.0362	0.0242	0.0961	0.4249	0.6918
	[0]	[0]	[0.464]	[0.193]	[0]	[0]	[0]
cons	2.0118	1.2868	1.3768	0.8555	0.7772	1.1536	2.1058
	[0]	[0]	[0]	[0]	[0]	[0]	[0]
Obs	295	292	330	337	1,254	1,254	1,254

Note: Probability that estimate =0 is shown in brackets below estimate.

Table 15: Estimation. Effect of personal characteristics on choices.

The only robustly significant variable seems to be knowledge of game theory, which has a positive effect on AD. Furthermore, it is significant in both treatments in which strategic (*i.e.*, game theoretic) interactions took place. This may indicate that some degree of indoctrination may have played a role and so that training can breed “sophistication”. This suggests that a typical population (in which knowledge of game theory is negligible for most people) could make choices quite different from the ones suggested by the Global Games theory. However, it is reasonable to assume that firms (the targeted population in Sanchez Villalba (2015)) are sophisticated, as they are used to take strategic interactions into account when making financial, marketing, logistic, ... and *tax-related* decisions. Therefore, the theory would be a good predictor of behavior for firms. Moreover, a similar result

[27]Subjects were not asked what type of experiments they took part in, so previous experience may not have been useful for solving the decision problem of this experiment. Medical or psichological experiments, for example, usually do not provide much help in solving economic problems. I am grateful to Silvia Martínez Gorricho for pointing this out.

could be achieved if individual taxpayers had access to sophisticated professional advice, something that is indeed likely to occur (especially for wealthy individuals).

5 Conclusions

The empirical analysis of tax evasion is problematic because of the reluctance of both taxpayers and tax agencies to provide the relevant information. This study, therefore, uses experimental data as a second-best alternative and focuses on the testing of some of the theoretical predictions obtained in Sanchez Villalba (2015), though the richness of the dataset also allows for the investigation of other interesting hypotheses related to decision-making processes and the global game theory, so that the results found can be extrapolated to other similar games ("Global Inspection Games"), e.g., the allocation of welfare benefits or the awarding of bonuses based on peer-evaluations.

Results are strongly supportive of the main prediction of the GIG model, namely, that a tax agency using a *relative auditing strategy* would do better than if it used the standard *cut-off* one. The negative externality between taxpayers generated by the relative policy and the associated strategic uncertainty it creates seem to be the powerful forces behind this result.

Also supported by the data are the predictions derived from the comparative statics of global games: evasion is higher in E treatments than in C ones, evasion is a decreasing function of signals, and errors decrease with agency's "toughness".

The picture, so encouraging in qualitative terms, is however radically different when considering it quantitatively: in general, the numerical predictions of the theory are rejected by the data. This is true for the medium cases (when the signal is medium), but not for the extreme ones though: in the latter, data fit the predictions and support the idea that people are, at least, "Rudimentary" and (intuitively) understand the concept of dominance in simple scenarios. Medium cases, on the other hand, show that most people do not use higher-order beliefs when making their decisions (not even in this simple experiment, in which only two iterations are needed). In spite of this, many times they do choose the actions predicted by the theory of global games, usually after playing the game a few times. This "learning" result is not so surprising, as it was already hinted by Carlsson and van Damme (1993) and found experimentally by Cabrales et al. (2007). Other factors also seem to affect decisions, like the tension between the risk-dominant and payoff-dominant equilibria, with their predicted effects closely mimicked by the data. More worrying, however, is the apparently pervasive presence of a significant group of people ("chance maximizers") who choose their strategies without taking into account all the available information (in this particular experiment, the payoffs in different scenarios) and that lead to the largest differences between observed and predicted actions (treatment LC). This concern is connected to the main result derived from the analysis of questionnaire data, which suggests that those with knowledge of game theory are more likely to play according to predictions than those without that knowledge . This result can have an impact on policy-making, as one would expect higher degree of "sophistication" among firms than among individual taxpayers (though the latter group can change their status if they have access to sophisticated professional advice).

The bottom line is, therefore, that though people may not use higher-order beliefs, many times they end up choosing the same actions than the ones predicted by the Global Games theory. Consequently, this ensures that predictions are usually supported in qualitative terms (comparative statics and inter-treatment comparisons) but rejected in quantitative ones. Nonetheless, the latter problem can be deemed as a minor one because of two mitigating factors: First, the discreteness of the model can work against it because it amplifies small differences and thus make the data-predictions matches more difficult (something already highlighted by Heinemann et al. (2002)). And second, the parameters of the model were explicitly chosen to discourage said matches. Thus, the fact that the data does support (qualitatively) the predictions in these most demanding conditions suggests that the theory would be an even better predictor in more favorable environments.

References

Alm, J. and M. McKee (2004). Tax compliance as a coordination game. *Journal of Economic Behavior and Organization 54*, 297–312.

Bosch-Domenech, A., J. G. Montalvo, R. Nagel, and A. Satorra (2002, December). One, two, (three), infinity,...: Newspaper and lab beauty-contest experiments. *American Economic Review 92*(5), 1687–1701.

Cabrales, A., R. Nagel, and R. Armenter (2007). Equilibrium selection through incomplete information in coordination games: An experimental study. *Experimental Economics 10*, 221–234.

Carlsson, H. and E. van Damme (1993). Global games and equilibrium selection. *Econometrica 61*(5), 989–1018.

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

Harsanyi, J. C. and R. Selten (1988). *A General Theory of Equilibrium Selection in Games*. Cambridge, MA. MIT Press.

Heinemann, F., R. Nagel, and P. Ockenfels (2002, June). Speculative attack and financial architecture: Experimental analysis of coordination games with public and private information. *FMG Discussion Papers, LSE, UK* (DP 416).

Heinemann, F., R. Nagel, and P. Ockenfels (2004, September). The theory of global games on test: experimental analysis of coordination games with public and private information. *Econometrica 72*(5), 1583–1599.

Heinemann, F., R. Nagel, and P. Ockenfels (2009). Measuring strategic uncertainty in coordination games. *Review of Economic Studies 76*, 181–221.

Holt, C. and S. K. Laury (2002, December). Risk aversion and incentive effects. *American Economic Review 92*(5), 1644–1655.

Morris, S. and H. S. Shin (2002). Social value of public information. *American Economic Review 92*(5), 1521–1534.

Reinganum, J. and L. Wilde (1985). Income tax compliance in a principal-agent framework. *Journal of Public Economics 26*, 1–18.

Sanchez Villalba, M. (2015). Global inspection games. *Journal of Public Economics 128*, 59–72.

Selten, R., A. Sadrieh, and K. Abbink (1999). Money does not induce risk neutral behavior, but binary lotteries do even worse. *Theory and Decision 46*, 211–249.

Stahl, D. O. and P. W. Wilson (1994). Experimental evidence on players' models of other players. *Journal of Economic Behaviour and Organization 25*, 309–327.

A Instructions for treatment GC[28]

Introduction

First of all, thank you very much for taking part in this experiment. It is important to start by saying that, though part of a serious research programme, this experiment is NOT a test. There are no "right" or "wrong" answers.

How it works

Before we do anything, we have to run through a few ground rules and instructions. After that we will move to the experiment proper, where you will be asked to make decisions in a number of economic situations presented to you. Finally you will get paid: on top of a show-up fee of £5, you will get a sum of money that will depend on your performance in the situations mentioned before.

The experiment consists of 5 stages:

- Instructions
- Trial rounds
- Experiment rounds
- Questionnaire
- Payment

We will go through these in detail below.

Ground rules

For the experiment to work we need to run it according to fairly strict rules, but there are not too many:

[28] Instructions for the other treatments were similar to these ones, with the logical changes in rules and parameters needed in each case.

- From now until the end of the experiment, please do not talk (it will not take long!)
- If there is something you need to ask about the way the experiment works just raise your hand -the experimenter will come to your desk.
- Please do not use the computer until you are told to.

The Six Stages

1 Instructions

The experimenter will read out the instructions. If you have questions, this is the time to deal with them. Just raise your hand and the experimenter will answer them privately.

2 Short quiz

This is to ensure that you understand the instructions.

3 Trial rounds

The experiment is organised in a series of rounds. Each round is a period in which you interact –via the computer only– with the other participants and make decisions that determine the amount of money you will get at the end of the session.

As a warm-up you will first take part in 2 trial rounds. These trial rounds are identical to the experiment rounds in every respect with one exception: the effect on payment. Trial rounds do NOT affect your reward at the end of the experiment. They allow you to check out the interface and familiarise yourself with the screen tables, buttons and commands. They also allow you to make mistakes without losing money.

4 Experiment rounds

This is the real thing. What you do during these rounds will determine the total amount of money you will get.

The following "Frequently Asked Questions" will lead you through the basic mechanics of the rounds.

4.1. What is this all about?

Let us start by saying that the experiment will consist of 20 experiment rounds. In each one of them the computer will pair you up with one other participant. Each of the other participants in the room is equally likely to be paired up with you.

4.2. What do I have to do?

You have to choose one of two possible actions, namely Y or Z. You choose one or the other by clicking on your preferred option in the bottom left panel of the choice screen (see figure 1) and then pressing the "OK" button in the same panel.

Figure 1: Choice screen

4.3. How is my payoff for the round determined?

Your payoff for the round depends on your own action, the action of the other participant, and an unknown parameter called q.

4.4. But exactly how *is my payoff for the round determined?*

There are two cases to consider:

a. If you choose action Z, your payoff is 654 "experimental points" with certainty.

b. If you choose action Y, your payoff depends on both the value of q and the action of the other participant, as shown in the table below (and also in the top-left panel of the choice screen (see figure 1)):

		Value of q		
		A	B	C
Other participant's choice	Y	1000	655	579
	Z	658	651	0

That is, if you choose Z, you always get 654 "experimental points", regardless of what the other participant does and what the value of q is. But if you choose Y, then there are several cases to consider. Let us see some of them (remembering that in all of them you choose Y and your payoff is measured in "experimental points"):

If the other participant chooses Y and q equals A, then your payoff is 1000.

If the other participant chooses Y and q equals B, then your payoff is 655. And so on.

4.5. So how much money do I get then?

Your payoffs are transformed into money at a rate of: 1000 "experimental points" = 50 pence

That is, if your payoff for the round is, for example, 655 "experimental points", your corresponding money earnings are $655 \times 50/1000 = 32.75$pence.

Your session earnings are computed by adding up the money you got during the 20 experiment rounds.

4.6. But, what is q?

q is a parameter that can only take one of 3 values: A, B or C. In any given round, your computer will choose one of these 3 values, with probabilities 0.20, 0.60 and 0.20, respectively.

Intuitively, you can think of these probabilities in the following way: Consider an urn with 100 balls. 20 of them are labelled "A", 60 "B" and 20 "C". The value of q will be determined by the label of one of the 100 balls in the urn, chosen randomly (by the computer).

4.7. Is there anything I could use to make a more informed decision?

Yes, there is. Before you make a decision you will get a "hint". This hint will be known only to you and can only take one of 3 values: a, b or c. It provides some information about the value of the unknown parameter q, as shown in the following table (and in the top-right panel of the choice screen (figure 1)):

If hint is...	...then q is...	...with probability...
a	A	1.000
	A	0.125
b	B	0.750
	C	0.125
c	C	1.000

For any given round, your hint can be found immediately below this table in the choice screen (figure 1).

The table may seem a bit complicated but do not worry, it is not. It simply says that if your hint is equal to a, then you can be sure that q is equal to A. Analogously, if your hint is equal to c, then q is equal to C. When your hint is equal to b, however, you do not know for sure what the value of q is, but you can tell how likely each value is: q is equal to B with probability 0.750, while it is equal to A or C with probabilities 0.125 and 0.125, respectively.

Important note: Although q is the same for you and the other participant, your hints may differ from each other.

4.8. Anything else I should know before making my choice?

If you want to make some computations before choosing your action, you can press the calculator button on the choice screen (the small square button just above the darker area (see figure 1)). Pens and paper are available for those who prefer them: raise your hand and an experimenter will take them to your desk.

Also, it is worth mentioning that there is no "Back" button, so please make your decisions carefully and only press the "OK" or "Continue" buttons when you are sure you want to move to the next screen.

4.9. So I made my decision, what now?

After you submit your decision, you will be shown the action you chose and the payoff you got for the round, as well as the value that q took (see figure 2). By clicking on the "Continue" button you will move to a new round (if there is any still to be played).

Figure 2: Results screen

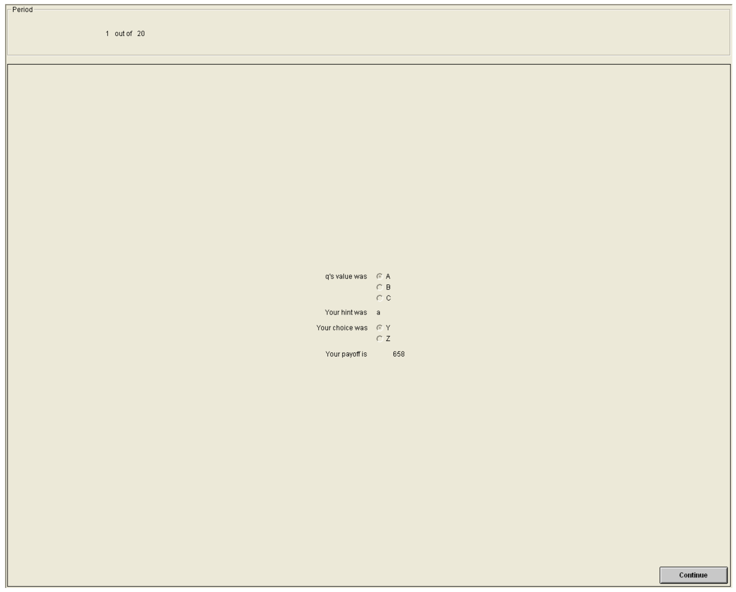

Figure 11:

4.10. And then? Is it the same over and over again?

Basically, yes. In every round, the structure is identical to the one described above: first a new q will be selected by the computer and you will be paired up with another participant, then you will be assigned a hint and will have to make a decision, and finally your payoff will be shown on the results screen.

You can check what happened in previous periods by taking a look at the darker area in the bottom-right panel of the choice screen (see figure 1). It includes information about the values adopted by q, the hints you got and the actions you chose in earlier rounds.

Important note: Every period is like a clean slate: the value of q, the participant you are paired up with and the hint you get may vary from round to round, but the RULES that determine them (explained in questions 4.6., 4.1. and 4.7.) do not. In short, rounds are independent: for example, you can think that in every round a new urn with 100 balls -20 “As”, 60 “Bs” and 20 “Cs” - is used to determine the value of q, as explained in question 4.6. Similarly, the pairings and hints of a given round are independent of the pairings and hints of previous rounds.

5 Questionnaire

We will ask you a few questions that will help us to further understand the data collected in the session.

6 Payment

Finally! You will be paid a show-up fee of £5 plus the sum earned during the session, as explained in question 4.5.

And that is it. Once again, thank you very much for participating!

SHORT QUIZ

1. What is your payoff (in "experimental points") if you choose Y, the person paired-up with you chooses Z and q is equal to A? . . . .

2. What is your payoff (in "experimental points") if you choose Z, the person paired-up with you chooses Y and q is equal to C? . . . .

3. If your hint is equal to b, what is the probability that q is equal to A? . . . .

B Rudimentary and Sophisticated Dominance

Let us start by defining the concepts of Soft, Medium and Tough games, which are simply the games played by the members of a group when the agency is soft, medium and tough, respectively (*i.e.*, they are like the game shown in table 2, with $q \in \{A, B, C\}$). Clearly, these games $g \in \mathcal{G} := \{S, M, T\}$ depend on the type of the agency, and so both g and q are subject to the same probabilistic process.

Based on this taxonomy of games and on the conditional probability distribution of q (shown in table 3), two different scenarios can be identified: one in which the signals give perfect information about the game being played (when $s = a$ or $s = c$), and another one in which precision is less than perfect (when $s = b$).

In the first iteration, therefore, a player who receives a soft signal ($s = a$) knows for sure that she is playing the Soft game ($g = S$). Furthermore, because of equation 6, she can immediately realize that Evasion *strictly dominates* Compliance, the very result indicated in part 1 of proposition 1. Following a similar argument and using equation 7, part 3 is also proved.

When the signal is medium ($s = b$), though, the person does not know the actual game g that is played, but she does know its conditional probability distribution $\Pr(g(q)|b) = \Pr(q|b)$. Thus, the game that she faces is depicted in figure 12, and her expected utility from evasion is given by equation 9, where s is replaced by b. This expression is an increasing function of the beliefs $k'(s')$, $\forall s' \in \mathcal{S}$, because of the nature of the relative policy (equation 5). The *worst-case scenario* for the optimizing person occurs, therefore, when $\mathbf{k}'(\mathbf{s}') = \mathbf{0}$, $\mathbf{0} := (0, 0, 0)$, such that the expected utility from evasion is $Eu\,(Y, \mathbf{0}|b)$. Analogously, the *best-case scenario* occurs when $\mathbf{k}'(\mathbf{s}') = \mathbf{1}$, $\mathbf{1} := (1, 1, 1)$ and expected utility is $Eu\,(Y, \mathbf{1}|b)$. It is not difficult to see that the no-strict-dominance condition of proposition 1 (part 2) requires

$$Eu\,(Y, \mathbf{0}|b) < u(Z) < Eu\,(Y, \mathbf{1}|b) \tag{13}$$

and if it is satisfied, a Rudimentary player will act as predicted by proposition 1.[29]

[29] Alternatively, this equation can be interpretated as follows. Let us construct a new, artificial 2x2 game

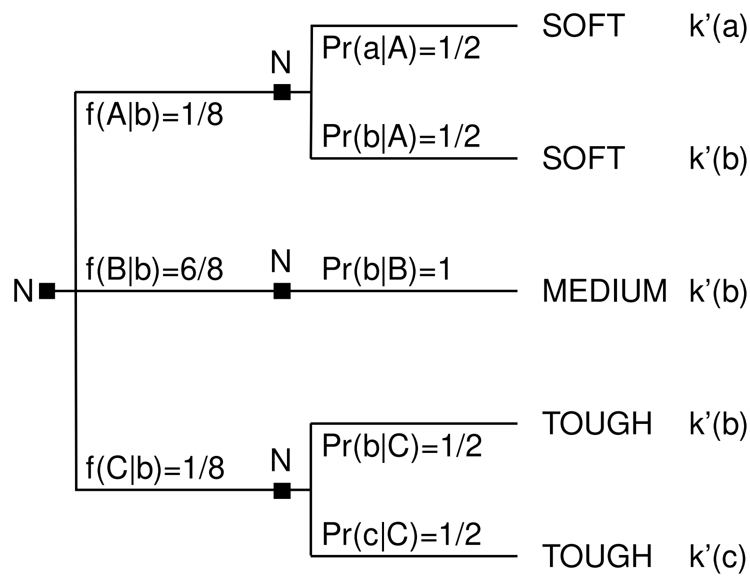

Figure 12: Game tree if signal is medium ($s = b$)

A Rudimentary player would stop her analysis here, but the Sophisticated one will continue to the next iteration. Furthermore, the sophisticated player will realize that, if the other member of her group is (at least) Rudimentary, then (by symmetry) she would have also worked out that Evasion (respectively, Compliance) is the strictly dominant strategy when the signal received is soft (a) (respectively, tough (c)). Formally, the sophisticated player's beliefs about the other person's choices will have precise numbers attached to them, namely, $k'(a) = 1$ and $k'(c) = 0$. The expected utility will reflect this: $Eu\left(Y, \left(1, k'\left(b\right), 0\right)|b\right)$ and *new worst-* and *best-case scenarios* can be computed: $Eu\left(Y, \mathbf{c}|b\right)$ and $Eu\left(Y, \mathbf{e}|b\right)$,where $\mathbf{c} := (1, 0, 0)$ and $\mathbf{e} := (1, 1, 0)$.

Depending on the position of the safe utility $u(Z)$ with respect to the latter two, three cases

like the one in table 2, but which is a weighted average of the Soft, Medium and Tough games defined above, $A := \sum_{q\in\mathcal{Q}} f\left(q|b\right) \cdot g\left(q\right)$, so that the corresponding (expected) utility in each of its cells is

$$u\left(d, d', E\left(q|b\right)\right) := \sum_{q\in\mathcal{Q}} f\left(q|b\right) \cdot u\left(d, d', q\right) \tag{14}$$

It can then be shown that $u\left(Y, Z, E\left(q|b\right)\right) = Eu(Y, \mathbf{0}|b)$, $u\left(Y, Y, E\left(q|b\right)\right) = Eu(Y, \mathbf{1}|b)$, and $u\left(Z, Y, E\left(q|b\right)\right) = u\left(Z, Z, E\left(q|b\right)\right) = u(Z)$, so that equation 13 implies that this "Average game" is a coordination game.

can arise, of which we are interested only in the following two:[30,31]

$$
\begin{aligned}
Eu\left(Y,\mathbf{c}|b\right) &> u(Z) && (17)\\
u(Z) &> Eu\left(Y,\mathbf{e}|b\right) && (18)
\end{aligned}
$$

The first one indicates that even in the *new worst-case scenario*, the expected utility from Evasion is higher than that of Compliance or, equivalently, that Evasion *strictly dominates* Compliance. The second one, on the other hand, implies that, even in the *new best-case scenario*, the expected utility from Evasion is lower than that of Compliance, and so that Compliance *strictly dominates* Evasion.

By definition these two conditions are mutually exclusive, and which one of them is satisfied determines the player's optimal strategy: either $\boldsymbol{\sigma}^* = (Y, Y, Z)$ if equation 17 holds or $\boldsymbol{\sigma}^* = (Y, Z, Z)$ if the one that holds is equation 18. These strategies are of the "threshold" type (Heinemann et al. (2004), Heinemann et al. (2009)) but can be indexed by their second component, which is the only one that differentiates one strategy from the other and corresponds to the optimal choice when the signal is medium, $\sigma^*(b)$. The value of this component, therefore, is the one that defines the Evasion, $\sigma^*(b) = Y$, and Compliance, $\sigma^*(b) = Z$, treatments. This is exactly what states the second part of proposition 2.

C Classification of subjects based on questionnaire data

The questionnaire also asked participants about the strategies they followed and the rationale behind them. This information was then used to classify them according to some stylized characteristics, in a fashion similar to the one used by Bosch-Domenech et al. (2002). The distribution of subjects in terms of categories and treatments is shown in table 16.

The different categories are defined as follows:[32]

Expected payoff maximizers (EPM): Those who indicated they played Y (Z) in E (C) treatments, based on expected-payoff maximisation. Note that this category includes everyone who played according to the OS strategy, even though they did not use HOBs.

[30] The third one does not lead to a unique solution, which goes against the spirit of the theory of global games. The reason for the non-uniqueness is the discreteness of the model. Having continuous choices may have avoided this problem, but at the cost (considered to be too high) of increasing the complexity of the game and thus the noise in the observations.

[31] For L-treatments, the analysis is greatly simplified since other people's choices do not affect one's decisions. Then, the equivalents of equations 17 and 18 are, respectively,

$$
\begin{aligned}
Eu(Y,b) &> u(Z) && (15)\\
u(Z) &> Eu(Y,b) && (16)
\end{aligned}
$$

[32] Appendix D shows literal transcripts of questionnaire comments made by some subjects that are characteristic of each one of these categories.

Category	GC	GE	LC	LE	All
Expected payoff maximizers (EPM)	10/11	8/11	5	5/13	28/40
Chance maximizers (CM)	1/2	0/3	6/7	0/8	7/20
Learners (L)	0	3	1	1	5
Mixers/Experimenters (M/E)	1	2	0	2	5
Non-independent (NI)	1	0	4	3	8
Randomizers (R)	1	2	1	0	4
Confused (C)	1	0	1/2	1	3/4
Risk-lovers (RL)	2	0	1	0	3
All	18	18	20	20	76

Note: Cells with two numbers separated by "/" reflect uncertainty about the allocation of some subjects to specific categories.

Table 16: Questionnaire. Classification of subjects.

Chance maximizers (CM): Those who only considered the probabilities of outcomes being higher or lower than the safe option, without weighting them using the associated payoffs.

Learners (L): Those whose decisions varied in the first periods, but chose always the predicted action afterwards.

Mixers/Experimenters (M/E): Those that deviated just once or twice from the predictions of the OS hypothesis but, unlike the Ls, did so at times other than the first periods (Experimenters). An alternative rationale could be that they followed a strategy such that they evaded and complied with probabilities that usually replicated the relevant odds ($(1/8, 7/8)$ in C treatments and $(7/8, 1/8)$ in E ones), and so could be labelled "Mixers".

Non-independent (NI): Those who (despite the instructions clearly stating that rounds were independent from each other) followed some kind of history-dependent strategy.

Randomizers (R) (also "Guessers" (G)): Those who chose randomly between Y and Z.

Confused (C): Those who seemed to be (or acknowledge they were) confused.

On top of these strategies, the degree of risk aversion is expected to play a role as well. In particular, risk aversion fosters compliance (*ceteris paribus*) and hence makes the Global Game predictions easier to be satisfied in C treatments, but works against them in E ones. Combining the strategies defined above and the degree of risk aversion, one can usually categorize all subjects and find some interesting stylized facts.

The first one that can be stated is that categories seem to order themselves in three "Dominance bands" according to their degree of coincidence with the Global Game predictions (see figures 13, **??**, 14 and **??**). Near the top we can find the $EPMs$ (high dominance). In the middle-ground there is a mixed bag of types (M/E, L, NI and C) who chose different actions in different periods, even though they always got the same signal b. Risk lovers (RL) are close to the top in E treatments and to the bottom in C ones, and the opposite is true for risk averse (RA) people.

All these results, however, are not surprising. The category that is really exciting to analyse in detail, on the other hand, is that of the CMs, since it seems to be behind the case with the largest deviations from predictions (that is, the LC treatment). Now, the first thing to notice is that in some cases CMs cannot be distinguished from $EPMs$, because the observed data are consistent with the predictions of both criteria (expected-payoff and probability maximisation) and the questionnaire information is vague (this is the rationale for the ambiguity in table 16). For this very reason, the most interesting scenarios are those where the two criteria prescribe different actions, as is the case in C treatments (the *Global Game theory/Expected Payoff Maximization* predicts Compliance, *Chance Maximization* predicts Evasion). Focusing on these treatments, it can be seen that significant deviations from the Global Game predictions take place, thus confirming the results of the tests that compare the levels of dominance in C and E treatments (table 10). Also, since *Chance Maximization*'s prescription to evade depends on what the other person does in GC but not in LC, it is not surprising that the degree of dominance in the former is greater than in the latter: the uncertainty about the other person's action in GC works against the incentives to evade and (as seen in figure 13) only RLs end up evading all periods. Since this interdependence does not play a role in LC, the number of subjects that evade all periods is far greater (and most of them are CMs –see figure 14), and explains the huge divergence between predictions and data (and confirms the ranking of treatments according to Dominance found in the previous section).

D Examples of categories

Expected-Payoff Maximizers (EPM): "If the hint was a, I selected action Y; otherwise, I selected action Z. There are only three outcomes that generate more than 654 points, and two of them only generate a negligible increase (relative to their risk). The only way to "gamble and win" is to play Y when the hint is b, and in that case, I am gambling that either my "opponent" has a hint of a (very unlikely), or my opponent has a hint of b, is risking that q is really A, and is right (also very unlikely). My risk is that my opponent plays Z, which is safer, and that q is B or C, which is likely. The risk/reward is far too high. When my hint is a or c the correct play is obvious - in the former case, playing Y always nets me more than 654, and in the latter, playing Y always nets me less than 654, no matter what my opponent does." (Subject #10, GC).

Chance Maximizers (CM): "If the hint is c, the best decision is always Z with a higher payoff. If the hint is b, it worths choosing Y, because thereis a probablility of 0.875 getting A or B, which are both higher than Z(654). If the hint is a, my decision is definitely Y." (Subject #20, LC).

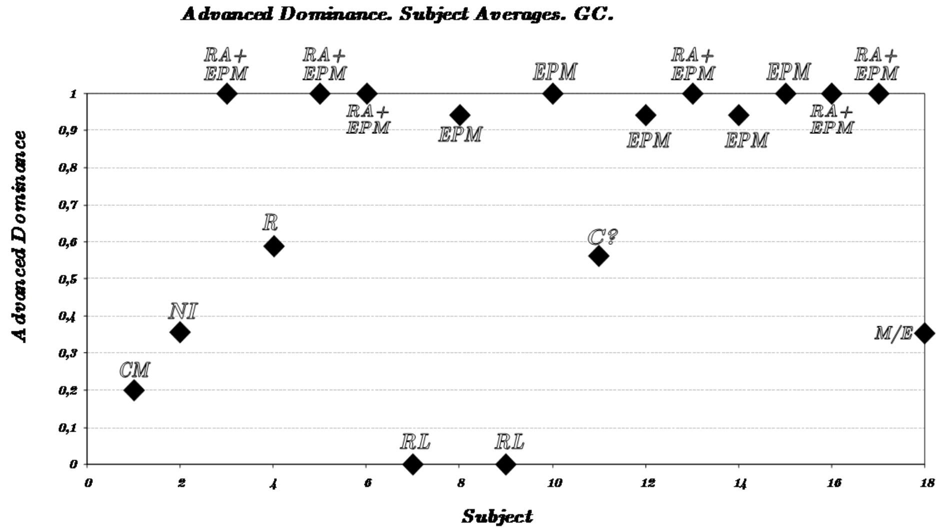

Figure 13: Advanced Dominance. Subject averages. GC treatment.

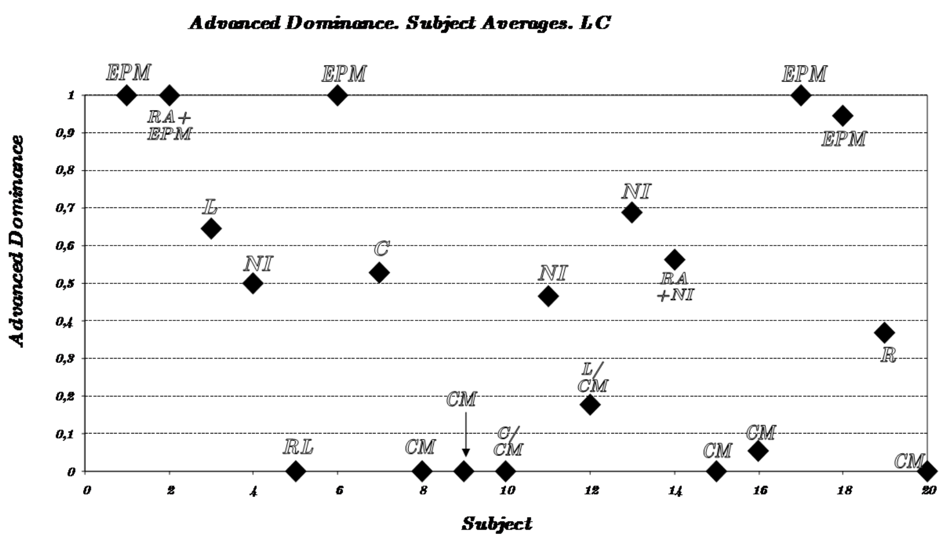

Figure 14: Advanced Dominance. Subject averages. LC treatment.

Learners (L): “At first i played it safe and went with the guarantee button z and then i took more of a risk by chosing the y button every time i got the hint "a" or "b". because there was a higher probability of gaining more points.” (Subject #18, GE).

Mixers/Experimenters (M/E): “If the hint was A, choice was Y. If the hint was C, choice was Z. If the hint was B, 80% of the time choice was Y and 20%, B.” (Subject #15, LE).

Non-independent (NI): “If the hint came up as A i always selected choice Y as I would be better off (ie gaining more money) through doing so regardless of what the other participant chose. Conversely, if the value of q was C i always chose Z since I would be worse off if i choice Y despite what the other person selected. If the value of q came up as b i would go systematically throught the choices Y,Y,Z. This was my order since if q=b and q=a i would be better off selecting Y and if q=c i would be better off selecting Z. Since the probability of q=b was the highest i put Y at the beginning of the order. I used my knowledge of maths and probablities to calculate the order in which to place my choices.” (Subject #2, GC).

Randomizers (R): “If the hint was a then i chose Y if the hint would have been c then i would have chosen Z. apart from this i just guessed randomly. the last 3 i thought i may as well take the risk as it was the end of the experiment.” (Subject #19, LC).

Confused (C): “If the probability was lower than the other option, i chose the other option. I did not take risks in the cases where the probability could also go for the lowest amount. Becasue i dont know much about the probability theory so i decided to go for the safest method.” (Subject #7, LC).

Risk-lovers (RL): “I chose Y every time unless I knew it was C. I was not given the hint a at any time. The difference between playing it safe and gambling with the Y option was small enough to make the experiment slightly more fun. I knew that I could lose 579, but only gain 421, but preferred the gamble.” (Subject #7, GC).

E Extra Tables

	Da	Db	Dc	RD	AD	SD
LC	.	0.0000	.	1.0000	0.0000	0.0000
LE	0.3231	0.0000	.	0.3180	0.0000	0.0000
GC	.	0.0000	0.1578	0.1552	0.0000	0.0000
GE	.	0.0000	0.0390	0.0418	0.0000	0.0000
LC=GC	.	0.0000	0.1578	0.1552	0.0000	0.0000
LE=GE	0.3231	0.0042	0.0390	0.1920	0.0042	0.0029
LC=LE	0.3231	0.0000	.	0.3180	0.0000	0.0000
GC=GE	.	0.0005	0.2170	0.4359	0.0005	0.0014

Note: Values of F-tests. Values below 5% imply the null hypothesis is rejected. Dots mean there is no variability in data as to compute the statistics.

Table 17: Dominance tests. Predictions and inter-treatment comparisons.

	ERRA	ERRB	ERRC	ERR
LC	0.3456	0.0000	0.0000	0.0518
LE	1.0000	0.1450	0.0004	0.3515
GC	0.2939	0.0000	0.0000	0.0147
GE	1.0000	0.1450	0.0010	0.2445
LC	0.0000	0.0000	0.0000	0.0000
LE	0.0038	0.0000	1.0000	0.1441
GC	0.0092	0.0000	0.1575	0.0000
GE	0.0006	0.0000	0.2612	0.0092
LC=GC	0.9556	0.0000	0.0000	0.0000
LE=GE	0.1082	0.0000	0.0034	0.0000
LC=LE	0.0000	0.0000	0.0000	0.3555
GC=GE	0.0135	0.0000	0.2160	0.0112

Note: Top panel: Predicted values of dependent variable. Middle and bottom panels: Values of F-tests. Values below 5% imply the null hypothesis is rejected.

Table 18: Errors tests. Predictions and inter-treatment comparisons.