Ambiguity in Criminal Punishment

Timothy C. Salmon†
Southern Methodist University

Adam Shniderman‡
Texas Christian University

January 2017

Abstract

There has been substantial prior field research on how the incidence of criminal behavior responds to changes in the probability of punishment. The results of this research is at best mixed in regards to its conformance to the predictions of standard expected utility theory. One possible cause for these mixed results is that punishment probabilities in the field are ambiguous rather than uncertain. The presence of this ambiguity could in part explain some of these conflicting results. As a step towards investigating this link, we conduct an experiment on tax compliance intended to try to understand how individuals respond to ambiguous punishment probabilities and in particular to how they respond to shifts in ambiguous versus known probabilities. We find that when probabilities are known and shift, the standard model works well to explain the response. When the probabilities are ambiguous and shift, the behavioral response is minimal. We also use these experiments as a means of testing whether ambiguity aversion might be present in sufficient degree to be exploitable in how enforcement procedures are advertised to increase their effectiveness at minimal cost. We find at best weak evidence in favor of ambiguity aversion and thus little support for the notion that enforcement regimes could take advantage of ambiguity aversion.

JEL Codes: C91, K42

Key Words: Ambiguity; Criminal Behavior; Deterrence

1 Introduction

Policy makers have three well known levers to use in attempting to deter crime - the severity, certainty and celerity (speed) with which punishments are enforced. There has been substantial research regarding which of these levers is most effective at deterring crime. There is often, however, little agreement in the scientific literature on even the most basic of issues

†Funding for this project was provided by the TCU Research and Creative Activities Fund and by SMU Dedman College of Humanities and Sciences. We thank seminar participants at the 2016 North American Meetings of the ESA and the 2017 ASSA Meetings for many helpful suggestions. We in particular thank our discussant at the ASSA meetings, Sera Linardi, for a very useful report. Approval for all experiments in this paper was obtained from the IRB’s at both SMU and TCU. All subjects signed consent forms prior to participation.

‡Southern Methodist University, Department of Economics, 3300 Dyer Street, Suite 301 Umphrey Lee Center, Dallas, TX 75275-0496. tsalmon@smu.edu, Phone: 214-768-3547, Fax: 214-768-1821.

‡Texas Christian University, Department of Criminal Justice, 2855 Main Drive, Scharbauer Hall, Suite 4200, Fort Worth, TX 76129. adam.shniderman@tcu.edu, Phone 817-257-6420.
related to crime deterrence. This study is aimed at trying to better understand one potential explanator for that disagreement, the presence of ambiguity in regard to punishments, and to investigate the nature of decision making in the presence of ambiguity.

A rational choice model of criminal behavior involves a potential criminal weighing the benefits of a criminal act against the potential consequences. Becker (1968) laid out the rational choice model that has served as the basis for much of the subsequent work on crime deterrence. If one assumes a risk averse individual, this standard model predicts that making the punishment for a crime more severe would have a stronger deterrent effect than increasing the probability that the individual will be apprehended or that the punishment will be realized. Perhaps paradoxically, Becker and the criminologists whose work is rooted in the Beckerian model have often concluded the opposite to be true based on field investigations. They often find that changing the severity of punishment has little deterrent effect (Decker and Kohfeld (1990); Doob and Webster (2003); MacCoun and Reuter (1998); Nagin (1978); Nagin (1998); von Hirsch, Burney, and Wikstrom (1999)), while increasing the probability of apprehension is sometimes found to have a deterrent effect (Eide (2000); Witte (1983)).

However, this issue remains unsettled as other studies often find no effect or even the opposite effect (Huff and Stahura (1980); Humphries and Wallace (1980); Loftin and McDowall (1982)). Despite the mixed evidence, many criminologists have concluded that changes in certainty of apprehension have a much greater deterrent effect than changes in the severity of punishment (Durlauf and Nagin (2011); Nagin and Pogarsky (2001); Paternoster (1987); Witte (1980); Zimring and Hawkins (1973)).

Our focus in the current study is to try to better understand the effects of changing the probability of apprehension on behavior to see if we can provide an explanation for some of these mixed results. While there have been a large number of studies on this issue, there is still no clear conclusion regarding whether or not such policies are effective. Some of the more prominent examples of these studies involve examining policing policies, such as attempts to increase the size of the police force, which one imagines increases the probability that criminals will be caught and prosecuted. Beginning in 1994, President Clinton took steps to add 100,000 new officers by the year 2000 through the Community Oriented Policing Services (COPS) Office. Over the course of his presidency, President Clinton awarded nearly $10 billion in federal grants to police agencies in all 50 states to put more police on the streets. This policy was implemented with great confidence regarding its ability to reduce crime, however, its actual effects have been the subject of much debate and controversy. In short, the results on the effectiveness of the policy are mixed at best.

Eck and Maguire (2000) reviews 27 studies examining the effects of changing the number of police on crime rates and found no consistent relationship between the number of police on the streets and the levels of violent crime. Studies in this sample provided evidence suggesting both that increasing the number of police is associated with a reduction in crime and that increasing the number of police is associated with an increase in crime. However, these studies suffer from endogeneity problems, making it difficult to clearly identify a causal relationship between the policy changes and the incidence of crime. There have also been several field experiments conducted on this issue that one would expect would resolve the causality questions. These field experiments have generally found no evidence that increased police presence deters criminal conduct. The Kansas City Preventive Patrol Experiment examined the effect of changes in police patrol (in vehicles) on crime rates. The
experiment found that changes in police strength had no significant effect on crime rates (Kelling, Pate, Dieckman, and Brown (1974)). The Newark Police department examined the effect of changes in foot patrols on crime rates and residents’ perceptions of crime in the city. The results in Newark were identical to those in Kansas City – changes in foot patrols had no impact on the crime rate (Sherman, Gottfredson, MacKenzie, Eck, Reuter, and Bushway (1998)). However, as George Kelling, director of both studies, notes, these experiments examine the impact of particular policing techniques on crime rates, rather the number of police (Kelling, 1995). It therefore isn’t clear that these changes should have had much of an impact on the probability of a criminal being caught. Additionally, these experiments have been criticized for both design, implementation, and evaluation shortcomings (Sherman, Gottfredson, MacKenzie, Eck, Reuter, and Bushway (1998)).

Taking an alternative approach, Levitt (2002) employs a variety of instrumental variables to try to obtain a causal result. The instruments for police strength tested include mayoral and gubernatorial election cycles as well as firefighter hiring. However, Kovandzic and colleagues have criticized his use of these techniques and proxy measures (Kovandzic et al., 2016). They contend that the use of mayoral and gubernatorial elections are weak instruments of police growth and therefore are unreliable for use in assessing the impact of police levels on crime.

There is also a literature that attempts to use lab experiments to examine how behavior responds to shifts in enforcement regimes, which should be expected to have high levels of internal validity and allow researchers to be able to observe clearly the behavioral mechanisms behind any effect from changing probabilities in punishment. While this too is a substantial literature with contributions across several decades, (Hill and Kochendorfer (1969); Schildberg-Hörisch and Strassmair (2012); Mungan and Klick (2015); Nagin and Pogarsky (2003); Vitro and Schoer (1972)), the picture one gets from this literature is that identifying these effects of increasing punishment probability on crime incidence is quite difficult and that to date, no clear conclusions can be drawn. DeAngelo and Charness (2016) provides a more recent examination of how changing enforcement regimes affects behavior with a special focus on how the fact that individuals have chosen an enforcement regime respond while being monitored in the way they chose. They find that individuals do respond in the manner predicted by the Becker model though they also respond to complexity in how the enforcement regime is explained by transgressing less often than when the enforcement regime is explained in a clearer manner.

In addition to the endogeneity problems in identifying these effects, we propose that there is a deeper specification issue that may be driving some of the conflicting field results which has to do with the model of decision making used at the foundation of deterrence research. The Becker model of criminal behavior is based on a classical model of decision making under uncertainty. In this model, an agent chooses among actions knowing that conditional on his choice, different events may then occur according to specific and known probabilities. An important difference between this model and the world a potential criminal operates in is that the probabilities of getting caught and punished are not well defined, and may even be unknowable. This changes the decision making paradigm to one involving ambiguity, rather than uncertainty. The key difference between the two environments is that in an uncertain environment a decision maker knows the probabilities with which events will occur, while in an ambiguous environment the decision maker does not
know these probabilities. The latter seems quite a reasonable assumption about the way criminal justice operates. In the face of ambiguity such as a criminal not understanding the probability of getting caught or not understanding how an increase in police enforcement could change that probability, it isn’t clear that one should observe the classical response predicted by the Becker model. Individuals could respond to the greater complexity in the choice environment by ignoring shifts they don’t understand or perhaps, in a manner that could be considered consistent with the results in DeAngelo and Charness (2016), by over responding to the presence of the additional complexity.

The notion that individuals might over respond to the complexity of an ambiguous choice environment is accounted for in part by a concept known as ambiguity aversion (Gilboa and Schmeidler (1989); Camerer and Weber (1992); Fox and Tversky (1995); Keren and Gerritsen (1999)). This concept is best explained using a canonical experiment known as the Ellsberg paradox (Ellsberg (1961)). As a demonstration of the core idea, consider an urn consisting of 30 red balls and then 60 additional balls that are either black or yellow but in unknown proportion. If you ask an individual to first bet on whether a ball drawn at random will be black or red then that will measure the person’s belief regarding whether the number of black balls is greater than or less than the number of red, which is 30. If the individual chooses to bet on red, then that suggests that they must think there are fewer than 30 black balls in the urn. If the person believes there are fewer than 30 black balls then he must believe there are more than 30 yellow balls in the urn. Consider then offering that same individual the option to bet on a draw from that same urn as to whether the draw will be red or yellow. They should now choose to bet on yellow as they previously revealed a belief that there are more yellow balls than red. If they instead again bet on a draw for a red ball then the person is deemed to be ambiguity averse. This notion of ambiguity aversion essentially involves an individual always forming a pessimistic belief about an event for which they do not know the probability. Substantial evidence exists showing that individuals express preferences of this sort (Becker and Brownson (1964); Butler, Guiso, and Jappelli (2014); Curley, Eraker, and Yates (1984); Dimmock, Kouwenberg, Mitchell, and Peijnenburg (2016); Easley and O’Hara (2009); Fellner (1961); MacCrimmon (1968); Hoy, Peter, and Richter (2014); Oechssler and Roomets (2014); Slovic and Tversky (1974)).

The reason that ambiguity aversion is important is that if individuals do act according to such preferences then the manner in which the increased presence of capable guardians is presented to individuals can alter the effectiveness of interventions on deterrence. For example, if it is generally known that the probability of getting caught for a crime has some minimum and maximum values then an ambiguity averse individual’s choice is likely to be most impacted by the maximum value. Policies aimed at adjusting the average or minimum value should be expected to have limited effectiveness since in an ambiguous situation, agents won’t be able to accurately assess the average and their pessimism would lead them away from paying attention to the minimum. Alternatively, the deterrence effect of a program is maximized by raising the possible upper end of enforcement possibilities even if the average isn’t shifted. If ambiguity is the correct depiction of the field choice environment then that insight could be helpful in designing enforcement regimes. In regard to the field evidence, the lack of a consistent response to policies aimed at shifting enforcement probabilities could be attributed to ambiguity as potential criminals may be paying attention to things other than what the researchers expected in choosing how to respond to shifts in
enforcement regimes. In this study, we will use a laboratory experiment that will examine how individuals respond to changes in apprehension probabilities when those probabilities are known and when they are ambiguous. The experiment will also examine the degree to which ambiguity aversion might drive behavior in the ambiguous environment. Although we noted that prior research has established that there are situations in which individuals will behave in a manner consistent with ambiguity aversion, these studies are mostly framed in regard to potential gains. When dealing with criminal punishment, the issue should be in regard to potential losses and it isn’t clear that individuals will treat gains and losses similarly. There is in fact other recent evidence that suggests substantial skepticism on the broader applicability of ambiguity aversion, Kocher, Lahno, and Trautmann (2015), which is certainly an important counter to the prior literature suggesting strong support for the notion. We use a laboratory experiment as the basis for our examination as it is necessary to obtain the clear identification we need of how changes in detection probability affect behavior. We see this as an initial small scale proof of concept that is necessary to establish the validity of these ideas before one would find it worthwhile to engage in testing equivalent policy changes in the field.

Our experiment is modeled on a situation involving potential tax evasion. Our subjects engage in a task for which they earn piece rate pay. Earnings are subject to a tax that is pure deadweight loss to the subjects. We allow the subjects to self-report their earnings with any reported earnings being taxed at a high rate. The self-reporting option allows them to misreport their earnings to obtain a lesser tax bill. We include the possibility that their report may be audited with misreported earnings incurring a fine. In our treatments, we vary what the subjects know regarding the audit probability. In some cases they know the audit probability and thus they are dealing with uncertainty. In other cases they only know the set from which the probability might be drawn but they have no information about how the probability is drawn from that set. This latter situation is one with ambiguity. Our findings will show that subjects respond as predicted to changes in known probabilities which provides vindication for the classical Becker model and is consistent with the results from DeAngelo and Charness (2016). When probabilities shift in an ambiguous way, we find no clear behavioral response. This is consistent with the field studies which also provide evidence of a muddled response to changes in enforcement regimes. Finally, we examine the behavior of the subjects to determine if they exhibit loss aversion in an exploitable degree but we find only weak and limited support for loss aversion which is contrary to most of the prior literature on ambiguity aversion but roughly consistent with the results from Kocher, Lahno, and Trautmann (2015).

2 Experimental Design

We seek to determine how ambiguity in the probability of apprehension affects behavior. This involves allowing subjects to engage in some form of transgressive behavior leading to possible losses of earnings. To make certain that subjects see the losses as real, the experiment design needs to maximize the subjects feelings of entitlement to the money they risk losing through apprehension and punishment. Consequently, our experiment involves subjects engaging in a real effort task to earn money and includes an element in which they
could lose some of their earnings in a fine for engaging in dishonest behavior. We explain each of these elements, in turn, as well as our treatments.

There are two types of subjects in our experiments: workers and tax collectors. There is a single tax collector per session; all other subjects are workers. In each round of the experiment, workers engage in a task for piece-rate earnings. This task is based on one used in Erkal, Gangadharan, and Nikiforakis (2011) and then in Ku and Salmon (2012). Subjects are asked to take random sequences of 4 letters and transform them into a numerically coded version using a provided code. Every encoding workers complete correctly earns them 1.5 ECU (Experimental Currency Units) which translated into $1.50 as the exchange rate in the experiment was 1 ECU=$1. Workers would have 4 minutes to complete as many encodings as they wished to in the production phase of a period.

At the end of the production phase, workers are asked to report to the tax collector how much they earned. Workers are allowed to enter any amount they wish. The amount they enter is subject to a 40% tax. Therefore, subjects could under-report their earnings to save on their tax bill. The downside to doing so is that there is a chance that the worker might be audited. This chance and how it is presented to the subjects depends on the round and the treatment, further explained below. If a worker is audited and found to have reported earnings at least as high as he or she actually earned, they simply pay the 40% of their reported earnings in taxes. If a worker under-reports, then he or she must pay the entire tax bill owed plus a fine - equal to the difference in the tax they claimed to owe and the amount they actually owed multiplied by 1.25. If a worker is not audited in a round, then they simply pay the 40% tax on whatever income they chose to report. The tax revenue is deadweight loss to the subjects. It is not redistributed in any way. Of course real taxes are collected by a government and spent on programs that do distribute the money in some way back to the populace and so real tax fraud could be seen as harming individuals. Since our experiment doesn’t possess this element, this might lead to more tax fraud than if the taxes benefitted other subjects in some way. While this is true, that effect should be expected to be orthogonal to the treatments and therefore not impact our comparative statics which are the measurements of interest, not the overall level of tax fraud.

The role of the tax collector is largely passive. He or she doesn’t perform the audits but the results of all audits performed in a round are shown to the tax collector. The only action of the tax collector is to then press a button labeled “Levy all Fines.” The purpose of the subject in this role is so that when workers lie about their earnings, they see it as lying to another subject in the experiment, rather than the experimenter. At a minimum, workers will know that another subject will observe their lying, so that it might increase the possibility that they see it as a moral transgression. The tax collector is paid a fixed fee of 10 ECU for their performance of this job plus they are also allowed to complete encodings at a reduced rate of 0.5 ECU per encoding. These earnings are not subjected to taxes.

The nature of audits depend on the treatment as outlined in Table 1. Subjects participate in only one treatment. Our first treatment is the Known Probability treatment (Known). In it there were three possible audit probabilities, 30%, 40% and 50% with the probability being associated with a round color. The color of a round was determined at random and the color could be either blue, red or green. The audit probability is 30% in blue rounds, 40% in red and 50% in green. In this treatment, subjects are told the round color when making their income report and so they know exactly what the audit probability
is when making their choice. They do not know the audit probability in the production phase, only at the income reporting phase.

In our Ambiguous Probability Treatment (Ambiguous), again there are three possible round colors. In red rounds, the audit probability is known to be 40%. In green and blue rounds, subjects know that all blue rounds have a common probability as do all green rounds with one being 50% and the other 30%. We do not inform them of which is which nor did we provide them any explanation of a mechanism that would have made the choice. Thus the audit probability is ambiguous in rounds with these two colors as subjects do not know the probability of the audit nor do they know the likelihood of any particular audit probabilities.

Our final treatment is called the Ambiguous Shift Treatment (Shift). In this one, the subjects face ambiguity in all rounds. In Red rounds the probability is either 10%, 40% or 70%. In blue the possibilities are 30%, 40% and 50%. Green rounds have an audit probability of either 30%, 50% or 70%. Again, subjects are told that one of the three possibilities has been chosen in advance for each color but they have been provided no details regarding how those choices were made. The idea in which this treatment represents a shift in ambiguity is to consider red rounds a baseline condition in which the probability of punishment is 10%, 40% or 70%. If one wishes to increase the enforcement level, one possible shift would involve keeping the potential average probability the same but increasing the lowest possible punishment percentage. This would involve the change in probabilities to the blue rounds in which the average is still 40% but the lowest possible probability is now 30% with the top probability falling. Alternatively, one could shift up the overall possibility of punishment as in the green rounds by moving the bottom and middle probabilities but leaving the top probability unchanged. As we will explain below, these two shifts allow us to identify which element subjects pay attention to when making their income reporting decisions.\(^1\)

In all treatments, subjects earn income and report earnings in 9 rounds. While round colors were randomly ordered between subjects, each subject faced each round color three times. At the end of the experiment one of the 9 rounds was chosen at random to generate final payment to the subjects. Subjects earned on average $30.24 including a $10 show up fee. We conducted a total of 12 sessions for these treatments leading to a total of 41 subjects from 5 different sessions in the worker role for the Known treatment, 34 from 4 sessions in the Ambiguous treatment and 38 from 3 sessions in the Shift treatment.

\(^1\)Given that the only thing that varies across round colors is the audit probability, one could be worried that this might lead to some experimenter demand effect in which subjects respond to the round colors by changing behavior more than they would if the difference weren’t so salient. While possible, the fact that most of our important results demonstrate a lack of any change in behavior between round colors suggest that there does not appear to have been much of a demand effect.
3 Hypotheses

There are several issues we need to test in the data in order to provide an answer to the underlying questions. In this section, we will provide a theoretical characterization of the choice behavior, as well as explain the hypotheses regarding how that choice behavior might vary according to the treatments.

For this analysis we will take the earnings, $E$, of an individual from the encoding task to be fixed. This should not vary with the round color as subjects do not know the color of a round when generating their income. They only know the round color when choosing how much to report. Our main interest is in that reporting decision and how it varies with round colors and so we model that choice rather than the choice of $E$. We will let $t$ represent the tax rate (40% in the experiments) and $c \in [0, 1]$ be the fraction of total earnings that a worker claims in their income report where $c = 1$ means the worker is reporting truthfully and $c < 1$ implies under reporting income.\(^2\) We will then let $p$ be the audit probability. Given this, the choice problem of a subject when facing their incoming reporting decision is as follows:

$$\max_{c \in [0, 1]} (1 - p)u[(1 - tc)E] + pu[E(1 - t - 1.25(t - tc))]$$

The first term represents the utility for the worker should the audit not occur. In this case, the worker receives their utility from their base earnings, $E$, less what they pay in taxes based on how much of their earnings they report, $ctE$. If the audit occurs, then the worker receives their utility from their base earnings less their full tax bill less the fine owed which is 1.25 times the difference between their owed and claimed taxes. This leads to an optimality condition of

$$\frac{1 - p}{1.25p} = \frac{u'[E(1 - t - 1.25(t - tc))]}{u'[(1 - tc)E]}$$

While there is no general analytical solution for $c$, we can assume a functional form for the utility function and solve for $c$ computationally. We will assume a standard CRRA utility function, $u(x) = x^\alpha$ and provide a characterization of how $c^*$ varies with $\alpha$ and $p$. Under the CRRA specification, $c^*$ does not depend on $E$. The characterization of optimal choices can be found in Figure 1 as we show the optimal $c^*$ for the various audit probabilities used in the experiment and for risk aversion parameters on the range of $[1, 1]$. The figure demonstrates that at very low audit probabilities, i.e. 10%, then no individuals should report any income unless they are exceedingly risk averse and even then they report essentially 0. Once the audit probability gets to 50%, then under reporting is not worthwhile for anyone. For the intermediate probabilities of 30 and 40%, the optimal percentage of income reported should be decreasing in $\alpha$.

We have chosen not to obtain a separate risk measure for our subjects as the Known treatment essentially provides this. We saw little value in obtaining a separate measure to compare with this one. Given that we can expect a range of risk aversion parameters in the population, the analysis above provides us with the basis for our first hypothesis regarding

\(^2\)In the experiment, subjects actually could report more income than they earned and a few subjects did so. We allowed for this simply because we didn’t want to impose any restrictions on reporting. Theoretically this is dominated so we won’t consider it in the analysis.
data from the Known treatment.

**Hypothesis 1** In the Known treatment, we expect the propensity for and level of income under reporting to be highest at the 30% audit probability and decreasing as the audit probability increases.

The Known treatment is designed as a test of whether behavior adjusts in the manner suggested by the standard Becker model when punishment is uncertain and this hypothesis reflects the standard predictions from that model; people will be more honest when the likelihood of punishment is higher. We do note, however, that we specifically chose the parameters of the experiment in a range such that this hypothesis should hold. As the analysis above shows, at low levels of the audit probability, the theory would predict no relation between it and the propensity or severity of under reporting as all will completely under report while at high levels of audit probability no one will under report. We chose probabilities in an intermediate range to allow us to observe a behavioral shift. This fact underscores the importance in designing experiments with parameters that allow for the behavioral shifts one wants to observe.

**Hypothesis 2** In the Ambiguous treatment, workers will under report less in blue and green periods than red periods.

Our second hypothesis refers to our Ambiguity treatment. This treatment is intended to help us understand how individuals will choose in an environment possessing ambiguity.
in the audit probabilities. Our statement of this hypothesis is premised on the notion that subjects will be ambiguity averse. If the subjects are ambiguity averse, when faced with the blue and green periods and knowing that the audit probability could be 30 or 50% subjects should place more weight on the possibility of the 50% audit probability and respond accordingly with less under reporting. If we compare their baseline in those two treatments to the red periods when the audit probability is known to be 40%, the subjects should transgress less often in the periods when the audit probability could be 50%. This hypothesis then is a basic test of whether our subjects are ambiguity averse to a strong enough degree that it will impact their behavior.

**Hypothesis 3** When comparing the Known and Ambiguous treatments by color, there should be equivalent behavior in Red rounds and Green rounds but more transgressions in Blue rounds in Known than Ambiguous.

Our third hypothesis is based on a more detailed test of the ambiguity aversion concept allowing us to determine the strength of any potential ambiguity aversion. When comparing behavior by round color between Known and Ambiguous, the red rounds in both treatments should lead to equivalent behavior as in both the audit probability is known to be 40%. For blue rounds, the audit probability is known to be 30% in Known but could be 50% in the Ambiguous. Therefore, it is expected that this possibility should lead to fewer transgressions in the Ambiguity treatment. For green rounds, the audit probability is known to be 50% in Known but could be 30% in Ambiguous. Again though, if subjects are ambiguity averse their decisions should be based on a pessimistic view of which probability could be in effect and assume it should be the 50% probability, at least in the case of extreme ambiguity aversion. This means that we might see the same behavior in green rounds between both treatments. Of course this is a very strong view of inequality aversion which is unlikely to hold. It will not be surprising if we instead find more transgressions in the Ambiguity treatment but how many fewer will be a useful gauge of the level of ambiguity aversion present.

**Hypothesis 4** In the Shift treatment, we expect that there will be fewer transgressions in red rounds than in blue rounds and there should be equal number between red and green rounds.

The Shift treatment serves several purposes. The first is as a test for how behavior might be affected by ambiguous shifts in enforcement levels in a manner we believe comparable to how shifts in police enforcement levels might change expectations of punishment likelihood in the field. In the field and in this case, our subjects will know that some shift has happened between one regime (round) and the next but they won’t know what the shift actually is or even have a good way of forming expectations over what the new probability is since they have no knowledge of the process which determines the probability. They won’t even know that the probability of enforcement has gone up only that there is a shift in the potential probability of enforcement.\(^3\) This treatment is also designed as a way to try to identify what element of announcements regarding regime changes might be most likely to impact

---

\(^3\)One might object that in the field, one should know that the probability has gone up even if one doesn’t know what the original probability was or the magnitude of the shift. For some policies that may be true
behavior, whether it is the minimum possible level, the maximum possible level or some naive average. Hypothesis 4 is one approach to testing that and it stated assuming that subjects are strongly inequality averse and therefore respond only to the highest possible probability of detection. Since in both red and green rounds, the maximum detection probability is 70%, we hypothesize that behavior should be equivalent. This should hold if subjects are strongly ambiguity averse and this is despite the fact that if one takes a naive expectation of the probability of audit, the probability is higher in the green rounds than the red ones. When comparing red rounds to blue ones, the maximum probability is lower in blue than red despite the fact that the naive average is the same. Again, if subjects respond according to ambiguity averse preferences this means that the red rounds should have a lower likelihood of transgressions than the blue.

While hypothesis 4 is stated under the assumption of ambiguity aversion, there are other possible bases for how subjects will respond to the audit possibility. For example, it is possible that workers take an optimistic view of the probabilities and respond to the minimum possible probability. If so, then we expect red rounds to generate the most under reporting given that the lowest audit probability in those rounds is 10% while the other two have the same minimum of 30% meaning we should see the same behavior in both. If subjects just naively average the probabilities, then in both blue and red rounds, the average audit probability is 40% and we should observe the same behavior. In the green rounds the average is 50% and so we should observe fewer transgressions. The point being that while our hypothesis is stated on the premise that our subjects are ambiguity averse, it is the case that this treatment was constructed not just to verify the effects of ambiguity aversion but more broadly to provide a clear test of whether subjects pay attention to the max, min or average of a set of possible probabilities when they have no information regarding how those probabilities will be selected.

4 Results

We will provide properly constructed regressions to test each of our stated hypotheses but it is useful to first get some understanding of the structure of the data by examining summary statistics regarding the choice behavior and these are shown in Table 2. For all three treatments, we provide: the average pre-tax earnings, the propensity to under report earnings and the percentage by which subjects under report conditional on under reporting. To be clear, this last summary statistic examines only those individuals who have chosen to under report their earnings. We have provided these summary statistics separately for those subjects whose pre-tax earnings are above average and those whose pre-tax earnings are below average. While theoretically the earnings level is unimportant to the choice to under report in the CRRA specification, we still wanted to examine the issue given that CRRA might not be the behaviorally correct specification and under other forms of risk aversion the optimal level to under report would depend on earnings level.

Examining these statistics provides quite suggestive indications for what our formal results will be. To quickly summarize; in the Known treatment, there is some indication that...
subjects do under report their income most often in Blue rounds and least often in Green which is as expected. In the Ambiguous treatment, above average earners under report most often in Red rounds and less so in Blue and Green which is suggestive of ambiguity aversion but it is not clear that the difference will be large enough to be significant. Below average earners demonstrate no such tendency. In the Shift treatment, subjects tend to under report least in the Green rounds with the propensity in the Red and Blue being similar. This is consistent with what might be seen to sluggish response to regime changes perhaps due to subjects paying attention to the naive average probability rather than to the maximum or minimum probabilities in the range. Our regressions will allow us a more detailed and careful look at each of these issues.

Our first set of regressions is not designed to test one of our hypotheses, but rather to ensure that a particular element of the design worked as intended. Since subjects did not know the color of a round when producing, their pre-tax income levels should not depend on round color. Also, in order for us to compare across treatments it is helpful for pre-tax income to also not depend on treatment. The experiment was designed to deliver the property that the only thing that should differ between rounds that would impact decisions is the audit probability. Table 3 provides a set of regressions to test whether pre-tax earnings were effected by treatment or round color. The regressions are all random effects panel regressions with standard errors clustered at the individual level. To allow for ease of interpreting the results, we provide one specification which includes all treatments. This specification allows for determining whether earnings depends on treatments. We also provide separate regressions for each treatment to determine if earnings depend on round color inside of the treatments. We find treatment and color variables all to be insignificant indicating that the experiment design was successful in delivering the property that pre-tax earnings depend on neither treatment nor round color.

Our first formal hypothesis concerned the Known treatment and simply stated that the

### Table 2: Summary Statistics

<table>
<thead>
<tr>
<th></th>
<th>Below Avg Earners</th>
<th>Above Avg Earners</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Blue</td>
<td>Red</td>
</tr>
<tr>
<td>Known</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pre-Tax Earnings</td>
<td>29.000</td>
<td>28.420</td>
</tr>
<tr>
<td>Propensity</td>
<td>0.632</td>
<td>0.589</td>
</tr>
<tr>
<td>% Under Report</td>
<td>0.355</td>
<td>0.312</td>
</tr>
<tr>
<td>Number</td>
<td>19</td>
<td></td>
</tr>
<tr>
<td>Ambiguous</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pre-Tax Earnings</td>
<td>26.458</td>
<td>27.458</td>
</tr>
<tr>
<td>Propensity</td>
<td>0.500</td>
<td>0.528</td>
</tr>
<tr>
<td>% Under Report</td>
<td>0.288</td>
<td>0.423</td>
</tr>
<tr>
<td>Number</td>
<td>12</td>
<td></td>
</tr>
<tr>
<td>Shift</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pre-Tax Earnings</td>
<td>28.063</td>
<td>28.000</td>
</tr>
<tr>
<td>Propensity</td>
<td>0.438</td>
<td>0.438</td>
</tr>
<tr>
<td>% Under Report</td>
<td>0.333</td>
<td>0.298</td>
</tr>
<tr>
<td>Number</td>
<td>16</td>
<td></td>
</tr>
</tbody>
</table>

Note: % Under Report is measured conditional on under reporting.
propensity to under report should depend on the round color with the ranking being Blue ≥ Red ≥ Green. Table 4 provides the regressions to formally test this hypothesis. We provide three different specifications of this regression. The first is a random effects logit regression with standard errors clustered at the individual level and a dependent variable being a binary variable for whether or not the individual reported earnings below their true earnings in a round. The second specification is a linear probability version of the same specification. The third version examines only the set of individuals who chose to under report and uses as the dependent variable the percentage by which they under reported. The first two specifications speak to the question of whether the individuals under report while the third speaks to the issue of the degree to which they under report if they do. We also provide a second set of these regressions that include a dummy variable to separate out those who earned less than the average in pre-tax earnings and then interactions with the round colors. This allows us to determine if low and high income earners behaved differently. This leads us to our first result.

**Result 1** In the Known treatment, the propensity of subjects to under report responds to round color as hypothesis 1 predicted.

Not surprisingly, hypothesis 1 is not rejected as overall subjects under report more often in Blue rounds than Red and then under report less often in Green rounds than Red. When we split out how low versus high earners behave, we see that high and low earners do respond differently. High earners under report more often in Blue rounds compared to Red but they show no change in behavior between Red and Green rounds. For the low earners, they show exactly the opposite in that they evidence no difference in behavior between Red and Blue rounds but do under report significantly less often in Green rounds.
Table 4: Regressions examining Known treatment.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Blue</td>
<td>1.614***</td>
<td>0.149***</td>
<td>0.099***</td>
<td>2.562***</td>
<td>0.214***</td>
<td>0.132***</td>
<td>(0.481)</td>
<td>(0.045)</td>
<td>(0.031)</td>
</tr>
<tr>
<td>Green</td>
<td>-1.136**</td>
<td>-0.099**</td>
<td>0.008</td>
<td>-0.613</td>
<td>-0.046</td>
<td>-0.032</td>
<td>(0.520)</td>
<td>(0.046)</td>
<td>(0.030)</td>
</tr>
<tr>
<td>LowEarn</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3.124**</td>
<td>0.265*</td>
<td>0.047</td>
</tr>
<tr>
<td>LE*Blue</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(1.517)</td>
<td>(0.140)</td>
<td>(0.079)</td>
</tr>
<tr>
<td>LE*Green</td>
<td>-2.037**</td>
<td>-0.162*</td>
<td>-0.066</td>
<td></td>
<td></td>
<td></td>
<td>(0.908)</td>
<td>(0.083)</td>
<td>(0.058)</td>
</tr>
<tr>
<td>Revenue</td>
<td>0.004</td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
<td>0.033</td>
<td>0.002</td>
<td>&lt;0.001</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>-0.896</td>
<td>0.425***</td>
<td>0.245**</td>
<td>-3.238*</td>
<td>0.244</td>
<td>0.222</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Robust clustered standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1

These results indicate that in a world in which punishment probabilities adjust in a known manner, then the Becker model should be expected to do a good job of explaining/predicting behavioral changes.

Our second hypothesis concerned the Ambiguous treatment and involved trying to determine if our subjects evidenced behavior consistent with ambiguity aversion. Table 5 provides the regressions to test this hypothesis and the regressions have the same structure as before but include only data from the Ambiguous treatment. This leads to our second result.

Result 2 In the Ambiguous treatment, we find no change in propensity to under report for low or high earners based on round color. We do find that conditional on underreporting, subjects under report slightly less in the blue/green ambiguous rounds.

Our regression specifications here compare behavior in Red rounds to Blue and Green combined as subjects should have seen little difference between the two. Overall, we find no significant differences in behavior between Red rounds and the other two. We continue to find no difference when we separate out effects for low and high earners.\(^5\) We do find that conditional on underreporting, the subjects are underreporting less in Blue and Green.

\(^4\)These claims can be verified by checking the significance of the combined terms of Blue + (LE*Blue) and Green + (LE*Green) which result in p-values of 0.329 and 0.027 for specification 4 and 0.336 and 0.030 for specification 5.

\(^5\)The test of the signficance for the linear combination of (Blue or Green) +(LE*BorG) results in a p-value of 0.464 for specification 4 and 0.549 for specification 5.
Table 5: Regressions examining Ambiguous treatment.

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blue or Green</td>
<td>-0.301</td>
<td>-0.032</td>
<td>-0.084***</td>
<td>-0.795</td>
<td>-0.066</td>
<td>-0.061*</td>
</tr>
<tr>
<td></td>
<td>(0.568)</td>
<td>(0.046)</td>
<td>(0.030)</td>
<td>(0.769)</td>
<td>(0.059)</td>
<td>(0.031)</td>
</tr>
<tr>
<td>LowEarn</td>
<td>-0.769</td>
<td>0.020</td>
<td>-0.118</td>
<td>1.340</td>
<td>0.107</td>
<td>-0.080</td>
</tr>
<tr>
<td></td>
<td>(2.058)</td>
<td>(0.178)</td>
<td>(0.112)</td>
<td>(1.065)</td>
<td>(0.091)</td>
<td>(0.073)</td>
</tr>
<tr>
<td>LE* BorG</td>
<td>0.193**</td>
<td>0.018**</td>
<td>0.004</td>
<td>0.202*</td>
<td>0.020**</td>
<td>0.002</td>
</tr>
<tr>
<td></td>
<td>(0.097)</td>
<td>(0.009)</td>
<td>(0.005)</td>
<td>(0.104)</td>
<td>(0.009)</td>
<td>(0.005)</td>
</tr>
<tr>
<td></td>
<td>-5.277</td>
<td>0.021</td>
<td>0.412**</td>
<td>-5.265</td>
<td>-0.034</td>
<td>0.511***</td>
</tr>
<tr>
<td></td>
<td>(3.275)</td>
<td>(0.280)</td>
<td>(0.164)</td>
<td>(3.808)</td>
<td>(0.342)</td>
<td>(0.181)</td>
</tr>
<tr>
<td>Obs</td>
<td>333</td>
<td>333</td>
<td>202</td>
<td>333</td>
<td>333</td>
<td>202</td>
</tr>
<tr>
<td>Clusters</td>
<td>37</td>
<td>37</td>
<td>28</td>
<td>37</td>
<td>37</td>
<td>28</td>
</tr>
</tbody>
</table>

Robust clustered standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1

Table 5: Regressions examining Ambiguous treatment.

rounds than in Red rounds which could be seen as very minor evidence in favor of ambiguity aversion but that would be very weak support for ambiguity aversion.

For our third hypothesis we want to test how behavior compares between Known and Ambiguous treatments to provide a more detailed look at the possibility of ambiguity aversion. Table 6 provides the regressions to provide that comparison. Data from both treatments are included and the structure of the regressions are again the same as in the prior two tables though we find it more convenient to present the results in the form of separate regressions for Low and High Earners. These regressions provide the support for our third result.

**Result 3** In comparing behavior in the Known and Ambiguous treatments, low earners treat each colored the round the same in both treatments. The high earners under report much more often in Red rounds in the Ambiguous treatment than the Known treatment and under report much less often in the Blue rounds in the Ambiguous treatment than in the Known treatment.

As explained before, the Red rounds are identical in the Ambiguous and Known treatments, so identical behavior is expected. This holds for Low Earners but not for High Earners, as High Earners under report much more frequently in the Ambiguous treatment than in Known. This is unexpected. In the blue and green rounds, Low Earners still show no response to treatment. The High Earners on the other hand are less likely to under report in Blue rounds in the Ambiguous treatment than in the Known treatment. Again the difference in this case is that in Known, the subjects knew the audit probably was 30% while in the Ambiguous treatment it could have been 30 or 50%. The indication is that the High Earners are responding to either the maximum possible audit probability or just the increase in the naive expectation of an increase in the audit probability. In Green rounds, High Earners behave equivalently between the two treatments. Had they been making their
decisions on the naive average probability, the fact that the audit probability could have been 30% or 50% in the Ambiguous treatment should have led to increased transgressions relative to the Known treatment in which the audit probability was certain to be 50%. This behavior is therefore consistent with the idea that they are responding to the top possible audit probability rather than the minimum or average. This provides another minimal bit of evidence suggesting that ambiguity aversion could be present in a small degree but, again, not to a substantial degree.

Our final hypothesis focuses on the Shift treatment and is based on trying to test more explicitly if behavior is a function of maximum, minimum or the naive average audit probability. The analysis of this treatment also provides more generally an indication for how individuals respond to ambiguous shifts in probabilities. The regressions to test these issues are shown in Table 7 which again have a similar structure to previous regression tables but are based on data only from the Shift treatment. This leads to our fourth and final result.

**Result 4** Hypothesis 4 is rejected. Subjects evidence equivalent behavior in all round colors in the Shift treatment.

While we find a few indications of ambiguity aversion for High Earners in some of the above tests, our Shift treatment shows us that this is not enough to drive how subjects respond to shifts in the ambiguity regarding the audit probability. In the Red and Blue rounds, the naive average audit probability is 40% but the high and low end possibilities are different. We find that behavior in the Blue and Red rounds are equivalent which is
consistent with the notion that the subjects are responding primarily to the naive average rather than the maximum possible probability. The naive average audit probability for the Green rounds is 50% but the top possible probability is 70% for both the Green and Red rounds. As such, if subjects were responding to the maximum possible probability, they would behave in the same manner between Green and Red rounds while if the naive average drives their behavior, they would transgress less often in Green rounds than Red. We find no response in behavior between Red and Green rounds. In the regressions that separate out high and low earners we find no significant response from either group.\footnote{To test for the total effect on Green vs Red rounds for the low earners we have to test the linear combination of Green +LowEarn + LE*Green. The p-value on the test for columns (4) and (5) are 0.356 and 0.313 respectively.}

While one could see the latter result as perhaps in keeping with Ambiguity Aversion that would only hold if there were a difference in behavior when comparing the Blue to the Red and Green rounds which still yields no difference. Taking these results together, the picture from the Shift treatment is that subjects are simply exhibiting no response to any of the differences in the enforcement regimes when they are presented in an ambiguous manner.

The fact that the behavior doesn’t change much across enforcement regimes in the Shift treatment is important. As described in the introduction, field studies on changes in enforcement regimes also generally find either counterintuitive behavioral changes or a lack of a change in behavior due to enforcement regime changes. Alternatively, in the Known treatment we find sharp shifts in behavior due to regimes shifts. The comparison of these two findings suggest that when a policy change leads to an ambiguous shift in an already
ambiguous enforcement regime then the response is not as strong as when the shift is from one clear probability to another.

5 Conclusion

The intent of this study was to determine the effect of ambiguity in punishment regimes on the tendency for individuals to engage in dishonest acts. Our interest in ambiguity was twofold. First, there exists substantial prior literature suggesting that individuals are ambiguity averse and so we wanted to examine the degree to which ambiguity aversion might drive the behavior in a frame regarding rule breaking. Second, we wanted to examine how shifts in ambiguous enforcement regimes might affect behavior.

Our first treatment is a baseline case designed to examine how individuals would react to shifts in enforcement regimes under the standard assumption of uncertainty, or when the probability of detection is known. In this case, our subjects responded as predicted by a standard model in that as the probability of detection was increased; the propensity to engage in dishonest behavior decreased. Our second treatment made the detection probabilities ambiguous in the sense that the subjects knew the detection probabilities were drawn from a specific set but they didn’t know how leaving them not knowing the actual detection probabilities they faced. This was designed to determine if the subjects were ambiguity averse or if they would react based on a pessimistic belief that the highest detection probability in the set was the one most likely to be in effect. We found little evidence consistent with that hypothesis. The individuals in our experiment who earned more than the average might be seen as mildly ambiguity averse but the effect is not statistically significant and there is definitely no effect on those earning less than the average. This is a very important result from the perspective of designing or perhaps in advertising enforcement regimes. Were individuals ambiguity averse, then an enforcement agency could be well served by advertising enforcement regimes in an ambiguous manner in hopes that people take the pessimistic belief regarding their likelihood of apprehension and respond accordingly. Our results suggest that such a strategy is unlikely to elicit the desired effect.

Our final treatment looked at how shifts in an ambiguous enforcement regime would impact behavior. Our treatment was designed to determine if individuals respond more strongly to the max, min or naïve average detection probability in the ambiguous set. Our findings again suggest that ambiguity aversion is not driving the behavior as they were not responding in a manner consistent with pessimistic beliefs regarding the enforcement regime. In fact, we found that our subjects exhibited minimal responses as we shifted between ambiguous enforcement regimes as their behavior remained about the same between each different regime.

This lack of a response to shifts in ambiguous regimes is important and it is informative to compare it carefully to how behavior shifted in the treatment involving only uncertainty. In the treatment where probabilities are known, one of our shifts in detection probability was to move from a 40% detection rate to 50%. We observed a substantial reduction in dishonest behavior. In our treatment involving a shift in ambiguous regimes, we had two cases in which one might see the average detection probability to be 40% and a third where it was 50%. We do not, however, find a reliable response to that shift overall nor if we look at high versus low earners separately. The implication is that individuals respond at
best more sluggishly, if at all, when enforcement regimes are changed from one ambiguous regime to another compared to how they respond to similar shifts in regimes involving only uncertainty. This result may help explain why many shifts in enforcement regimes in the field have been less effective than anticipated and provides important insights for how one might design and frame future regime changes in a way that might make them more effective.

Our results imply that simply announcing changes in an enforcement regime in an ambiguous manner, such as “more police will be on patrol” or “IRS agents will be carefully scrutinizing more tax filings,” should not be expected to be effective. Indeed we cite several field studies in the introduction that demonstrate the principle that simply announcing ambiguous changes in enforcement regimes has unpredictable and uncertain effects. The issue may well be that simply announcing a shift from a regime with an unknown detection probability to a new regime with a detection probability that will also be unknown is difficult for individuals to process and respond to. If those in the target population have no concrete knowledge about the new detection probability, how is it supposed to have an impact on deterrence? We can return to the notion of ambiguity aversion to see that one perhaps perceived benefit of leaving the detection rates ambiguous is the hope that individuals will take a pessimistic prior and over respond to such an announcement. However, we find no such response in our data which is corroborated by the mixed evidence in the field.

If a law enforcement authority wants to announce a regime shift in a way that should be expected to be effective, our results suggest that quantifying the detection rates or at least the change might be more effective. For example, instead of announcing that the IRS would be evaluating an increased number of tax filings, the agency could announce specific targets for how many additional filings would receive deeper scrutiny by an agent. This would communicate to potential tax evaders the actual increase in their chance of detection, or at least a closer approximation, which allows them to respond in the expected way. Police departments could not realistically pre-announce the patrol patterns of new officers to convey increases in crime detection and apprehension rates; however, they could make a point of announcing increases in apprehension rates to try to make the impact of the increased enforcement quantifiable. Other enforcement agencies could likely find similar ways of presenting any alterations to enforcement regimes aimed at increasing deterrence.

While our results are suggestive that providing policies with more specifics would lead to enhanced deterrence, we note that it is important to proceed with caution when interpreting new and unexpected results like ours to the field. It is important to solidify findings by replicating them in other domains and populations prior to implementing such suggestions with great confidence. We see our results as an important step to shifting policy makers’ focus to the ambiguity inherent in enforcement regimes and encourage both policy makers and researchers to consider carefully how this ambiguity might be limiting the effectiveness of deterrence policies. As the issue of how ambiguity in enforcement influences deviant behavior becomes better understood and the recognition that enforcement regimes involve ambiguity rather than uncertainty gains greater recognition, we expect that it will help in refining the design and framing of enforcement policies.
References


