

Is one judging head the same as three: a natural experiment on individuals vs teams

Theodore Alysandratos¹

¹Royal Holloway, University of London, Department of Economics

Abstract

This paper investigates the differences between decisions taken by individuals and teams. It does so analysing a natural experiment. In 2012 a judicial reform in Greece replaced three member judicial panels with single member ones. The assignment of judges to panels is exogenous, via a lottery. The assignment of defendants to panels is based on a date cut-off. Crucially, cases are tried by both types of panels in parallel. Our novel and unique dataset contains 1723 observations, the universe of decisions regarding drug trafficking offences over 18 months from the court of Athens. We find that single member panels are between 14 and 17 percentage points more likely to convict a defendant. There is no difference in the length of sentences conditional on being convicted, the probability of imposing a fine or the height of the fine. The variance of sentences imposed by single member panels is lower than that of three member panels. This finding is more in line with the model of group polarisation due to information aggregation by Roux and Sobel (2015).

1 Introduction

Every decision in life is taken either by individuals or groups. Yet, we have only limited knowledge regarding whether and how the decisions of groups differ from those of individuals. This paper utilises a natural experiment to study how the decisions of individuals differ by those of teams. We do so using a unique field experiment that took place in Greece. In 2012 judicial panels comprised of three judges were replaced by single member ones. The assignment of judges to panels takes place via a lottery and we are able to follow judges trying cases alone or in a group. The assignment of cases to the panels is based on a date cut-off. We estimate that single member panels are between 14 and 17 percentage points more likely to convict a defendant. This result is robust to a number of checks. The way in which the reform was implemented rules out explanations such as the changing of norms among the judges or the political and social instability that the country experienced during that period. We provide suggestive evidence that are in line with the model of Roux and Sobel (2015). This theory suggests that groups aggregate the information of their members, in contrast to models of preference aggregation (Fischman, 2011 and 2015; Anwar, Bayer and Hjalmarsson, 2015).

This discussion is of great social and political significance in the context of judicial decision making. Countries faced with increased demand for judicial services and binding fiscal constraints may be tempted to create single member panels in order to increase the efficiency of their judicial system. It is important to know how the decisions of those panels would differ and what may be the consequences for the defendants. Our findings suggest that single member panels decisions differ systematically from those of three member ones. Policy makers need to be aware of the differences and take them into account when planning reforms. More generally, our results are useful to designers of decision making mechanisms.

More precisely, in 2012 the Greek parliament passed a law that created single member judicial panels. Offenses that are relatively common and with few difficulties with respect to proof and evidence were assigned to those panels. The existing three member panels did not cease operations. They continued to try cases that had been assigned to them before the reform came into effect. The way that the reform was implemented was similar to quasi-random assign-

ment to treatment. Offenses that were indicted under the same statutes were tried during the same period by both types of panels. To make this an even better case to study the differences between single and three member panels, the assignment of judges to each panel takes place through a lottery. Thus, we observe the same judges trying the same offenses on their own and in a group. Our set up is very similar to that of laboratory experiments in which the experimenter only observed the decision of a team.

We collected 1723 cases, the universe of drug trafficking related offenses, from the court of Athens from June 2012 to January 2014. We find that single member panels are between eleven and fourteen percentage points more likely to convict a defendant. In order to interpret the results we assume that in every case judges are trying to avoid two types of errors: Type I and Type II. Sir William Blackstone, a famous English jurist of the 18th century formulated this dilemma as follows: "It is better that ten guilty persons escape than that one innocent suffer". Famously, Chancellor Bismarck, a century later, is rumoured to have expressed the opposite view: "it is better that ten innocent men suffer than one guilty man escape". The aphorisms of the two men illustrate the dilemma the judges face. If they are too lenient, they respect the rights of individuals, but they jeopardise public welfare by increasing the number of guilty persons they acquit. Should they are too strict, on the other hand, they safeguard public welfare but they will inevitably harm more wrongly convicted individuals. The reaction of the judges in our sample shows that they side more with Bismarck than with Sir Blackstone.

We find no difference in the length of the sentences conditional on conviction. The effects are not just statistically significant, but their size shows their economic and social significance as well. Whereas the policy maker in this case may have achieved the goal of expediting the delivery of justice, the cost of worsening the overcrowding in the state prisons needs to be taken into account. In addition, the variance of the sentences, both conditional on being convicted and the unconditional, is larger in three member panels. This finding is in line with theories that suggest groups deliberate and aggregate the information of their members. This is in contrast with most of the theoretical models used in the literature so far to describe judicial decision making. In information aggregation models groups synthesise the information of their members in order to

form a group posterior. In these models groups reach, mechanically, no worse and usually better decisions than individuals. In the context of judicial decision making, this implication has added weight with respect to the validity of the judicial judgements.

Our paper contributes to the current literature in several ways. This is the first study examining the differences between individuals and groups using field data and exogenous allocation to groups. Moreover, our subjects are highly qualified experts taking decisions in high stakes situations. Even though the cost or the reward to the judge is mostly psychological, the cost to the defendant might be more than five years in prison. All of the experimental literature uses small to moderate stakes and students as their subjects.

The rest of the paper is organised as follows: section 2 presents an overview of the relevant literature and discusses differences of this paper to it. Sections 3 and 4 describe the judicial system in Greece and give a detailed description of the reform that created the single member panels. Sections 5 and 6 present the data and the main results. Section 7 contains additional evidence and robustness checks. Section 8 concludes.

2 Literature Review

This paper draws and builds on two strands in the literature. The first is the investigation of differences in the decision making of individuals vs groups. Traditional approaches consider the individual as the decision making unit. However, in practice, decisions are often taken by groups. This discrepancy spurred a research agenda that investigated whether groups arrive at different decisions from individuals. With respect to risk aversion and choices under risk a number of studies describe the groups as more closely resembling the risk neutral, utility maximising benchmark (Charness et al., 2007; Rockenbach et al., 2007). Others show that groups are better at avoiding biases such as loss aversion (Sutter, 2007), they make fewer mistakes in social learning situations (Fahr and Irlenbusch, 2011). Adams and Ferreira (2010) obtained a very large dataset containing the bets of individuals and endogenously formed groups in a large betting market in Alaska. The purpose of the gamblers was to guess when the ice in a river would break. Betting on the ice break up is a

longstanding tradition in Alaska. The actual timing of the event is affected by exogenous, environmental factors. Their findings suggest that groups' bets are more moderate and more accurate than those of individuals. Similarly Bär et al. (2011) report that mutual funds managed by individuals follow more extreme investment styles and are more industry concentrated.

On the other hand, Bone et al. (1999) find that their groups of two violate expected utility maximisation in ways very similar to those of individuals. Prather and Middleton (2006) and Prather et al. (2001) evaluate field data from investment funds managed either by individuals or teams. They find that both perform similarly and there are no differences in favour of groups. In their review of the literature Kugler et al. (2012) conclude that the jury is still out on whether individuals are more or less risky than teams. They also note that groups tend to perform better in tasks where a clear, objectively correct answer exists (Laughlin and Ellis, 1986). Additionally, Leana (1985) show that when a group has a designated leader, they tend to comply with the leader's suggestions, hence groupthink is limited. It is important to note that in our case it is difficult to establish an objectively correct outcome with respect to the verdict of the trial or the length of the incarceration. Moreover, the three member panels have a clear leader, the presiding judge, who is more experienced and more highly ranked. Therefore, it is even less clear whether the three member panels will perform different than the single member ones.

Moving on from decision making tasks, several lab experiments have examined how groups behave relative to individuals in games. In their review paper (Charness and Sutter, 2012) conclude that groups are more rational than individuals. Teams exhibit a higher level of depth of reasoning in beauty contest games with (Kocher et al., 2006) and without selection (Kocher and Sutter, 2005), as well as in normal form games (Sutter et al., 2013). Groups also make fewer mistakes in tasks that involve the calculation of conditional probabilities and result in counterintuitive actions (Charness et al., 2007). Furthermore, Charness et al. (2010) show that groups are less susceptible to the conjunction fallacy. Finally, groups avoid more frequently the winner's curse in takeover games (Casari et al., 2009) and overbidding in contests (Sheremeta and Zhang, 2010). Teams have also be found to conform more closely to the rational benchmark in Ultimatum games (Bornstein and Yaniv, 1998; Robert and Carnevale,

1997). Related to the ultimatum game is the dictator game. It has been used in the economic literature to distinguish between expectations and pro-social motivations in the allocation of the pie. Cason and Mui (1997) present the results of an experiment where teams tended to be more generous. In order to explain this discrepancy with the rest of the literature, Luhan et al. (2009) conducted an experiment with bigger groups, relative to Cason and Mui (1997) and anonymity of the group members. They show that under those conditions, the teams were offering less than individuals. Translating these findings in our context, it is possible that three member panels are better at evaluating the defendants. If groups are less generous than individuals, one would expect single member panels to assign shorter sentences and to convict the accused less frequently. However, the explanation in Luhan et al. (2009) relies partially on the need of individuals to obey social norms. It is possible that the prevailing norms among judges are such that do not lead to this outcome.

The main difference of the aforementioned papers to ours is that, with few exceptions, they involve laboratory studies with students. In contrast, we examine the decisions of very experienced subjects in high stakes situations, at least for one of the parties involved, with significant real life consequences. The senior judges have several years of experience trying similar cases and, if they find the defendants guilty, the minimum sentences are at least five years long. In addition, the field studies involve endogenously formed groups (Adams and Ferreira, 2010) or clients who chose mutual funds accounts managed by individuals or teams (Bär et al., 2011; Prather and Middleton, 2006; Prather et al., 2001). In our case we have exogenous assignment of both judges to cases and cases to judges. Therefore our field data are more similar to a lab experiment and allow to more cleanly identify the effect of deciding in a group or individually, without the influence of self-selection.

The second strand of the literature that is related to this paper is the one discussing the process of judicial decision making. Most of the literature has focused on decisions of multi-member panels and attempts to elicit the mechanism via which the panel reaches a decision. Posner (2010) mentions that judges do not engage in much collective deliberation and argues in favour of dissent aversion. Epstein et al. (2011) provide evidence in favour of that model. They show that federal judges in the United States avoid dissenting when the caseload is

high or there are few judges in the circuit. They argue that dissenting under these conditions would increase the costs on colleagues, and therefore they are avoided. Fischman (2015) compares two models of judicial panel voting, one where the characteristics of the judges matter and one where their votes influence the decision of other judges. He uses several datasets and finds empirical support for the latter model. Anwar et al. (2015) analyse the voting patterns of judicial panels in Sweden. In those panels sit politically appointed lay judges. Anwar et al. (2015) find that the decisions of the panels differ systematically when judges appointed by far-left or far-right parties sit in them. Fischman (2011) develops a model of consensus voting and shows that it fits the data from a United States federal circuit better than a model where judges simply vote based on their ideological preference. Regarding single member panels, Eren and Mocan (2015) recently demonstrated that emotions influence judges' decisions. More precisely, they find that unexpected defeats of the football team of the State University increase the probability of finding a defendant guilty. In a widely cited study Danziger et al. (2011) show that the decisions of judges in single member parole panels in Israel are affected by the time of the day that the case is tried. The authors attribute their finding to mental depletion. However, Glöckner (2016) argues that this effect may be an artifact of rational decision making and further studies are required to verify the existence of mental depletion of judges. Another study (Kip Viscusi, 1999) demonstrates that judges are closer to the rational benchmark and fall prey to fewer biases compared to jurors and the general population. Kirchgassner and Pommerehne (1993) discuss the decision making of judges from a public choice perspective. They caution that judges, especially those with limited career concerns, faced with the possibility of making a mistake, may be tempted to take the decision that results in the least bad mistake for them. They argue this would lead to more convictions.

Finally, Roux and Sobel (2015) propose a formalised psychological theory of deliberation. Their model assumes that groups aggregate the private signals of their members. By construction in their model, groups are better decision makers than individuals. They also exhibit larger variance. Roux and Sobel use their theory to explain experimental results, such as the ones in Schkade, Sunstein and Kahneman (2000). Their model differs fundamentally from the

preference aggregation models that have been used in the literature to describe the decision making of judicial panels (Anwar et al., 2015; Epstein et al., 2011; Fischman, 2015). Our findings are the first field evidence in line with the deliberative model of Roux and Sobel (2015).

This review suggests a lack of field evidence regarding the differences between individual decision makers and exogenously assigned groups. To the best of our knowledge, this is the first study that attempts to do so.

3 The judicial system

In Greece judges are not politically affiliated. Their functional and personal independence is enshrined in the constitution. Since 1995, in order to be appointed they need to pass the examinations for the National School of Judges. The study in the National School of Judges lasts for sixteen months. In order to be eligible to sit the entry exam the candidates need to be at least twenty eight years old, full members of a bar association and to have practiced law for at least two years or have a PhD in law and have practiced law for at least one year. It is also possible to be allowed to take the exam if one has been a Justice of the peace or have worked as legal clerks in the courts for at least five years and possess a Bachelors in law. Justices of the peace hold a legal degree, have legal practice experience and must pass a different examination in order to be appointed.

Once a probationary period of two years is completed, judges cannot be fired unless they are convicted for a criminal offence or commit a serious disciplinary offence or their health status does not allow them to carry out their duties. Judges who are not promoted to the Supreme Court are retired when they have turned sixty five. Supreme Court Justices are retired when they turn sixty seven. The ranks of judges is as follows: First, they are appointed to the rank of Judge of first instance. They are promoted to Presiding Judge of first instance, Judge of appeals, Presiding judge of appeals, Supreme Court Judge and, finally, Vice President and President of the Supreme Court. Promotions, up to the rank of Supreme Court Judge, are decided by Court Councils, consisting of judges. These rights and procedures are constitutionally protected to ensure the independence of the judiciary. It is important to note that "Judge of appeals"

is a rank and does not refer to whether a judge only tries appeals. In fact, as is the case in our dataset, Presiding judges of appeals issue decisions of first instance. In order to avoid confusion, henceforth "Presiding judge" refers to Presiding judge of Appeals and "judge" to Judge of appeals. According to the law, one cannot be promoted to the rank of presiding judge with fewer than ten years of experience in lower ranks.

The members of the judiciary are allocated to panels via a lottery that takes place approximately one month before the day of the trial. Regarding the three member panels, there are two non-overlapping sets from which the members of the panel are drawn. The first consists of all the presiding judges, except for the most experienced ones whose services are reserved for the hardest cases.¹ The presiding judge is legally required to sit between the two other judges and directs the discussion. The second set is comprised of all the judges. A simple majority is required for the panel to take a decision and a decision must be reached. Regarding the single member panels, from the beginning of Autumn 2013 the presiding judges are drawn from the exact same set with the presiding judges of the three member panels. Before Autumn 2013 they were drawn from a list of twenty presiding judges. The Court of Appeals in Athens lists sixty three presiding judges in total in its roster at any given time. The lottery for the single member panels takes place one month before the trial, too. The rules allow for judges and presiding judges to exempt themselves from the lotteries for a limited period of time if their name has been drawn too many times in the past. No other exemptions from the lottery are allowed. Jury panels are not involved in the proceedings of the kind of cases we examine. The burden of the decision lies entirely on the shoulders of the trying judges.

4 The reform

In 2012, the Greek parliament passed a law that reformed its judicial system. The expressed goal of the lawmaker's was to expedite the delivery of justice. Among other measures that reformed civil courts, new judicial panels comprised by a single judge were created for a number of felonies. The list includes, among

¹According to the number of individual presiding judges we observe in our sample and the number of judges who are affiliated with the Court of Appeals in Athens, approximately two thirds of the presiding judges take part in this lottery.

others, burglaries, thefts, drug trafficking, illegal use or selling of weapons, illegal gambling, and violations of intellectual property. The only exception were those offenses that carry a maximum sentence of life imprisonment. The aforementioned offenses are rather common and without major difficulties in terms of evidence and proof.²

A particular feature of this reform was the way that cases were assigned to the new, single judge, panels. The allocation depended not on when the trial was scheduled, but on when the subpoena was served. If a subpoena to a defendant was served on or before 1/4/2012, she would be tried by a three member panel, as was the case before the reform. If it was served on or after 2/4/2012, she would be tried by a single member panel. In other words, we have a discontinuity that creates quasi-random assignment to treatment. Due to scheduling issues, the cases are not tried in strict chronological order according to when the subpoena was served. The defendants are assigned a panel on a first come, first served basis. However, the panels are usually overbooked. The cases that cannot be tried on the scheduled day due to time constraints are re-assigned to a later day. Therefore, cases for which subpoenas were served before the reform came into effect were tried at the same time or even after cases that were referred to the new courts. Thus, any extrajudicial confounding factors, such as the general economic conditions in the country, that may influence the decision of the courts did not differ between the two panels at the time of the trial.

In a sense our setting is very similar to that of Sutter et al. (2013), who only observed the decision of the team and did not know what exactly took place during the group deliberations. Our advantage is that we can observe the same agents, namely the presiding judges, when they are randomly allocated to both single- and three-member panels. Moreover, our subjects are not university students taking decisions for small to moderate stakes. Instead, they are highly qualified and experienced experts taking decisions with important consequences for the defendants.

It is important to note for the validity of our results that it is the state prosecutors who bring charges against the accused. If the court finds that

²Every draft bill that is put to vote is accompanied by a report of the sponsoring ministry regarding the rationale for each article of the bill. The report regarding the creation of the single judge panels explicitly stated as the reason for their creation the simplicity and the commonality of the offenses referred to such panels.

the accused should have been indicted for less serious crimes, they still have to try them for the lesser charges. They are not allowed to refer them to a different panel. This characteristic of the system combined with (partially) non-overlapping mandatory sentences ranges allow us to identify comparable cases in the two panels.

5 The data

We collected our data at the biggest penal court in Greece, the court of Athens. Approximately 20% of the annual load of cases in the country is tried there. About two hundred judges serve the court at any time, of which sixty three hold the appropriate rank to preside over cases in single and three member panels. Our dataset was collected with the help of a research assistant, an experienced and appropriately instructed lawyer. He collected the observations using dedicated terminals accessible from the site of the court.

In total we collected 1723 observations for the period from June 2012 to January 2014. In order to minimise noise and allow for easier identification of any effect, we restricted our sample to drug cases. Our dataset contains the universe of drug related cases that were tried in Athens during that period and for which a decision of guilt or acquittal was made. We did not collect data on cases that were postponed or for which the statute of limitation had expired or the indictment ceased for any other reason (e.g. the death of the defendant).

During the period of our sample there were two laws penalising drug trafficking. From September 2012 to 19th March 2013, simple drug related offenses carried a sentence between three and fifteen years and a fine varying from 29 to 290,000 Euros. Aggravated offenses had a mandatory minimum sentence of fifteen years and a maximum of life imprisonment and a fine of at least 15,000 and up to 588,235 Euros. From 20th March 2013 until January 2014 the law mandated a sentence between five and fifteen years and a fine of up to 500,000 Euros for the simple cases. Aggravated offences carried a minimum sentence of ten years to life imprisonment and a fine between 50,000 and 600,000 Euros. It is important to note that the applicable law depends on the day that the trial took place, regardless of when the defendant was indicted.

We have information on: the nationality of the defendant (Greek or for-

eigner), the type of panel in which they were tried(single or three member panel), the day of the trial and the day the written decision was published³, the verdict of the panel(guilty or not guilty), the length of the sentence, whether a fine was imposed and the height of the fine in thousands of Euros, whether the sentence was suspended, the name of the presiding judge and her sex. Due to reasons related to the law for the protection of private data we did not record the name or the surname of the defendants, even though that was visible on the database. Because of that we only have data on the sex of the Greek defendants. The sex of the defendants and the judges is easily identifiable in Greek from their name and surname, which have unique gender identifiable suffixes. Our variables are described in table 1.

A weakness of our data is that we cannot tell immediately under which article of the law the defendants were indicted. This is potentially a problem, since ideally we would like to compare the decisions on the same offense tried under the two panels. In order to make sure that we are comparing apples to apples, we exploit the mandatory minimum sentences for the more serious offenses. Those offenses have a minimum of fifteen or ten years, depending on the time when the trial took place⁴, and may carry a life sentence. We restrict our sample to offenses with an imposed sentence strictly below fifteen years for cases tried up to and including March 2013 or strictly below ten years for cases tried from April 2013 onwards. This splitting of the sample guarantees that the cases we include were tried for the less serious offenses and are, thus, comparable between panels. It is possible for a defendant to be indicted for an aggravated offense, but to be found guilty of lesser charges only. We consider that instance to be substantially the same as cases in which the accused was indicted for a lesser offense from the start. Another concern is that the composition of the defendants changed over time. In other words, perhaps criminals changed their behaviour over time. Even though we do not have access to more detailed data that would allow us to directly assess that suspicion, we have two ways to address it. First, we use the mandatory minimum sentences as a cutoff for the seriousness of the offense. Second, since we know the time of the trial, we can

³The decision of the panel is announced orally in court on the day of the trial and is effective immediately. However, the written decision is required for a number of measures of judicial recourse cannot be exercised without it. The written decision contains the opinion of the panel and its arguments.

⁴Until March 2013 the minimum sentence for an aggravated offense was fifteen years. From April 2013 the minimum sentence is ten years.

Table 1: Description of variables

Variable Name	Description
Verdict (Dummy)	The decision of the panel regarding the outcome of the trial. It takes the value 1 if the defendant was found guilty and 0 if she was found not guilty.
Punishment	The length of the sentence in months. The maximum sentence possible is 25 and the minimum is 0.
Fine (Dummy)	A dummy variable that is 1 if a fine was imposed and 0 otherwise.
Suspension (Dummy)	A dummy variable that is 1 if the sentence was suspended and 0 otherwise.
Panel (Dummy)	A dummy variable that is 1 if the case was tried by a three member panel and 0 otherwise.
Foreigner (Dummy)	If the name of the defendant was written using latin script, she is classified as foreigner and the value of the variable is 1. If her name was written using greek script, she is considered Greek and the value is 0.
Trend	The number of the week in which the case was tried relative to the first week in which we observe trials in our dataset. The value for the first week is set to 1.
New Law (Dummy)	A dummy that is 1 if the case was tried under the new law that took effect after April 2013 and 0 otherwise.
Presiding Judge's sex (Dummy)	We utilise the suffixes in the names and the surnames of the presiding judges to assign them to a sex. The variable is 0 if the presiding judge is female and 1 if he is male.
Defendant's sex (Dummy)	A dummy variable that is 1 if the defendant is a woman and 0 if he is male. Data on this variable were collected only for Greek defendants.

control for time effects. Other potential confounding factors, such as a change in the norms or the attitude of the judges are addressed by our research design. We observe decisions on similar cases, tried concurrently by the same presiding judges allocated to one of the two types of panels randomly.

Table 2 shows the descriptive statistics of our dataset, divided by panel type

(single and three members). Our full sample contains 102 presiding judges, 38 of which appear in both panels at least once. A Wilcoxon rank sum test using the full sample of observations shows that the difference in the probability of conviction is statistically significant (p-value=0.000). The average punishment in single member panels is 91 months and in three member panels almost 104 months (Wilcoxon rank sum test, p-value = 0.000). Be advised however that more severe cases are included in the full dataset. Those cases are only tried by three member panels. When restricting ourselves to the subset of comparable cases, the average sentence imposed by single member panels is 86 months and that by three member panels is 75 months (Wilcoxon rank sum test, p-value=0.000). The average probability of imposing a fine conditional on convicting for the subset of comparable cases is 73 percent for single member panels and 69 per cent for three member ones. The difference is not statistically significant (Wilcoxon rank sum test, p-value=0.118). The average fine imposed by single member panels is approximately ten thousand euros, whereas three member panels on average impose a fine of eleven thousand euros (Wilcoxon rank sum test, p-value=0.818).

Figure 4 presents the cumulative frequency of the sentences. Panel A plots the cumulative frequency for the full sample and panel B only for the sub-sample of observations for which we are certain the cases are comparable. In panel B about half the cases in the three member panels receive sentences up to five years and seventy five percent of the cases are convicted to up to nine years in prison. In contrast, single member panels convict fifty percent of the cases tried in front of them to approximately nine years in prison and seventy five percent of the defendants in single member panels are sentenced to ten years of imprisonment. A Kolmogorov-Smirnov test shows that the distribution of punishments assigned by the two courts in this sub-sample differ significantly (p-value=0.000). Panel A shows that after the tenth year of imprisonment the picture is reversed. This is to be expected because more serious offenses that are tried in three member panels carry a sentence of at least ten years. Therefore, when considering the full sample the penalties imposed by the three member panels are biased upwards and attenuate the effect we observe in the sample of cases with sentences of up to ten years. Nevertheless, the difference in the two distributions remains statistically significant (Kolmogorov-Smirnov

Table 2: Descriptive statistics

Defendants	Full sample		
	Single member panels	Three member panels	Total
Number of cases	703	1020	1723
Prob. of conviction (%)	92.6	85.3	86
Avg sentence length Convicting (in months)	90.87	103.5	98
Prob. of fine Convicting (%)	74	74	74
Avg. fine height Convicting (in thousands of Euros)	11	18	15
Prob. of suspended sentence Convicting (%)	19	23	21

Defendants	Comparable cases		
	Single member panels	Three member panels	Total
Number of cases	637	831	1468
Prob. of conviction (%)	92	82	86
Avg sentence length Convicting (in months)	86	75	80
Prob. of fine Convicting (%)	73	69	70
Avg. fine height Convicting (in thousands of Euros)	9.8	10.7	10.2
Prob. of suspended sentence Convicting (%)	21	29	25

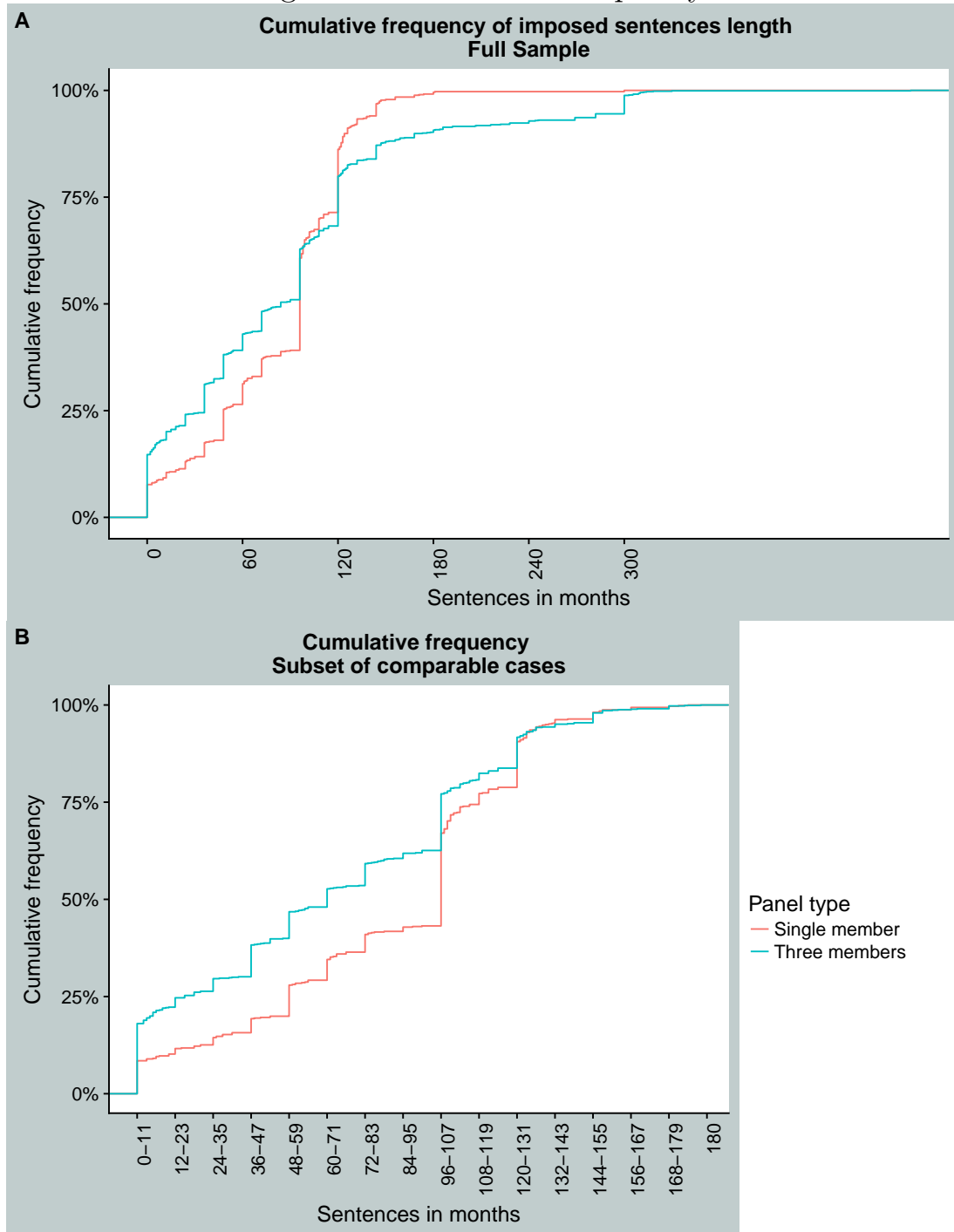
	Presiding Judges		
	Single member panels	Three member panels	Total
Total Number	49	91	102
Proportion Female	49	68	64
Cases per Judge	8.23	5.61	8.54

test, p-value=0.002).

6 Results

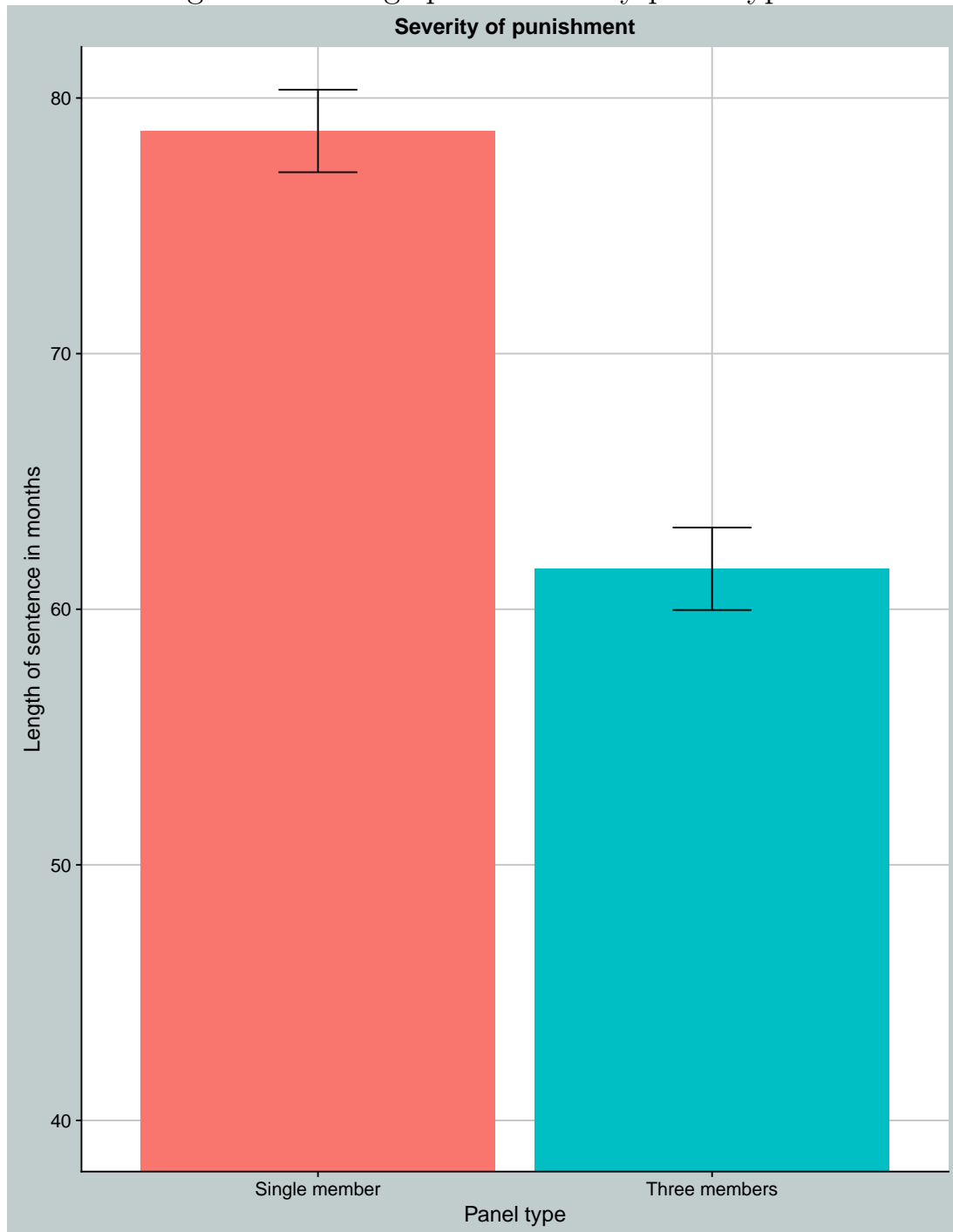
We now turn our attention to examining the differences between single and three member panels more formally. For the purposes of this section we will restrict our sample to cases with imposed sentence strictly below 15 years up to March 2013 and strictly below 10 years from April 2013. These criteria take advantage of minimum sentences for more serious offenses and exclude them. This way we ensure the comparability of our observations.

Figure 1: Cumulative frequency



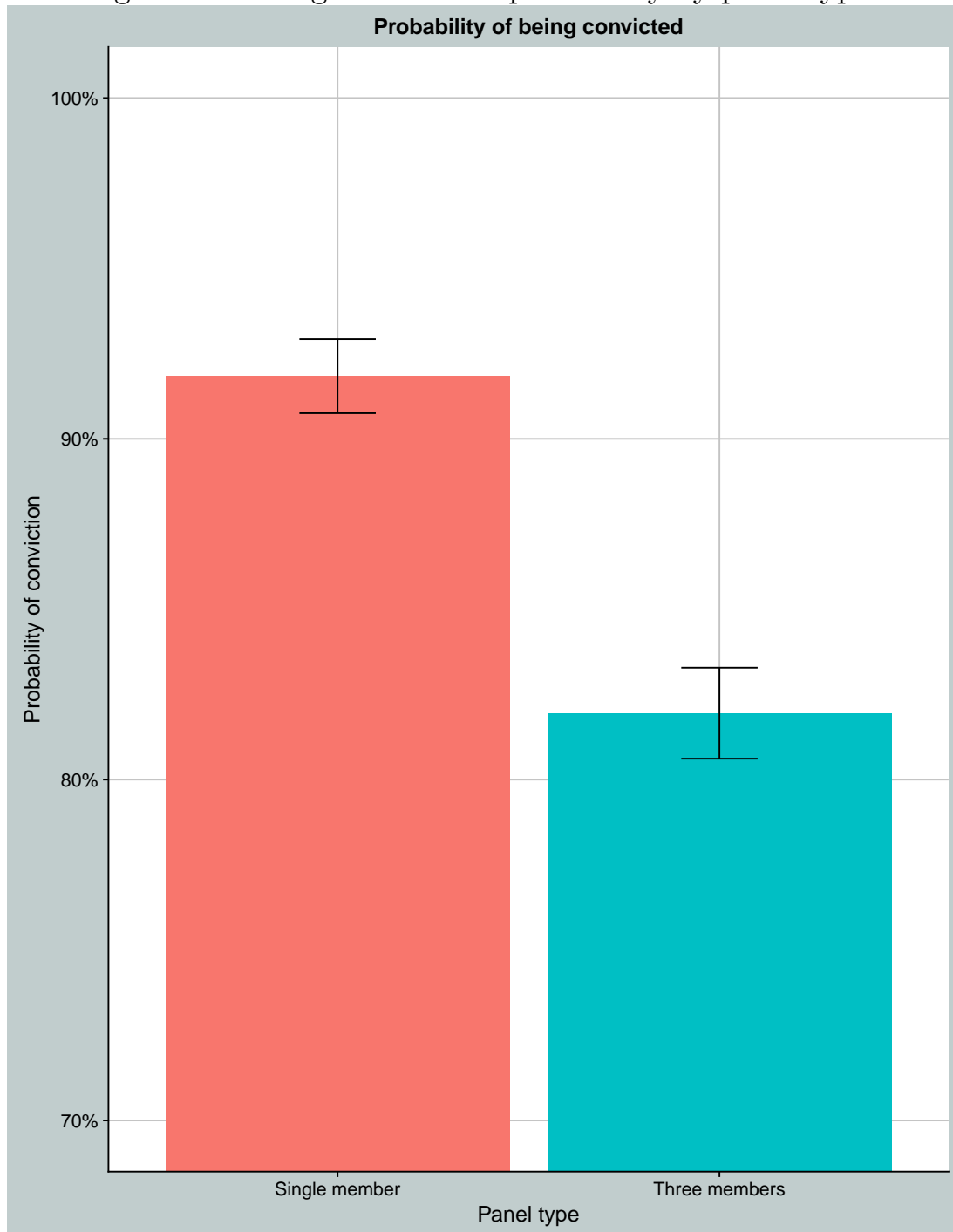
Notes: The x-axis shows the cumulative frequency and the y-axis the imposed sentences in months. Panel A presents the cumulative frequency for the full sample. Panel B presents the cumulative frequency only for the main subset which includes imposed sentence strictly less than 15 years up to March 2013 and strictly less than 10 years from April 2013. These criteria take advantage of minimum sentences for more serious offenses and exclude them. Hence, our observations are comparable.

Figure 2: Average punishment by panel type



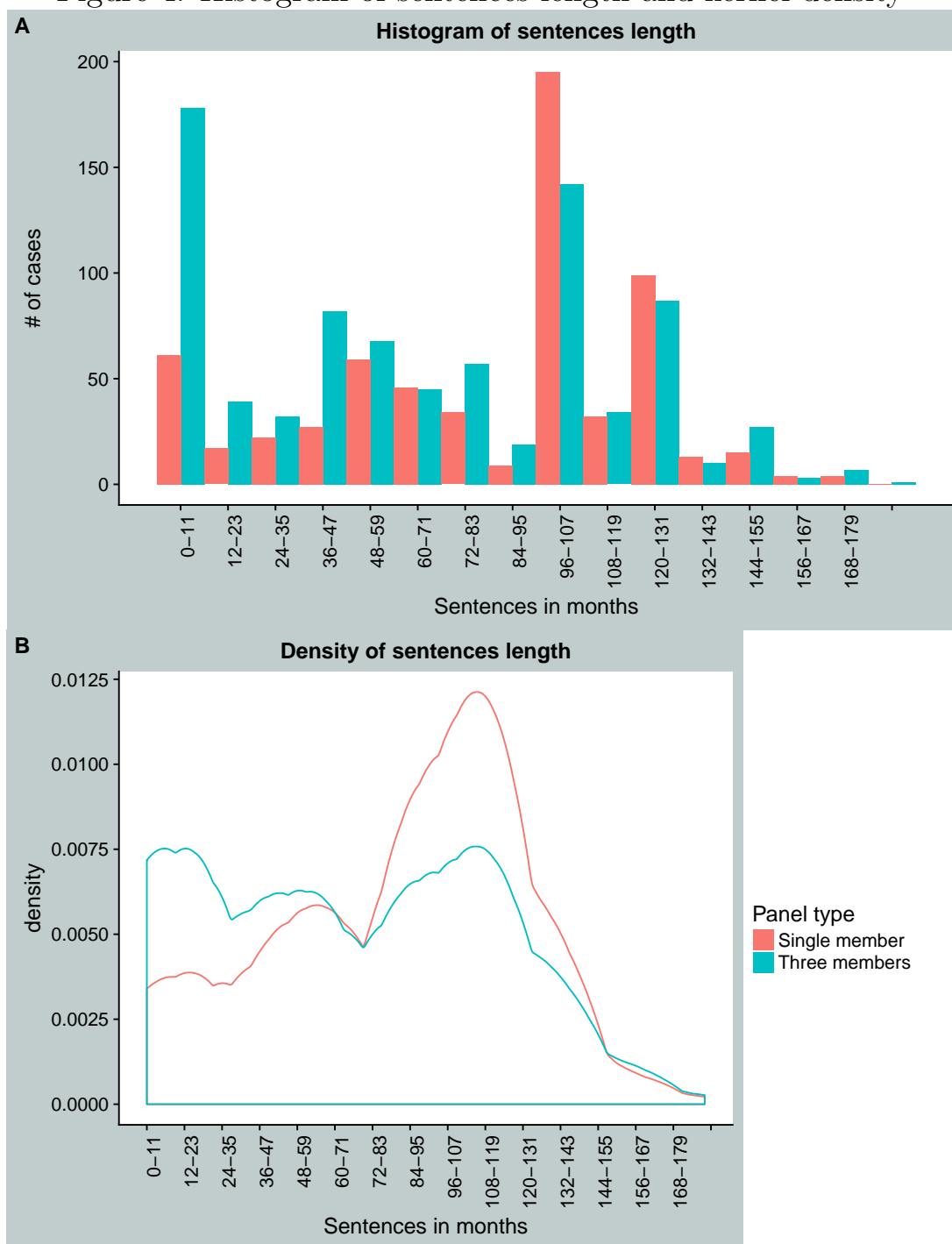
Notes: The y-axis shows the average probability of being convicted and the x-axis the type of panel to which it refers. The figure was produced using the main subset which includes imposed sentence strictly less than 15 years up to March 2013 and strictly less than 10 years from April 2013. These criteria take advantage of minimum sentences for more serious offenses and exclude them. Hence, our observations are comparable.

Figure 3: Average conviction probability by panel type



Notes: The x-axis shows the panel type and the y-axis the average imposed sentence in months. The figure was produced using the main subset which includes imposed sentence strictly less than 15 years up to March 2013 and strictly less than 10 years from April 2013. These criteria take advantage of minimum sentences for more serious offenses and exclude them. Hence, our observations are comparable.

Figure 4: Histogram of sentences length and kernel density



Notes: In panel A the y-axis shows the number of cases that fall into each bin. The width of the bins in the x-axis is 12 months. Panel B presents the estimated density of the imposed sentences. An Epanechnikov kernel was used with binwidth=12. The figures were produced using the main subset which includes imposed sentence strictly less than 15 years up to March 2013 and strictly less than 10 years from April 2013. These criteria take advantage of minimum sentences for more serious offenses and exclude them. Hence, our observations are comparable.

Result 1 *Three member panels are less likely to convict a defendant.*

In tables 3 and 4 we present the results of regressions regarding the conviction rates. Columns 1-2 present the output of OLS regressions. Columns 3-4 present the output of Logit regressions. In column 1 the conviction rate is estimated to be approximately ten percentage points lower in three member panels. The implied average marginal effect from column 3 is that the conviction rate in three member panels is about eight percentage points lower. This result is robust to expanding the sample using the full sample (columns 2 and 4), even though the average marginal effect is diminished in those regressions. These results seem to provide support for the proposition of Kirchgassner and Pommerehne (1993) who hypothesised that under conditions similar to ours the judges would turn to low cost mistakes. A judge has to weigh the cost to society from letting a guilty person walk free and the cost to the individual from convicting an innocent. It is consistent with our findings that the result of this weighing process favours the interest of the public against that of the individual. Another explanation could be that there are judicial peer effects which result in moderation of opinions. Later we present evidence that this is an unconvincing interpretation.

In tables 3 and 4 we also observe that the nationality of the defendant has strong explanatory power. Foreigners are about twelve percentage points more likely to be found guilty and the effect is relatively stable regardless of which partition of the sample or regression technique (OLS or Logit) is used. There are a number of potentially relevant differences between Greek and foreign defendants that may explain this phenomenon, other than judicial discrimination. For example, we do not know if the foreigners are more likely to rely on public defenders, their criminal history or their associates.

Moreover, we find some indications of gender effects in our regressions. Panels in which the presiding judge is female are more likely to convict a defendant. The effect however is only significant at the 10% confidence level when using a logit regression with the comparable sample or an OLS regression with the full sample of observations. Its significance increases in the full sample when using a logit regression. In any case the interaction of the presiding judges sex with the panel type is not significant with either sample and specification. This shows that the driving force of the gender differences is unlikely to be peer effects.

Table 3: Regressions regarding the conviction rate

	<i>Dependent variable:</i>	
	Verdict	
	(1)	(2)
Panel	-0.104 p = 0.0003***	-0.070 p = 0.005***
Foreigner	0.126 p = 0.000***	0.139 p = 0.000***
New Law	-0.006 p = 0.837	0.013 p = 0.637
Trend	-0.0004 p = 0.566	-0.001 p = 0.390
Senior Judge's sex	0.038 p = 0.155	0.040 p = 0.089*
Panel*Senior.Judge's sex	0.046 p = 0.201	0.031 p = 0.320
Constant	0.821 p = 0.000***	0.810 p = 0.000***
Observations	1,468	1,723
R ²	0.060	0.060
Adjusted R ²	0.056	0.056
Residual Std. Error	0.335 (df = 1461)	0.313 (df = 1716)
F Statistic	15.529*** (df = 6; 1461)	18.181*** (df = 6; 1716)

Note:

*p<0.1; **p<0.05; ***p<0.01

Notes: OLS regressions. Regression (1) utilises the main subset of cases. Regression (2) includes the full sample of cases. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Result 2 *The length of the sentence does not differ between the two panels.*

In table 5 we present the output of OLS regressions regarding the length of the imposed sentences. According to column (1) there is no difference between the two types of panels in the extensive margins. The sentences imposed in both panels are not statistically different from zero. When we use the full sample in column (2) we see that three member panels impose longer sentences. This is expected, since the full sample includes more serious offenses that could even carry a life sentence.

Result 3 *Both types of panels are equally likely to impose fines.*

Table 4: Regressions regarding the conviction rate

	<i>Dependent variable:</i>	
	Verdict	
	(1)	(2)
Panel	-0.765 p = 0.002***	-0.586 p = 0.017**
Foreigner	1.004 p = 0.000***	1.226 p = 0.000***
New Law	-0.041 p = 0.887	0.154 p = 0.587
Trend	-0.004 p = 0.497	-0.006 p = 0.340
Senior Judge's sex	0.522 p = 0.081*	0.600 p = 0.045**
Panel*Senior.Judge's sex	0.038 p = 0.914	-0.018 p = 0.958
Constant	1.687 p = 0.0000***	1.570 p = 0.0000***
Observations	1,468	1,723
Log Likelihood	-545.162	-573.731
Akaike Inf. Crit.	1,104.324	1,161.463

Note: *p<0.1; **p<0.05; ***p<0.01

Notes: Logit regressions. Regression (3) utilises the main subset of cases. Regression (2) includes the full sample of cases. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Result 4 *Both types of panels impose fines of similar height.*

In table 6 we show the results of regressions regarding the probability of imposing a fine. The panel type is not statistically significant in any regression. In columns (1) and (3) we report the results of an OLS and a Logit regression respectively, regardless of whether the defendant was found guilty. In columns (2) and (4) we report the results of an OLS and a Logit regression respectively conditional on being convicted. The sign reversal of the Panel variable between columns (1 & 3) and columns (2 & 4) is indicative of the lower probability of conviction in the three member panels. In any case the effect is not different from zero.

Similarly in table 7 we find no statistically significant difference between the

Table 5: Regressions regarding the length of sentences

	<i>Dependent variable:</i>	
	Punishment	
	(1)	(2)
Panel	0.029 p = 0.993	15.604 p = 0.002***
Foreigner	28.188 p = 0.000***	51.453 p = 0.000***
New Law	-0.055 p = 0.486	-0.201 p = 0.094*
Trend	-15.734 p = 0.0000***	-9.040 p = 0.088*
Senior Judge's sex	12.201 p = 0.0001***	18.254 p = 0.0001***
Panel*Senior.Judge's sex	-9.186 p = 0.022**	-6.190 p = 0.312
Constant	67.273 p = 0.000***	47.910 p = 0.000***
Observations	1,266	1,723
R ²	0.205	0.159
Adjusted R ²	0.201	0.156
Residual Std. Error	34.102 (df = 1259)	60.790 (df = 1716)
F Statistic	54.196*** (df = 6; 1259)	54.177*** (df = 6; 1716)

Note:

*p<0.1; **p<0.05; ***p<0.01

Notes: OLS regressions. Regression (1) utilises the main subset. Regression (2) includes the full sample. Note that the full sample includes more serious cases where the imposed sentences are expected to be higher, therefore it is not surprising that the effect disappears. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

amounts of fines imposed by the two types of panels. Column (1) presents the results of an OLS regression of the comparable subset on the height of the fine conditional on being convicted.⁵ Column (2) reports results unconditional on the verdict of the panel. In either case there is no statistically significant difference between single and three member panels. These findings reinforce the view that the presiding judges try to avoid the cost to the public from releasing a guilty person. If the difference in the length of the sentences and the probability of conviction were due to judicial peer effects, one would expect judicial moder-

⁵We had to drop six observations because the fine on record seems to be out of the legally allowed range. We attribute that to a clerical error. The result is not sensitive to the inclusion of those observations.

ation to affect not only the probability and the length of incarceration, but also collateral penalties, such as the probability of imposing a fine and the amount of the fine. The reason for that is that the presence of peers is hypothesised to moderate each individual’s biases and preferences. Even though it is not conclusive evidence and it remains a possibility, it is certainly not very convincing that biases manifest only with respect to the probability of the sentence. On the other hand, it seems more plausible to assume that while the outcome of weighing the individual versus the social cost results in favour of the latter with respect to the decision to convict or not, it favours the individual when they concern the the decision to impose a fine and its height.

Result 5 *The variance of the sentences in single member panels is smaller.*

In a preferences aggregation model group decisions are considered to be less likely to yield extreme outcomes than those taken by individuals. With respect to judicial decisions, peer effects are thought to moderate the sentences imposed. However, in our sample we observe the opposite: single member panels have lower variance. The variance of three member panels (variance=2163.54) which is larger than that of the single member panels (variance=1660.96). Both an F-test (p-value=0.000) and a classical Levene’s test based on the absolute deviations from the mean (p-value=0.000) find that the variances of the two panels are significantly different.⁶ The model of information aggregation in groups by Roux and Sobel (2015) predicts that under certain assumptions regarding the distribution of the signals and the individual priors, groups will have higher variance than individuals. Hence, we interpret this finding as a suggestive evidence that judicial panels aggregate information and not preferences. This explanation, as discussed in Roux and Sobel (2015), would also account for a number of irregularities observed in experiments.

7 Additional checks

In this section we discuss alternative explanations and conduct additional tests to tackle those concerns. The first concern is that by using only the cases in our

⁶A modified robust Levene-type test based on the absolute deviations from the trimmed mean (p-value=0.000) and a modified robust Brown-Forsythe Levene-type test based on the absolute deviations from the median (p-value=0.000) yield the same results.

Table 6: Regressions regarding the probability of imposing a fine

	<i>Dependent variable:</i>			
	Fine			
	<i>OLS</i>		<i>logistic</i>	
	(1)	(2)	(3)	(4)
Panel	-0.044 p = 0.259	0.032 p = 0.423	-0.197 p = 0.282	0.176 p = 0.413
Foreigner	0.289 p = 0.000***	0.247 p = 0.000***	1.264 p = 0.000***	1.197 p = 0.000***
New Law	-0.002 p = 0.042**	-0.002 p = 0.043**	-0.009 p = 0.049**	-0.010 p = 0.053*
Trend	-0.082 p = 0.057*	-0.088 p = 0.044**	-0.399 p = 0.053*	-0.496 p = 0.036**
Senior Judge's sex	0.075 p = 0.041**	0.052 p = 0.148	0.374 p = 0.038**	0.287 p = 0.148
Panel*Senior.Judge's sex	-0.012 p = 0.803	-0.046 p = 0.369	-0.094 p = 0.694	-0.264 p = 0.334
Constant	0.544 p = 0.000***	0.646 p = 0.000***	0.247 p = 0.259	0.733 p = 0.004***
Observations	1,468	1,266	1,468	1,266
R ²	0.125	0.102		
Adjusted R ²	0.121	0.098		
Log Likelihood			-885.843	-696.806
Akaike Inf. Crit.			1,785.686	1,407.612
Residual Std. Error	0.457 (df = 1461)	0.431 (df = 1259)		
F Statistic	34.814*** (df = 6; 1461)	23.959*** (df = 6; 1259)		

Note:

*p<0.1; **p<0.05; ***p<0.01

Notes: OLS and Logit regressions. The main subset is used. Columns (1) and (3) report results unconditional on being convicted. Columns (2) and (4) report results conditional on the defendant being convicted. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

subset, we inflate the number of acquittals artificially. Indeed if the acquittal rate for more serious crimes is higher than that of the less serious ones and we have misidentified some of our cases, we would be detecting an effect where there is none. Even though it is an unusual assumption, it is potentially valid. For that reason we performed additional regressions including the full sample. The results of these regressions have been presented in tables 3 and 4. In columns 2 and 4 the panel variable remains statistically significant. This underlines the robustness of our result.

The specific way in which this reform was implemented allows us to address several concerns that are common in cases such as this one. Since the trials in both panels take place during the same period, factors such as changing norms in the court or changes in the behaviour of criminals and any economic or political instability that the country might have experienced are constant between the two panels. Therefore, any differences cannot be attributed to

Table 7: Regressions regarding the height of the imposed fine

	<i>Dependent variable:</i>	
	Amount	
	(1)	(2)
Panel	2,148.427 p = 0.144	1,088.817 p = 0.389
Foreigner	5,854.555 p = 0.000***	5,913.335 p = 0.000***
New Law	12.702 p = 0.725	5.290 p = 0.867
Trend	-2,133.057 p = 0.180	-1,816.613 p = 0.195
Senior Judge's sex	1,012.526 p = 0.441	1,200.273 p = 0.313
Panel*Senior.Judge's sex	2,330.097 p = 0.205	2,256.891 p = 0.160
Constant	2,974.035 p = 0.086*	2,547.133 p = 0.090*
Observations	1,260	1,462
R ²	0.043	0.046
Adjusted R ²	0.038	0.042
Residual Std. Error	15,600.710 (df = 1253)	14,775.100 (df = 1455)
F Statistic	9.324*** (df = 6; 1253)	11.784*** (df = 6; 1455)

Note:

*p<0.1; **p<0.05; ***p<0.01

Notes: OLS regressions. The main subset is used. Column (1) reports results conditional on the defendant being convicted. Column (2) reports results unconditional on being convicted. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

such factors. Even if one assumes that different judges may react differently, we have a panel of judges who try cases in both panels.

In table 8 we present the results of panel regressions regarding the probability of conviction. In columns (1), (2) and (3) pooled, random and fixed effects models respectively confirm the results of the results of column (1) in 3. If anything the point estimates of the random and fixed effects regressions are higher than that of the OLS one. We should note that the Breusch-Pagan test (p-value=0.000) and the Hausman test (p-value=0.228) indicate that the fixed effects estimator is not efficient and the best model is random effects. In these panel regressions single member panels are about 10 to 17 percentage points

more likely to convict defendants.

In table 9 we present the results of panel regressions regarding the length of the imposed sentences. All regressions utilise the sample of comparable cases conditional on being convicted. The difference, regardless of the model used, is statistically not different from zero. Again, the Breusch-Pagan test (p-value=0.000) and the Hausman test (p-value=0.543) indicate that the best model is the random effects one.

Table 8: Panel regressions regarding the conviction rate

	Verdict		
	(1)	(2)	(3)
Panel	-0.104*** (0.028)	-0.142*** (0.035)	-0.173*** (0.043)
Foreigner	0.126*** (0.019)	0.120*** (0.019)	0.110*** (0.020)
New Law	-0.0004 (0.001)	-0.001 (0.001)	-0.001 (0.001)
Trend	-0.006 (0.032)	0.001 (0.033)	-0.0004 (0.036)
Senior Judge's sex	0.038 (0.027)	0.021 (0.040)	
Panel*Senior.Judge's sex	0.046 (0.036)	0.081* (0.046)	0.078 (0.061)
Constant	0.821*** (0.034)	0.866*** (0.043)	
Observations	1,468	1,468	1,468
R ²	0.060	0.114	0.044
Adjusted R ²	0.056	0.110	-0.028
F Statistic	15.529*** (df = 6; 1461)	31.318*** (df = 6; 1461)	12.481*** (df = 5; 1365)

Notes:

***Significant at the 1 percent level.

**Significant at the 5 percent level.

*Significant at the 10 percent level.

Notes: Panel regressions. Column (1) reports the results of a pooled panel. Column (2) reports the results of random effects. Column (3) reports the results of fixed effects. All the columns use the main subset of observations. Note that both the Breusch-Pagan and the Hausman test favour the random effects model. The outcomes of these tests indicate that the fixed effects estimator is not efficient. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

The presiding judges are experienced and have served in the courts for several years. However, it is possible that some of them are more familiar with drug trafficking cases than others and perhaps that difference in familiarity is

Table 9: Panel regressions regarding the length of sentences

	<i>Dependent variable:</i>		
	Punishment		
	(1)	(2)	(3)
Panel	0.029 (3.196)	-0.133 (3.819)	-0.682 (4.769)
Foreigner	28.188*** (2.165)	27.864*** (2.155)	27.727*** (2.234)
New Law	-0.055 (0.079)	-0.062 (0.087)	-0.099 (0.107)
Trend	-15.734*** (3.456)	-16.706*** (3.597)	-18.955*** (3.947)
Senior Judge's sex	12.201*** (2.854)	9.609** (4.199)	
Panel*Senior.Judge's sex	-9.186** (4.003)	-8.018 (4.978)	-3.264 (6.632)
Constant	67.273*** (3.758)	69.119*** (4.713)	
Observations	1,266	1,266	1,266
R ²	0.205	0.200	0.170
Adjusted R ²	0.201	0.196	0.097
F Statistic	54.196*** (df = 6; 1259)	52.279*** (df = 6; 1259)	47.564*** (df = 5; 1163)

Note:

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Notes: Panel regressions. Column (1) reports the results of a pooled panel. Column (2) reports the results of random effects. Column (3) reports the results of fixed effects. All the columns use the main subset of observations. Note that both the Breusch-Pagan and the Hausman test favour the random effects model. The outcomes of these tests indicate that the fixed effects estimator is not efficient. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

driving our results. In order to check whether our result is driven by the familiarity of the presiding judges with these types of offenses we ran additional regressions using only the judges who had tried more than the median number of cases per judge in our sample. The results are presented in 10 and are very similar to those presented in 3 and 5. It follows that it is not the less familiar judges who are afraid of making a mistake. On the contrary, it is those who are more familiar with the cases who drive the result.

Finally, we need to address one last point. So far our regressions indicate that there is a difference in the conviction rate between single and three member panels and that difference is not due to the presiding judges. It could be argued

however that the result is due to the two junior judges. More precisely, it is possible that the two junior judges are systematically more lenient and thus their preference is implemented in the three member panels. If the presiding judges are stricter, this would manifest in the outcomes of single member panels, where they cannot be outvoted by their more lenient colleagues. Even though this theory is plausible, there is a number of pieces of evidence presented already that do not fit in. First, if the presiding judges are stricter, one would expect that characteristic to manifest not only in higher conviction rates but also the other three outcomes examined here. As we have seen, the only difference is that with respect with the probability of conviction. Second, the judges pass through several selection processes and they are indoctrinated in the same lines of thinking, from the age of 18, when they enter law school, until at least the age of 29, when they finish their mandatory training at the National School of Judges. They all have similar backgrounds, they have similar experiences and interact with each other frequently, therefore one would expect their preferences to be relatively similar as well. Third, junior judges may want to impress the presiding judges in order to gain promotion faster or to build their reputation. Given that promotions are decided by the Presiding Judges, career concerns would indicate that their incentive is to pander to the Presiding Judges, not overrule them. Finally, if the two junior judges are systematically more lenient and overrule the presiding judges, one would expect that to result in higher variance for the single member panels. Yet, we observe the opposite. Finally, as previously mentioned, Leana (1985) show that in the presence of a clear leader, such as the presiding judge in our case, groups tend to agree with the leader's opinion. All these pieces of evidence taken together render the explanation that the junior judges are more lenient implausible.

8 Conclusions

In this paper we have utilised a reform of the judicial system in Greece to examine the differences in decision making between individuals and groups. In 2012 Greece established single member panels in order to replace three member ones, in an attempt to expedite the delivery of justice. The new panels were assigned common offenses, such as drugs, burglaries and thefts, which had no

Table 10: Results with subset of judges with high familiarity

	<i>Dependent variable:</i>	
	Verdict	Punishment
	(1)	(2)
Panel	-0.082 p = 0.006***	-0.741 p = 0.823
Foreigner	0.120 p = 0.000***	30.347 p = 0.000***
New Law	-0.015 p = 0.649	-15.777 p = 0.0001***
Trend	0.0001 p = 0.873	-0.057 p = 0.504
Senior Judge's sex	0.047 p = 0.095*	11.756 p = 0.0001***
Panel*Senior.Judge's sex	0.002 p = 0.963	-6.937 p = 0.102
Constant	0.805 p = 0.000***	66.261 p = 0.000***
Observations	1,258	1,088
R ²	0.055	0.222
Adjusted R ²	0.051	0.218
Residual Std. Error	0.333 (df = 1251)	34.117 (df = 1081)
F Statistic	12.246*** (df = 6; 1251)	51.416*** (df = 6; 1081)

Note:

*p<0.1; **p<0.05; ***p<0.01

Notes: OLS regressions. The subset of comparable cases is used. Only the judges who have tried more than the median number of cases in our sample are included. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

particular difficulties with respect to evidence and proof. The three member panels did not cease to exist, but they continued to try cases that had been assigned to them before the implementation of the reform. We utilise the quasi-random assignment of the defendants to panels and the lotteries through which judges are assigned to panels to study the effects of the single member panels on the decisions. In order to do that we collected a novel and unique data set of 1723 drug offenses from the court of Athens.

We find that single member panels are between 14 and 17 percentage points more likely to convict a defendant. However, they do not impose longer sentences, higher fines nor they are more likely to impose a fine. We provide

suggestive evidence that this behavioural response is not due to peer effects but due to the judges attempting to avoid a socially costly mistake. Moreover, we provide the first suggestive field evidence of information aggregation in groups.

We add to the literature by providing the first field evidence with respect to the differences of individuals and teams in high stakes situations. We complement previous studies by providing the first evidence of differences between exogenously formed groups and individuals.

References

- Adams, R. and D. Ferreira (2010). Moderation in groups: Evidence from betting on ice break-ups in alaska. *The Review of Economic Studies* 77(3), 882–913.
- Anwar, S., P. Bayer, and R. Hjalmarsson (2015, May). Politics in the courtroom: Political ideology and jury decision making. (21145).
- Bone, J., J. Hey, and J. Suckling (1999, Apr). Are groups more (or less) consistent than individuals? *Journal of Risk and Uncertainty* 18(1), 63–81.
- Bornstein, G. and I. Yaniv (1998, Jun). Individual and group behavior in the ultimatum game: Are groups more “rational” players? *Experimental Economics* 1(1), 101–108.
- Bär, M., A. Kempf, and S. Ruenzi (2011). Is a team different from the sum of its parts? evidence from mutual fund managers. *Review of Finance* 15(2), 359–396.
- Casari, M., C. Jackson, and J. Zhang (2009, December). Do Groups Fall Prey to the Winner’s Curse? (2009-18).
- Cason, T. N. and V.-L. Mui (1997). A laboratory study of group polarisation in the team dictator game. *The Economic Journal* 107(444), 1465–1483.
- Charness, G., E. Karni, and D. Levin (2007). Individual and group decision making under risk: An experimental study of bayesian updating and violations of first-order stochastic dominance. *Journal of Risk and Uncertainty* 35(2), 129–148.
- Charness, G., E. Karni, and D. Levin (2010). On the conjunction fallacy in probability judgment: New experimental evidence regarding linda. *Games and Economic Behavior* 68(2), 551 – 556.
- Charness, G. and M. Sutter (2012, September). Groups make better self-interested decisions. *Journal of Economic Perspectives* 26(3), 157–76.
- Danziger, S., J. Levav, and L. Avnaim-Pesso (2011). Extraneous factors in judicial decisions. *Proceedings of the National Academy of Sciences* 108(17), 6889–6892.

- Epstein, L., W. M. Landes, and R. A. Posner (2011). Why (and when) judges dissent: A theoretical and empirical analysis. *Journal of Legal Analysis* 3(1), 101–137.
- Eren, O. and N. Mocan (Forthcoming). Emotional judges and unlucky juveniles. *AEJ:Policy*.
- Fahr, R. and B. Irlenbusch (2011). Who follows the crowd—groups or individuals? *Journal of Economic Behavior & Organization* 80(1), 200 – 209. Special Section: Fish Markets.
- Fischman, J. B. (2011). Estimating preferences of circuit judges: A model of consensus voting. *The Journal of Law and Economics* 54(4), 781–809.
- Fischman, J. B. (2015). Interpreting circuit court voting patterns: A social interactions framework. *The Journal of Law, Economics, and Organization* 31(4), 808–842.
- Glöckner, A. (2016, November). The irrational hungry judge effect revisited: Simulations reveal that the magnitude of the effect is overestimated. *Judgment and Decision Making* 11(6), 601–610.
- Kip Viscusi, W. (1999). How do judges think about risk? *American Law and Economics Review* 1(1), 26–62.
- Kirchgassner, G. and W. W. Pommerehne (1993, September). Low-cost decisions as a challenge to public choice. *Public Choice* 77(1), 107–115.
- Kocher, M., S. Strauß, and M. Sutter (2006). Individual or team decision-making—causes and consequences of self-selection. *Games and Economic Behavior* 56(2), 259 – 270.
- Kocher, M. G. and M. Sutter (2005). The decision maker matters: Individual versus group behaviour in experimental beauty-contest games*. *The Economic Journal* 115(500), 200–223.
- Kugler, T., E. E. Kausel, and M. G. Kocher (2012). Are groups more rational than individuals? a review of interactive decision making in groups. *Wiley Interdisciplinary Reviews: Cognitive Science* 3(4), 471–482.

- Laughlin, P. R. and A. L. Ellis (1986). Demonstrability and social combination processes on mathematical intellectual tasks. *Journal of Experimental Social Psychology* 22(3), 177 – 189.
- Leana, C. R. (1985). A partial test of Janis' groupthink model: Effects of group cohesiveness and leader behavior on defective decision making. *Journal of Management* 11(1), 5–18.
- Luhan, W. J., M. G. Kocher, and M. Sutter (2009, Mar). Group polarization in the team dictator game reconsidered. *Experimental Economics* 12(1), 26–41.
- Posner, R. A. (2010). *How judges think*. Harvard University Press.
- Prather, L. J. and K. L. Middleton (2006). Timing and selectivity of mutual fund managers: An empirical test of the behavioral decision-making theory. *Journal of Empirical Finance* 13(3), 249 – 273.
- Prather, L. J., K. L. Middleton, and A. J. Cusack (2001). Are n+1 heads better than one? the timing and selectivity of Australian-managed investment funds. *Pacific-Basin Finance Journal* 9(4), 379 – 400.
- Robert, C. and P. J. Carnevale (1997). Group choice in ultimatum bargaining. *Organizational Behavior and Human Decision Processes* 72(2), 256 – 279.
- Rockenbach, B., A. Sadrieh, and B. Mathauschek (2007). Teams take the better risks. *Journal of Economic Behavior & Organization* 63(3), 412–422.
- Sheremeta, R. M. and J. Zhang (2010, Jul). Can groups solve the problem of over-bidding in contests? *Social Choice and Welfare* 35(2), 175–197.
- Sutter, M. (2007). Are teams prone to myopic loss aversion? an experimental study on individual versus team investment behavior. *Economics Letters* 97(2), 128 – 132.
- Sutter, M., S. Czermak, and F. Feri (2013, November). Strategic sophistication of individuals and teams. experimental evidence. *European Economic Review* 64, 395–410.