Cognition and incentives in plea decisions: Categorical differences in outcomes as the tipping point for innocent defendants

Rebecca K. Helm

University of Exeter

This research was funded by a UK Research and Innovation grant 'Putting Cognitive Science to Work in the Legal System' to Rebecca K. Helm. The author has no real or potentially perceived conflicts of interest. Researchers should contact the author to obtain materials and data for purposes of reproducing the results or replicating the procedure. Correspondence concerning this article should be addressed to Rebecca K. Helm, University of Exeter, Amory Building, Rennes Drive, Exeter, EX4 4RJ. Email: r.k.helm@exeter.ac.uk.

Abstract

Existing research suggests that incentives to plead guilty may influence guilty and innocent defendants differently. This study examines that possibility through testing theoretically-informed predictions relating to the interaction between different types of plea discount (sentence length and sentence type), guilt, and probability of conviction in predicting plea decisions, with a focus on the discounts in England and Wales. Participants (N=3,375) made plea decisions in vignettes that varied discount type offered, probability of conviction, and guilt between-subjects. Participants also answered questions about considerations that were important to them when making plea decisions. Results provide support for predictions, specifically at higher levels of probability of conviction, by showing that a discount resulting in a categorically different sentence type (probation rather than custody) encouraged both 'guilty' and 'innocent' participants to plead guilty, but that a discount resulting only in a shorter sentence of the same type (a 1/3 reduction in sentence-length) only encouraged 'guilty' participants to plead guilty. Participant reports of the considerations important to them when pleading suggest that the categorical discount reduced the importance of factual guilt or innocence in the decision-making of innocent defendants. Findings suggest that utilising plea discounts that vary sentences quantitatively but not categorically is important in maximising the extent to which plea discounts appeal to guilty but not innocent defendants.

Keywords: Guilty plea decision-making; legal decision-making; fuzzy-trace theory; sentencing; criminal procedure.

Cognition and incentives in plea decisions: Categorical differences in outcomes as the tipping point for innocent defendants

Accumulating psychological research has provided important insight into guilty plea decision-making. This research highlights the complexity of guilty plea decisions, combining assessments of risks and benefits that permeate risky decisions more generally with guilt and fairness related considerations (Helm, 2018; Redlich et al., 2017). This complexity arises in part because the sentence a defendant receives when pleading guilty is typically more lenient than the sentence they would receive if convicted at trial. The level of penalty discount that should be awarded to a defendant on the basis of a guilty plea is an important policy question. On the one hand, discounts must be sufficient to incentivise guilty defendants to plead. When a defendant pleads guilty it saves public time and money and saves victims and witnesses from having to testify (e.g. Sentencing Council, 2017; Missouri v Frye, 2012). In fact, many modern criminal justice systems would not be sustainable if many defendants did not plead guilty (e.g. R v Caley and others, 2012; Santobello v New York, 1971). On the other hand, discounts should not be so generous that they pressure innocent defendants to plead (Blume & Helm, 2014) or result in unduly lenient sentences for offenders (Sentencing Council, 2011). Research can assist in this policy debate through providing insight into guilty plea decision-making, and specifically how incentives to plead guilty might be tailored to appeal to guilty but not innocent defendants.

Existing research into guilty plea decision-making has confirmed that many of the factors that influence risky decision-making in problems such as the dread disease problem (Tversky & Kahneman, 1986) also influence plea decisions. Potential sentences at plea and trial, the probability of conviction at trial, and risk preferences have all been highlighted as important influences on plea decisions. Generally, research suggests that defendants will plead guilty more when the sentence at trial is more severe or the sentence when pleading is

less severe (Bordens & Basset, 1985), when the probability of conviction is higher (Tor, Gazal-Aval & Garcia, 2010; Wilford et al., 2020), or where they have a lower preference for risk (Bjerk, 2007). Each of these parameters will be influenced by other key variables, for example the strength of the evidence against the defendant will inform estimates of the probability of conviction.

The "shadow of trial" model of plea decision-making formalises these factors and the relationship between them to explain guilty plea decision-making. According to the model a defendant will compare the certain outcome from pleading guilty with the expected value of trial (meaning the probability of conviction at trial multiplied by the outcome if convicted at trial) and make a decision based on a comparison between these outcomes (Mnookin & Kornhauser, 1979). The model suggests that a risk-neutral defendant will plead guilty where their expected value of trial is worse (i.e. more severe) than the plea outcome (Bibas, 2004; Easterbrook, 2013; Mnookin & Kornhauser, 1979). Thus, according to the model, plea decisions are responsive to changes in the probability of conviction and the specific outcomes at plea and trial. Research has provided some support for this model in predicting long-run average plea decisions but has also shown that it is unlikely to be a reliable predictor of decisions in individual cases (Bushway & Redlich, 2012; Bushway et al., 2014).

Empirical research has also demonstrated systematic biases that are not accounted for by the shadow of trial model. For example, research suggests that common biases such as framing biases (preferring the sure option when a decision is framed as a gain and the risky option when a decision is framed as a loss) that influence risky decisions also influence plea decisions (Garnier-Dykstra & Wilson, 2019). This bias is not accounted for by the shadow of trial model since precise comparisons between plea and trial outcomes will be the same regardless of how the decision is framed. In addition, the model does not account for demonstrated complexity specific to plea decision-making.

Research has shown that plea decisions are, in at least one way, more cognitively complex than other risky decisions (e.g. Helm, 2018). This complexity is caused by the fact that plea decisions also involve consideration of factual guilt or innocence, and invoke strong related values (most obviously only wanting to plead guilty if you have actually committed a crime, see Helm & Reyna, 2017). Although defendants can be permitted to plead guilty without admitting guilt (e.g. through Alford pleas in some US states, see Diehm, 2015; Redlich & Özdoğru, 2009), research suggests that defendants are more likely to plead guilty when they are guilty, even when controlling for the probability of conviction (Helm & Reyna, 2017; Redlich & Shteynberg, 2016; Schneider & Zottoli, 2019; Tor et al., 2010; Wilford et al., 2020). In adult populations, research suggests that this is driven, at least in part, by value-based decision-making grounded in not wanting to plead guilty (and essentially admit to committing) a crime that was not committed (Helm & Reyna, 2017; Tor et al., 2010). Innocent defendants may also be more cognitively disposed to picking the "risky" option of trial due to innocence placing them in more of a loss frame (Tor et al., 2010). Despite this, both empirical evidence (including experimental work and examination of real plea decisions) and legal literature suggest that innocent defendants do, under the right conditions, decide to plead guilty (e.g. Blume & Helm, 2014; Covey, 2013; Dervan & Edkins, 2013; Henderson & Levett, 2018; Wilford et al., 2020; Wynbrandt, 2016; Zimmerman & Hunter, 2018; Zottoli et al., 2016). Thus, while innocence influences defendants who are deciding whether to plead guilty and leads to a reluctance to plead guilty, it is not dispositive.

Research also suggests that guilt or innocence can interact with other factors influencing defendant decision-making, including the probability of conviction at trial (Tor et al., 2010), the framing of plea decisions (Garnier Dykstra & Wilson, 2019), advocate recommendations (Henderson & Levett, 2018), and, in some cases, the plea discount

(Redlich & Shteynberg, 2016; but see Dervan & Edkins, 2011; Schneider & Zottoli, 2019). These studies indicate the potentially complex nature of plea decisions in both guilty and innocent defendants, and also suggest that it may be possible to structure plea discounts to encourage guilty but not innocent defendants to plead. However, existing research provides relatively little insight into how incentives to plead might be tailored to appeal to guilty but not innocent defendants and, relatedly, little insight into how defendants may behave differently when facing relatively small differences in outcomes between plea and trial and when facing more categorical differences.

In addition, to date little or no research has examined incentives to plead guilty outside the North American context, where generous incentives attract significant attention. Discounts offered to defendants who plead guilty outside of the North American context tend to be considered relatively modest comparatively (Roberts & Bradford, 2015; Vamos, 2009). Thus, it may be easy to presume that incentives will not influence innocent defendants in these contexts. However, the psychological complexities of plea decision-making mean that even seemingly modest incentives can lead to decisions being made in a way that is not normatively desirable from a legal perspective, most obviously when innocent defendants plead guilty.

This paper draws on psychological theory, specifically Fuzzy-Trace Theory (FTT), and existing research to develop and test predictions relating to the effects of different incentives to plead, and how such incentives are likely to interact with defendant guilt and the probability of conviction. These predictions are tested specifically in the context of incentives to plead guilty in England and Wales, and results provide specific insight into the appropriateness of incentives to plead currently offered in that context as well as into plea decision-making more generally.

Fuzzy-Trace Theory

FTT is a dual process theory of memory and decision-making (Brainerd & Reyna 2019; Reyna, 2012). According to FTT, when an individual encodes information (such as the potential outcomes as a result of plea and trial) they encode this information in two ways - as verbatim information (literal, veridical and detailed representations, e.g. one year of custody) and as gist (fuzzy, contextual, and meaning-based representations, e.g. a short period of custody) (Brainerd & Reyna 2019; Reyna & Brainerd, 1995). When an individual makes a decision, they then rely more on one of these types of representation. Reliance on more verbatim information is associated with "verbatim processing," which is likely to be superficial but precise and influenced by exact numbers and details. Reliance on more gist information is associated with "gist processing," which is meaning-based and non-precise (Helm & Reyna, 2017). Decision-making based on gist tends to be non-compensatory, meaning that certain risks (e.g. risks of a categorically more negative outcome) will not be taken regardless of potential benefits. This effect has been demonstrated in risky decision-making in the sexual health context, where decision-makers relying more on verbatim may be willing to trade risks and rewards in deciding whether to engage in unprotected sex, while decision-makers relying more on gist are more likely to conclude that no benefits outweigh the possible negative outcomes (Reyna et al., 2011). Reliance on gist can explain framing in risky choice problems, since options can boil down to a definite gain vs. a possible gain in the gain frame and a definite loss vs. a possible loss in the loss frame, leading to a preference for the sure option in the gain frame and the risky option in the loss frame (Reyna et al., 2014; Reyna & Brainerd, 2011).

Whether a person relies on gist or verbatim representations is determined by (a) the accessibility of those representations in memory (e.g. gist is more stable over time than verbatim representations); (b) individual differences in propensity to rely on gist or verbatim representations (Helm & Reyna, 2017); and (c) external factors (e.g. the specificity of

retrieval cues) promoting reliance on a particular type of representation (e.g. Mills et al., 2008; Reyna & Kiernan, 1994). Understanding the external cues that push decision-makers towards reliance on gist or verbatim in decision-making is important since legal systems have control over external factors relevant to the plea decision (e.g. decision options). According to FTT, adults will tend to rely on the simplest level of gist necessary to make a decision. So, where decision options are categorically different from one another in a meaningful way, adults will make decisions based on these categorical differences (Reyna, 2012). Where categorical differences fail to discriminate options, ordinal differences (e.g. more vs. less) will be used where possible. Fine-grained distinctions will typically only be relied upon where required by a task (e.g. in a mathematical problem requiring precise calculations to reach an answer, or in a memory task asking decision-makers to recognise exact numbers or words, see Reyna & Kiernan, 1995) or where simpler distinctions do not provide a solution.

Where qualitative values are important in decision-making, these values can create a meaningful distinction between options. For example, importance of the value "I would not want to risk death" would make an option with a risk of death and an option without that risk categorically different from one another in a meaningful way. Since there is a categorical distinction between options, individuals can rely on gist and will be likely to do so. As a result, they will be likely to make a decision based on this categorical distinction, at least in the absence of other relevant categorical distinctions. The relationship between values and gist occurs because values are typically represented in long-term memory as vague gists that are applicable to a range of scenarios and with insufficient specificity to be applicable in verbatim-based decision-making (Fukura et al., 2013; Reyna & Casillas, 2009). The relationship can work in both directions – a person relying on gist is more likely to access and decide in accordance with their values, and where values are highly important in a decision-making process, they are likely to push decision-makers more towards gist-processing.

The Relationship Between Innocence and Gist

When innocent defendants are confronted with a plea decision, strong fairness-related values are likely to be hugely important – most notably a desire to not plead guilty to something that you did not do. This desire has been shown to be important in existing work examining gist in guilty plea decision-making (Helm & Reyna, 2017). This importance of innocence is reflected in research showing the general reluctance of innocent defendants to plead guilty (e.g. Tor et al., 2010). This value is gist-based, will be represented categorically in long-term memory, and makes a simple, qualitative, distinction (factual guilt or innocence) central to the decision-making process. Thus, innocent decision-makers will typically have a categorical distinction upon which to base their decision. In the absence of other categorical distinctions, this distinction is likely to be determinative and decision-makers will not consider more fine-grained distinctions between options.

On the other hand, those who are factually guilty will typically have accepted some risk of conviction and are unlikely to be driven to the same extent by visceral and fairness-related concerns. The equivalent value in guilty defendants, wanting to admit to true wrongdoing, has not been shown to be such a key consideration in the literature and research suggests that guilty defendants are not reluctant to go to trial in the same way as innocent defendants are reluctant to plead guilty (e.g. Tor et al., 2010). As a result, guilty individuals are far less likely to have this meaningful guilt-related categorical distinction upon which to base their decision-making. In the absence of other categorical distinctions, guilty defendants will therefore have to consider more fine-grained distinctions in their decision-making and will be pushed towards more verbatim processing.

Existing research provides some support for the idea that innocent defendants are likely to be more driven by values and guilty defendants are more likely to be driven by outcomes. In one study examining rationales given for accepting or rejecting hypothetical

plea offers, researchers found that while innocent and guilty 'defendants' were both generally driven by outcomes in decisions to accept plea offers, innocent defendants were much more likely than guilty defendants to be driven by value-focused rationales when rejecting plea offers (Schneider & Zottoli, 2019).

Because factual guilt or innocence (a gist-based consideration) is likely to be important to innocent defendants, and creates a categorical distinction between decision options, FTT predicts that innocent defendants will be likely to rely on gist. Decisions will therefore be made in accordance with simple categorical distinctions. Where guilt / innocence is the only categorical distinction relevant this will therefore be determinative. Precise changes in expected value that do not influence the "gist" of options (e.g. because both sentences remain "short" custodial sentences) will not register in the decision-making process. However, where another meaningful categorical distinction is introduced, it will compete with factual guilt or innocence and force even innocent defendants to rely on more fine-grained distinctions.

On the other hand, guilty defendants will be far less likely to see such a clear categorical distinction between decision options, since guilt / innocence is less important to them. As a result, in the absence of another categorical distinction, there will be no meaningful categorical distinction to rely on and guilty defendants will be pushed towards more fine-grained distinctions and more verbatim processing.

Therefore, innocent defendants should make plea decisions in predictably different ways from guilty defendants. For innocent defendants, innocence will drive decision-making in the absence of other categorical distinctions, meaning that categorical but not purely quantitative distinctions between plea and trial will encourage guilty pleas. For guilty defendants, guilt will not drive decision-making in the absence of other categorical

distinctions, meaning that both categorical and quantitative incentives to plead will encourage guilty pleas.

Existing research provides some support for the prediction that innocent defendants are more likely to be reliant on gist-based processing (Tor et al., 2010). This research shows that in hypothetical vignettes when plea and trial options have an equal expected value (meaning the plea option is equivalent to the outcome if convicted at trial multiplied by the probability of conviction), "guilty" defendant's preferences do not change much based on the probability of conviction at trial. This finding would be expected if decisions were determined by precise expected value calculations, and risk preferences. On the other hand, "innocent" defendants given the same problems showed changes in decision-making despite absolute expected values remaining constant. Specifically, innocent defendants were less likely to plead when the probability of conviction was lower and more likely to plead when the probability of conviction was higher – suggesting an influence of factual innocence in addition to expected value (i.e. a general reluctance to plead until the probability of conviction became high).

Therefore, according to FTT, offering quantitative incentives to plead guilty but not incentives that change decision options from a categorical perspective has the potential to encourage guilty but not innocent defendants to plead guilty.

More Quantitative and More Categorical Incentives in England and Wales

In England and Wales, approximately two-thirds of defendants accused of serious crimes (crimes tried in the Crown Court rather than the Magistrates Court) plead guilty (Ministry of Justice, 2020, p9). Considering all crimes prosecuted by the Crown Prosecution Service, around 90% of convictions are obtained via guilty plea (Nobles & Schiff, 2019). The sentence for a criminal offence is set by a judge following conviction, in line with prescribed sentencing guidelines. Specific sentencing guidelines outline the discounts that can be made

to a sentence on the basis that a defendant entered a guilty plea rather than being convicted at full trial (Sentencing Council, 2017). This guideline notes that where a defendant pleads guilty at the earliest opportunity a sentence reduction of one-third should be made (subject to certain exceptions) (Section D1). This reduction decreases the further along in proceedings a plea is entered (Section D2). Importantly, the reduction can result in imposing a different "type" of sentence, for example by reducing a custodial sentence to a community sentence, or by reducing a community sentence to a fine (Section E1). Whether a different type of sentence is awarded is at the discretion of the sentencing judge. A different type of sentence would typically be awarded in borderline cases where a 1/3 reduction could be taken into account in this way rather than through reducing sentence length. The guidelines specifically note that where a court has imposed a different type of sentence there "should normally" be no further reduction on account of the guilty plea (although a further reduction may be justified based on other factors) (Section E1). As a result, where a sentence type discount is awarded, a sentence length discount will not normally be given on the basis of the guilty plea.

Thus, the system in England and Wales provides a good example of incentives that can be considered more quantitative (1/3 changes in length to a sentence) and more categorical or qualitative (changes in the type of sentence). The 1/3 change in sentence length is unlikely to alter the overall meaning of the plea options (since it is likely insufficient to change the more categorical classification of outcomes, for example as a "long" or "short" period in custody) but would change the strength of the deal in a more quantitative, expected value type analysis. However, the change in sentence type is likely to alter the plea decision in both more quantitative ways (e.g. because more days of freedom are lost as the result of custody) and more categorical ways. A different sentence type has a qualitatively as well as a quantitatively different effect on a person's life (see Helm & Reyna, 2017) and will change the "gist" of plea options. For example, it could change a risk of custody to the definite

avoidance of custody (a categorical distinction). The importance of the categorical difference generated by a change in sentence type can be seen by looking at the different values or principles that might become evoked in the decision-making process when a change in sentence type is offered. For example, where a person would need to risk custody by going to trial but not by pleading guilty a value like "I would never risk going to prison" or "I would never risk being taken away from my family" may become engaged in the decision-making process. Note that these types of values are unlikely to be evoked in a case involving a 1/3 sentence reduction only, where both plea and trial would involve prison and being away from family. It is unlikely that absent additional complicating factors someone would feel, for example, that they could accept two years away from family but could never risk three.

In the case of innocent defendants, thought to be relying more on gist, where a 1/3 sentence length reduction is offered, this will not alter the decision-making process, since gist remains the same. In such cases, the gist of the plea option is pleading guilty when innocent and receiving a certain "gist" sentence and the gist of the trial option is not pleading guilty when innocent, and risking the same "gist" sentence. This pushes the defendant towards trial (since not pleading guilty when innocent and receiving a possible penalty is better than pleading guilty when innocent and receiving a certain penalty). As a result, it is likely that even when a 1/3 sentence reduction is offered, innocent defendants will continue to go to trial. On the other hand, where an incentive is offered that changes the meaning of options, a new gist is introduced that might encourage the defendant towards a guilty plea. For example, where a custodial sentence is risked at trial but can be avoided by pleading, a value of not wanting to risk custody will begin to compete with the value of not wanting to plead guilty when innocent. In the case of guilty defendants, thought to be relying more on verbatim, both the 1/3 sentence reduction and sentence type reduction have the potential to influence the decision by influencing the precise risks and rewards involved in the decision.

Predictions

Based on previous work, and consistent with FTT, it is expected that people who are 'guilty' will plead guilty more often than those who are innocent (e.g. Tor et al., 2010; Helm & Reyna, 2017). This study also tests a number of predictions specific to FTT.

1. People who are guilty will be influenced by both more categorical and more quantitative incentives, even where these incentives are relatively modest. They will therefore be responsive to a 1/3 sentence-length reduction and a sentence-type reduction.

This prediction is based on the FTT-informed expectation that guilty defendants will not have a clear categorical driver in plea decisions (absent a categorical distinction introduced by the plea discount) and therefore decisions will be influenced by the introduction of a categorical discount (which will create a categorical distinction to be relied upon), or, where such a discount is not present, by more fine-grained distinctions between decision options such as differences in expected value (which will influence decisions in the absence of categorical distinctions).

2. People who are innocent will only be encouraged to plead guilty when a categorically more lenient sentence (i.e. a different type of sentence) is given when pleading guilty compared to trial.

This prediction is based on the FTT-informed expectation that innocent defendants will tend to have a categorical distinction central to their decision-making (pleading guilty when innocent vs. not pleading guilty when innocent) due to important relevant values. Thus, innocent defendants will be likely to rely on gist and will be driven by this categorical distinction in the absence of a categorical distinction to compete with it.

3. A categorical difference between plea and trial will introduce a gist that competes with not wanting to plead guilty when innocent that will make factual innocence less important in decision-making.

This prediction is based on the FTT-informed expectation that changing decision options in categorical ways (such as altering the charges involved in a plea decision, see Helm & Reyna, 2017) will influence the gist relied on in plea decisions, and thus ultimate decisions.

METHODS

Participants

Participants were 3,375 adults recruited from the Prolific platform.[1] This sample size included approximately 70 participants in each experimental condition in the study (35 participants who received a less detailed community service description, and 35 participants who received a more detailed community service description). This sample size was chosen to ensure assumptions of analysis methods used would be satisfied, and to roughly match sample size in previous work using FTT and applying similar paradigms (e.g. Helm & Reyna, 2017).

Participants completed the study online and participation took an average of between three and four minutes. All participants completed a Captcha verification prior to completing the study and received a small monetary incentive upon completion. Participants ranged in age from 19 to 83 (M = 36.04, SD = 13.27). All participants were British. Eighty-seven percent identified as White, 5.4% identified as Asian, 3.5% identified as Black, and the remainder identified as another racial group. This distribution is broadly representative of the racial distribution in the UK, where the most recent census results listed 86% of the population as White, 7.5% of the population as Asian, and 3.3% of the population as Black (Office for National Statistics, 2011). Approximately 34% of participants identified as male, 66%

[1] Post hoc analyses confirm $1 - \beta > 80\%$ for detecting an odds ratio of 1.3 (small; Chen, Cohen, & Chen, 2010) or an effect size f of .1 (small; Cohen, 1988). Analyses conducted using G*Power (Faul, Erdfelder, Lang, & Buchner, 2007).

identified as female, and the remainder identified as another gender group (e.g. gender fluid). This distribution was more weighted towards females than the UK population as a whole is. The most recent census described the population as approximately 51% female, and approximately 49% are male (Office for National Statistics, 2011). Analyses described below were conducted controlling for gender, and significant results remained the same.

Design Overview

Participants made decisions in hypothetical plea vignettes. Participants were randomly assigned to one of 48 experimental cells produced by a 2 Guilt (Innocent vs. Guilty) x 4 Discount (9-year custodial sentence reduced to 6-year custodial sentence, 3-year custodial sentence reduced to 2-year custodial sentence, 3-year custodial sentence reduced to 2-year community sentence, and 1-year custodial sentence reduced to 1-year community sentence) x 6 Probability of Conviction (30%, 40%, 50%, 60%, 70%, 80%). Half of the participants also answered questions about the values that were most important to them when making their decision. More information on the design and how conditions were chosen is provided below.

Materials and Procedure

Participants were presented with a hypothetical plea vignette in which they were told they had been accused of theft and had to make a decision about whether to plead guilty to the theft or go to trial. Theft was chosen due to the varying degrees of theft allowing realistic manipulations of sentence length, and to avoid as far as possible confounding from decisions being driven by social stigma (as they may be, for example, in cases involving sexual offences).

Participants were randomly assigned to a guilt condition (they were either told they did commit the theft that they were accused of, or they did not commit the theft they were accused of) and a probability condition (there was said to be either a 30%, 40%, 50%, 60%, 70%, or 80% probability of conviction at trial). This range of probabilities was included in order to vary the strength of deals participants were offered from an expected value standpoint, and to

capture potential changes in gist relied upon. If a principle such as "I would not plead guilty when innocent" is governing a decision then probability of conviction will be unlikely to have an effect on decision-making. On the other hand, if a principle such as "I would not want to risk custody" is governing a decision then decision-makers may plead more as the risk of custody becomes greater (i.e. as the probability of conviction increases).

Participants were also randomly assigned to one of the following four sentence discount conditions, designed to mirror possibilities under the sentencing guidelines in England and Wales:

1. Nine years of custody if convicted at trial and six years of custody if convicted via plea (referred to as the SentenceLength9v6 condition).
2. Three years of custody if convicted at trial and two years of custody if convicted via plea (referred to as the SentenceLength3v2 condition).
3. Three years of custody if convicted at trial and two years of community service if convicted via plea (referred to as the SentenceLength+Type condition).
4. One year of custody if convicted at trial and one year of community service if convicted via plea (referred to as the SentenceType condition).

Conditions were designed to be controlled experimentally, while also representing realistic scenarios in the England and Wales context. The SentenceLength9v6 and SentenceLength3v2 conditions were both included despite both involving only sentence length in order to ensure results were driven by the proportion of the discount (1/3) not changing the gist of the sentence (e.g. from long to short) rather than being driven by the absolute length of the discount (3 years and 1 year respectively). The decision in condition 3 is unlikely to occur in practice due to sentencing guidelines encouraging judges not to award a sentence length discount where a "type" discount is given (Sentencing Council, 2017). However, the condition was included to provide a comparison to condition 2. The number of years of sentence associated with each

option remain the same in both conditions (3 v 2), but in condition 3 the reduced sentence length is also combined with a change in sentence type. Table 1 provides an overview of the distinctions between conditions.

[Insert Table 1 Here]

Participants were given a brief description of community service. Half of the participants were simply told that this would mean they would not go to prison but that restrictions would be placed on their movements. The other half received the same description, but were also given examples of such restrictions, specifically needing to engage in community work or to be home by a set time in the evenings.[2] This description was added half way through the study to address a concern that participants may not have an understanding of what 'restrictions on movement' would mean. The level of detail included about community service was included as a factor in the primary regression analysis but was not significant and did not influence the significance of other variables.

After making their plea decisions, a subset of participants (n=1624) also answered questions about considerations when pleading (the other sub-set answered different questions following the plea decision, involving how participants perceived plea and trial, which are not analysed for the purposes of this study). Specifically, participants were asked what was most important to them when deciding whether to plead guilty: whether they committed the crime or not, the need to avoid custody, or another consideration. Participants were able to select one or multiple options.

RESULTS

[2] This resulted in approximately 35 participants in each condition receiving each of the community service descriptions.

Plea Decisions

Overall, participants pled guilty in 34.3% of cases and chose to go to trial in 65.7% of cases. Guilty defendants pled guilty more often than innocent defendants (13.1% of innocent defendants pled guilty compared to 55.7% of guilty defendants, $\chi^2 = 679.04$, $p < .001$). This finding confirmed expectations based on FTT and existing work, and suggested that the manipulation of guilt / innocence was effective.

A linear regression analysis was used to analyse data, due to research suggesting that a general linear model with an identity link rather than a logit link function is generally the best strategy to estimate causal effects of treatments on binary outcomes (Gomila, 2020), particularly in models including a number of interaction terms. A regression was preferred to an analysis of variance (ANOVA) since treating probability of conviction as a continuous variable made results easier to interpret. However, a supplementary ANOVA was also conducted. Results were very similar, and are reported in the Supplemental Materials.

The regression analysis included Probability of Conviction (30%, 40%, 50%, 60%, 70%, 80%), Guilt (0=innocent, 1=guilty), Discount (SentenceLength9v6, SentenceLength3v2, SentenceLength+type, Sentencetype), and all interactions between those variables. The discount variable was dummy coded into three variables comparing the Sentence Length 9v6 Discount condition to each of the other levels of Discount, as detailed in Table 2. The dependent variable was Plea Decision (0=plea, 1=trial). The results of the regression are reported in Table 2. Removing the three-way interaction terms from the model reduced the overall model fit (to R = .539, $R^2 = .290$, Adjusted $R^2 = .288$, *SE* = .40), so all interaction terms remained in the model.

Follow-up analyses were conducted to explore the nature of the three-way interactions among Guilt, Probability, and 9v6 / Length + Type comparison, and among Guilt, Probability, and 9v6 / Type comparison (which moderate all other significant effects).

Relationships between Guilt, Probability, and Discount are displayed in Figure 1 to assist in the interpretation of results.

Linear regression analyses were conducted separately for each Discount level, and included Guilt, Probability, and the interaction between Guilt and Probability as predictors. These results showed an interaction between Guilt and Probability in the Sentence Length conditions, although note that this interaction just missed two-tailed significance in the SentenceLength3v2 condition (SentenceLength9v6: B=-.063, SE = .015, β = -.311, t=-4.353, p <.001; SentenceLength3v2: B=-.027, SE = .016, β = -.134, t=-1.714, p =.087). Follow-up analyses in each of these conditions showed that the relationship between Probability and Plea Decision (pleading guilty more as the probability of conviction increased) was stronger in the guilty condition than in the innocent condition (SentenceLength9v6 Guilty, B=-.08, SE = .01, β = -.26, t=-5.5, p <.001, Not Guilty B=-.01, SE = .01, β = -.10, t=-2.07, p =.040; SentenceLength3v2 Guilty, B=-.05, SE = .01, β = -.26, t=-5.52, p =.001, Not Guilty B=-.02, SE = .01, β = -.17, t=-3.36, p =.001). This interaction was not significant in the two conditions including a sentence-type discount (p =.912 and p = .860, respectively), meaning that the influence of probability of conviction on plea decision was not significantly different based on guilt in the two conditions where a sentence 'type' distinction was present.

Values Analyses

The values participants rated as most important to them when making their plea decision were assessed to test prediction 2, that a categorical difference between plea and trial will introduce a gist that competes with not wanting to plead guilty when innocent, that will make factual innocence less important in decision-making.

Overall, 75.9% of participants said that whether they were guilty or not was the most important, or one of the most important, considerations for them when deciding whether to plead guilty. Thirty percent of participants said that the need to avoid custody was the most

important, or one of the most important, considerations for them when deciding whether to plead guilty. In the values analyses, unlike in the decision analyses reported above, Probability was not significant and did not significantly interact with other variables. For this reason, ANOVA results were most easily interpretable and an ANOVA was used for primary analyses.

The importance of being guilty or not.

First, analyses examined when participants rated whether they were guilty or not as the most important, or one of the most important, considerations in making their plea decision. A univariate ANOVA was conducted, with whether being guilty or not was rated as the most important, or one of the most important, considerations in the decision (1) or not (0) as the dependent variable, and Guilt, Probability of Conviction, and Discount as predictors.

Results revealed three significant effects. First, a significant main effect of Guilt ($F(1,1536)=75.823$, $p<.001$, $\eta_p^2 = .047$) and a main effect of Discount ($F(3,1536)=4.995$, $p=.002$, $\eta_p^2 = .010$). Finally, a significant interaction between Guilt and Discount ($F(3,1536)=4.361$, $p=.005$, $\eta_p^2 = .008$), which is displayed in Figure 2. For participants in the guilty condition, whether they were guilty was the most important, or one of the most important, considerations for approximately the same proportion of participants across the Discount conditions. For participants in the innocent condition, whether they were guilty or not was the most important, or one of the most important, considerations to significantly less participants in the two sentence type conditions, when compared to the two sentence length only conditions.

[Insert Figure 2 Here]

The risk of custody

Second, analyses examined when participants rated the need to avoid custody as the most important, or one of the most important, considerations in making their plea decision. A univariate ANOVA was conducted, with whether the need to avoid custody was the most important or one of the most important considerations in the decision (1) or not (0) as the dependent variable, and Guilt, Probability of Conviction, and Discount as predictors.

Results revealed two significant effects. First, a main effect of Guilt ($F(1,1536)=25.608$, $p<.001$, $\eta_p^2 = .016$) such that participants in the guilty condition were more likely to rate needing to avoid custody as the most important, or one of the most important, considerations to them when making their plea decisions ($M_{guilty} = .360$, $SE = .016$, $M_{innocent} = .244$, $SE = .016$).

Second, results revealed a main effect of Discount ($F(3,1536)=17.415$, $p<.001$, $\eta_p^2 = .033$). Unsurprisingly this consideration was rated as the most important, or one of the most important, considerations to participants in the two conditions in which it was possible to definitively avoid custody (Length+Type and SentenceType) significantly more than it was in the SentenceLength discount conditions ($M_{sentencelength9v6} = .205$, $SE = .022$, $M_{sentencelength3v2} = .238$, $SE = .024$ [$p=.307$], $M_{length+type} = .400$, $SE = .023$ [$ps<.001$], $M_{SentenceType} = .369$, $SE = .023$ [$p<.001$, $p<.001$, $p=.315$]).

DISCUSSION

Key Findings

The results of this study support previous research suggesting that factual guilt is likely to influence plea decisions, but also that under the right conditions innocent defendants will plead guilty (e.g. Helm & Reyna, 2017; Redlich & Shteynberg, 2016; Tor et al., 2010). Results also provide support for FTT-based predictions that while purely quantitative incentives to plead are likely to appeal only to guilty defendants, more categorical, qualitative, incentives to plead are likely to appeal to both guilty and innocent defendants.

Participants in the innocent condition in this study showed significant resistance to pleading guilty when the more quantitative incentive was offered even at high levels of probability of conviction. However, when a categorical incentive was offered (the change in sentence from custody to community service), they showed a similar trend to guilty defendants in terms of pleading guilty more as the probability of conviction increased (note that this relationship was also significant when the quantitative incentive was offered but was weaker in innocent defendants than in guilty defendants). An analysis of values showed that where this categorical incentive was offered the value of not wanting to plead guilty when innocent became less important, and the value of wanting to avoid custody became more important. Thus, results suggest categorical differences have the potential to lead innocent defendants to plead and to reduce the importance of factual guilt and innocence in their decision-making process. The more quantitative incentive did not have this effect, although guilty defendants did appear responsive to this incentive – pleading guilty at relatively high rates when the probability of conviction was high.

Limitations

These findings should be interpreted in light of a number of important limitations. First, the study used a nested rather than fully factorial design. This design was chosen to maximise ecological validity (by using real incentives that might be faced by defendants in England and Wales). Caution should be taken when extrapolating the findings to other contexts or incentives (including to the US plea-bargaining system, as discussed below). However, the findings are supported by theoretically founded predictions, and thus would be expected to replicate in circumstances where the same theoretical considerations apply (i.e. a comparison between a more quantitative and more qualitative incentive).

The participants in the study were not making real plea decisions and their decisions were considerably simpler than those made in practice. This simplicity was reflected in the

short time taken to make decisions in the study. In practice, an array of other considerations may be important to real defendants and influence their decisions, for example pressure from lawyers or the legal system (which has been shown to be important) (Baldwin & McConville, 1977; Henderson & Levett, 2019; Henderson & Shteynberg, 2020), remand in custody (Edkins & Dervan, 2018; Peterson, 2020), or informal charge discounts awarded (Helm, 2019). In practice even in the structured system in England and Wales, defendants can be faced with other incentives to plead. For example, defendants may also be incentivised to plead where the prosecutor agrees to drop or reduce charges if a plea is entered, or where they can gain informal benefits by pleading guilty, such as release from custody or lower costs (for more information see Baldwin & McConville, 1977; Helm, 2019). In addition, the vignettes presented to participants in this study were highly simplified in order to ensure participant understanding. This simplification was necessary experimentally but meant that in important ways vignettes differed from real decisions. For example, in real decisions defendants risking a custodial sentence would also consider the fact that they would be unlikely to serve the duration of their sentence in custody, and would likely be released early. The community service description given to participants in this study was also quite general, and future work should test predictions with more strenuous and specific requirements of community service to ensure generalisability for different defendants.

Therefore, this experimental research forms only part of the evidence that should be considered in evaluating guilty plea policy. However, it is an important and necessary part. The research was able to use a controlled experimental design that allowed close measurement of the effects of experimental manipulations and minimized confounding. While it is true that people may make decisions differently in an experimental study than they would in practice, emerging research on guilty pleas suggests that field work and experimental work are likely to converge on important findings (e.g. relating to children pleading guilty when innocent; Helm

et al., 2018a and Zottoli et al., 2016, and innocent defendants pleading guilty; Covey 2013 and Tor et al., 2010) and can support one another. Experimental research can therefore provide important insight in planning and interpreting field work and illuminating potential causal mechanisms for observed phenomena. Currently examining guilty pleas from innocent defendants is hard in practice due to the difficulty determining factual guilt or innocence, particularly in systems where appealing a conviction via plea is often difficult (e.g. Blume & Helm, 2014; Nobles & Schiff, 2019). However, this work might be drawn on in order to design instruments for interviews with defendants, and could bolster the reliability of findings relying on self-reported guilt or innocence. In addition, it should be noted that if anything the fact participants did not actually have to experience custody might have been expected to reduce the effects of our manipulations (since people may be more willing to risk custody in a vignette than in reality), meaning that effects could even be stronger in practice.

Ultimately, the findings in this study are unlikely to reflect magnitudes that would be expected in real cases, but still capture an important part of the cognition underlying plea decision-making. There is also a strong argument to suggest that the system should not be offering defendants incentives that members of the public (such as the participants here) would expect to encourage innocent people to plead guilty.

Implications for Policy

The findings in this paper contribute to the policy debate in England and Wales concerning whether pleading guilty should influence the "type" of sentence that a defendant receives. While current Sentencing Guidelines allow for a different type of sentence to be imposed where a defendant pleads guilty compared to what would be imposed at trial, this is not uncontroversial (e.g. Roberts & Harris, 2017; Helm, 2018). The research reported in this paper suggests that allowing a guilty plea to influence sentence type might be leading innocent defendants to plead guilty and undermining the role of guilt or innocence in their decision-

making process, thus undermining important normative goals of the criminal justice system. On the other hand, the research suggests that the current 1/3 sentence length discount may well be appropriate in encouraging guilty but not innocent defendants to plead guilty, and so supports the retention of this discount. The importance of ensuring appropriate discounts when defendants plead guilty can be seen in the recent Post Office Scandal, almost all of the sub-post masters and mistresses acquitted of false accounting and theft charges by the Court of Appeal had initially pleaded guilty to charges against them (Hamilton & Others v Post Office Limited, 2021). Following their acquittal, some of the successful appellants stated that they had pled guilty to avoid the risk of custody (e.g. Ashworth, 2021).

In addition, results suggest that defendant autonomy may be undermined where sentence type discounts are introduced, since considerations stemming from those discounts may undermine a desire to maintain innocence, even in innocent defendants. Thus, such discounts undermine the ability of an individual who has not committed a crime to act in accordance with their core goals and values by contesting guilt at trial (for more information on this concept of autonomy, see Bratman, 2007; Dworkin, 1994). This paper therefore provides important support for retaining the 1/3 sentence length discount in England and Wales but removing the ability of this discount to change sentence type imposed.

The findings from this paper can also influence the regulation of incentives to plead guilty in other jurisdictions. However, findings do not necessarily directly generalize to systems that differ significantly from the system in England and Wales, such as the United States plea-bargaining system. In that context, discounts are largely determined by individual prosecutors rather than judges and are much more varied. Sentence-length reductions are often of more than a third of sentence and are therefore much more likely to categorically change decision options for defendants (e.g. from a long custodial sentence to a short custodial sentence). As a result, sentence discounts may not be so distinguishable from the “type”

discounts studied in this paper, and may similarly lead innocent defendants to systematically plead guilty. This suggestion should be explored in research examining the US system directly. If confirmed, and if such discounts are still retained, their appeal to innocent defendants should be recognised and justified. Some legal literature does suggest that innocent defendants pleading guilty may not be problematic because innocent defendants are convicted at trial too and should be able to benefit from plea reductions, since they face the possibility of conviction and a harsher penalty at trial (Garrett, 2016). If it is acknowledged that incentives to plead are systematically appealing to the innocent, this should be recognised through an appeals system that does not make it unduly difficult for defendants to appeal following a guilty plea (e.g. Blume & Helm, 2014).

Implications for Research

These findings may also provide useful context for plea researchers more generally examining the relationship between guilt and the probability of conviction at trial. For example, some studies suggest that innocent defendants are influenced by the probability of conviction, potentially even more than guilty defendants where the baseline position of guilty defendants is to plead guilty (e.g. Wilford et al., 2020; Zimmerman & Hunter, 2018). However, these studies have provided choices in which plea outcomes differ categorically from outcomes if convicted at trial (lab work vs. a charge of academic dishonesty, and a long vs. short period of custody respectively in the studies cited). This research suggests that the probability of conviction may influence innocent defendants far less when outcomes that do not differ categorically are offered (e.g. two short periods of lab work, or two short periods of custody, differing in specific magnitudes). Future research should build on the work in this paper, and specifically examine other more quantitative reductions (e.g. reductions in the length of a community sentence) and other more categorical reductions (e.g. reductions from a community sentence to a fine) in order to extend the generalizability of results. Future research should also

examine other factors likely to influence plea decisions in practice and interact with findings of this work. Such factors should include factor that might be race (since research suggests that in practice those in different racial groups might be making plea decisions differently; e.g. Hood, 1992), gender (which influence decisions in a way not captured by this study), and attorney influence (which has been shown to be important; e.g. Henderson & Shteynberg, 2020). Future research should also test the predictions of this study in a population who have really made plea decisions, to ensure that results replicate among the people who actually need to make plea decisions in practice.

Empirical research has the potential to ensure that convictions obtained via plea are normatively acceptable in practice, most notably through maximising accuracy of convictions via plea. This paper provides an important contribution to such research by showing that ensuring gist-equivalence in plea and trial options so far as possible is important in ensuring respect for the presumption of innocence, the accuracy of convictions, and the autonomy of those making plea decisions.

REFERENCES

Ashworth, A. (2006). Four threats to the presumption of innocence. *International Journal of Evidence and Proof, 10*(4), 241-279.

Ashworth, J. (2021). South Warnborough woman's Post Office conviction overturned. The Basingstoke Gazette. https://www.basingstokegazette.co.uk/news/19254393.south-warnborough-womans-post-office-conviction-overturned/.

Baldwin, J. & McConville, M. (1977). *Negotiated Justice: Pressures to Plead.* Robertson Publishing.

Bibas, S. (2004). Plea bargaining outside the shadow of trial. *Harvard Law Review, 117,* 2463-2547.

Bjerk, D. (2007) Guilt shall not escape or innocence suffer? The limits of plea bargaining when defendant guilt is uncertain. *American Law and Economics Review, 9*(2), 305-329.

Blume, J. H., & Helm, R. K. (2014). The unexonerated: Factually innocent defendants who plead guilty. *Cornell Law Review, 100*(1), 157-191.

Bordens, K. S., & Bassett, J. (1985). The plea bargaining process from the defendant's perspective: A field investigation. *Basic and Applied Social Psychology, 6,* 93-110. doi: 10.1207/s15324834basp0602_1.

Brainerd, C. J., & Reyna, V. F. (2019). Fuzzy-trace theory, false memory, and the law. *Policy Insights from the Behavioral and Brain Sciences*, *6*(1), 79-86. doi: 10.1177/2372732218797143.

Bratman, M. (2007). *Structures of Agency: Essays*. Oxford University Press.

Bushway, S. D., & Redlich, A. D. (2012). Is plea bargaining in the shadow of trial a mirage? *Journal of Quantitative Criminology, 28,* 437-454. doi: 10.1007/s10940-011-9147-5.

Bushway, S. D., Redlich, A. D., & Norris, R. J. (2014). An explicit test of plea bargaining in the shadow of trial. *Criminology, 52,* 723-754. doi: 10.1111/1745-9125.12054.

Chen, H., Cohen, P., & Chen, S. (2010). How big is a big odds ratio? Interpreting the magnitudes of odds ratios in epidemiological studies. *Communications in Statistics – Simulation and Computation*, *39*(4), 860– 864. doi: 10.1080/03610911003650383.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Erlbaum.

Covey, R. D. (2013). Police misconduct as a cause of wrongful convictions. *Washington University Law Review, 90,* 1133-1190.

Crown Prosecution Service, Code for Crown Prosecutors (2018). Retrieved from https://www.cps.gov.uk/publication/code-crown-prosecutors#:~:text=4.7%20The%20finding%20that%20there,on%20which%20they%20might%20rely.

Dervan, L. E., & Edkins V. A. (2013). The innocent defendant's dilemma: An innovative empirical study of plea bargaining's innocence problem. *Journal of Criminal Law & Criminology, 103,* 1.

Diehm, J. W. (2015). Pleading guilty while claiming innocence: Reconsidering the mysterious Alford plea. *University of Florida Journal of Law & Public Policy, 26,* 27-52.

Dworkin, R. (1994). *Life's Dominion.* Vintage Books.

Easterbrook, F. H. (2013). Plea bargaining is a shadow market. *Duquesne Law Review, 51,* 551-558.

Edkins, V. A. & Dervan, L. E. (2018). Freedom now or a future later: Pitting the lasting implications of collateral consequences against pretrial detention in decisions to plead guilty. *Psychology, Public Policy, and Law, 24*(2), 204-215. doi: 10.1037/law0000159.

Faul, F., Erdfelder, E., Lang, A., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175–191. doi: 10.3758%2FBF03193146.

Fukukura, J., Ferguson, M. J. & Fujita, K. (2013). Psychological distance can improve decision making under information overload via gist memory. *Journal of Experimental Psychology, 142*(3), 658. doi: 10.1037/a0030730.

Garnier-Dykstra, L. M., & Wilson, T. (2019). Behavioral economics and framing effects in guilty pleas: a defendant decision making experiment. *Justice Quarterly*, 1–25. doi: 10.1080/07418825.2019.1614208.

Garrett, B. L. (2016). Why plea bargains are not confessions. *William and Mary Law Review, 57*(4), 1415-1444.

Gomila, R. (2020). Logistic or linear? Estimating causal effects of treatments on binary outcomes using regression analysis. *Journal of Experimental Psychology: General.* doi: 10.1037/xge0000920.

Hamilton & Others v Post Office Limited [2021] EWCA Crim 577.

Helm, R. K. (2018). Cognitive theory and plea-bargaining. *Policy Insights from the Behavioral and Brain Sciences*, *5*(2), 195–201. doi: 10.1177/2372732218786974.

Helm, R. K. (2019). Constrained waiver of trial rights? Incentives to plead guilty and the right to a fair trial. *Journal of Law and Society, 46*(3), 423-447. doi: 10.1111/jols.12169.

Helm, R. K., (2021). Guilty pleas in children: Legitimacy, vulnerability, and the need for increased protection. *Journal of Law and Society.*

Helm, R. K., & Reyna, V. F. (2017). Logical but incompetent plea decisions: A new approach to plea bargaining grounded in cognitive theory. *Psychology, Public Policy, and Law*, *23*(3), 367–380. doi: 10.1037/law0000125.

Helm, R. K., Reyna, V. F., Franz, A. A., & Novick, R. Z. (2018a). Too young to plead? Risk, rationality, and plea bargaining's innocence problem in adolescents. *Psychology, Public Policy, and Law*, *24*(2), 180-191. doi: 10.1037/law0000156.

Helm, R. K., Reyna, V. F., Franz, A. A., Novick, R. Z., Dincin, S., & Cort, A. E. (2018b). Limitations on the ability to negotiate justice: Attorney perspectives on guilt, innocence, and legal advice in the current plea system. *Psychology, Crime & Law*, *24*(9), 915-934. doi: 10.1080/1068316X.2018.1457672.

Henderson, K. S., & Levett, L. M. (2018). Investigating predictors of true and false guilty pleas. *Law and Human Behavior, 42*(5), 427–441. doi: 10.1037/lhb0000297.

Henderson, K. S., & Levett, L. M. (2019). Plea bargaining: The influence of counsel. In *Advances in Psychology and Law* (pp. 73-100). Springer.

Henderson, K. S., & Shteynberg, R. V. (2020). Plea decision making: the influence of attorney expertise, trustworthiness, and recommendation. *Psychology, Crime & Law*, *26*(6), 527-551. doi: 10.1080/1068316X.2019.1696801.

Mills, B., Reyna, V. F., & Estrada, S. (2008). Explaining contradictory relations between risk perception and risk taking. *Psychological Science, 19*(5), 429–433. doi: 10.1111/j.1467-9280.2008.02104.x.

Ministry of Justice, Criminal Law Statistics (2020). Retrieved from https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/910530/criminal-justice-statistics-quarterly-march-2020.pdf

Missouri v Frye, 566 US 134 (2012).

Mnookin, R. H. & Kornhauser L. (1979) Bargaining in the shadow of the law: The case of divorce. *The Yale Law Journal, 88*(5), 950-997.

Natsvlishvili v Georgia, App. No. 9043/05, 29 April 2014.

Nobles, R. & Schiff, D. (2019) Criminal justice unhinged: The challenge of guilty pleas. *Oxford Journal of Legal Studies, 39*(1), 100-123. doi: 10.1093/ojls/gqy036.

Office for National Statistics, Ethnicity and National Identity in England and Wales: 2011 (2011). Retrieved from https://www.ons.gov.uk/peoplepopulationandcommunity/culturalidentity/ethnicity/articles/ethnicityandnationalidentityinenglandandwales/2012-12-11.

Peterson, N. (2020). Do detainees plead guilty faster? A survival analysis of pretrial detention and the timing of guilty pleas. *Criminal Justice Policy Review, 31*(7), 1015-1035. doi: 10.1177/0887403419838020.

Quirk, H. (2006). The significance of culture in criminal procedure reform: Why the revised disclosure scheme cannot work. *The International Journal of Evidence & Proof, 1*, 42-59. doi: 10.1350/ijep.2006.10.1.42.

R v Goodyear [2005] EWCA Crim 888.

R v Caley and Others [2013] Crim LR 342.

Redlich, A. D., Bibas, S., Edkins, V. A., & Madon, S. (2017). The psychology of defendant plea decision making. *American Psychologist, 72*(4), 339–352. doi: 10.1037/a0040436.

Redlich, A. D., & Shteynberg, R. V. (2016). To plead or not to plead: A comparison of juvenile and adult true and false plea decisions. *Law and Human Behavior*, *40*(6), 611- 625. doi: 10.1037/lhb0000205.

Redlich, A. D., & Özdoğru, A. A. (2009). Alford pleas in the age of innocence. *Behavioral Sciences & the Law*, *27*(3), 467-488. doi: 10.1002/bsl.876.

Reyna, V. F. (2012). A new intuitionism: Meaning, memory, and development in Fuzzy-Trace Theory. *Judgment and Decision Making, 7*(3), 332–359.

Reyna, V. F., & Brainerd, C. J. (1995). Fuzzy-trace theory: An interim synthesis. *Learning and individual Differences*, *7*(1), 1-75.

Reyna, V. F., & Brainerd, C. J. (2011). Dual processes in decision making and developmental neuroscience: A fuzzy-trace model. *Developmental review*, *31*(2-3), 180-206. doi: 10.1016/j.dr.2011.07.004.

Reyna, V. F., Estrada, S. M., DeMarinis, J. A., Myers, R. M., Stanisz, J. M., & Mills, B. A. (2011). Neurobiological and memory models of risky decision making in adolescents versus young adults. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37*(5), 1125-1142. doi: 10.1037/a0023943.

Reyna, V. F., Chick, C. F., Corbin, J. C. & Hsia, A. N. (2014). Developmental reversals in risky decision making: Intelligence agents show larger decision biases than college students. *Psychological Science, 25*(1), 76-84. doi: 10.1177/0956797613497022.

Reyna, V. F., Helm, R. K., Weldon, R. B., Shah, P. D., Turpin, A. G., & Govindgari, S. (2018). Brain activation covaries with reported criminal behaviors when making risky choices: A Fuzzy-Trace Theory approach. *Journal of Experimental Psychology General, 147*(7), 1094-1109. doi: 10.1037/xge0000434.

Reyna, V. F., & Kiernan, B. (1994). Development of gist versus verbatim memory in sentence recognition: Effects of lexical familiarity, semantic content, encoding instructions, and retention interval. *Developmental Psychology, 30*(2), 178–191. doi: 10.1037/0012-1649.30.2.178.

Reyna, V. F., & Kiernan, B. (1995). Children's memory and metaphorical interpretation. *Metaphor and Symbol*, *10*(4), 309-331. doi: 10.1207/s15327868ms1004_5.

Roberts, J.V., & Bradford, B. (2015). Sentence reductions for a guilty plea in England and

Wales: exploring new empirical trends. *Journal of Empirical Legal Studies, 12*(2), 187–210. doi: 10.1111/jels.12069.

Roberts, J. V., & Harris L. (2017). Reconceptualising the custody threshold in England and Wales. *Criminal Law Forum* 477-499.

Santabello v New York 404 Us 257 (1971)

Sentencing Council, Attitudes to guilty plea sentence reductions (2011). Retrieved from https://www.sentencingcouncil.org.uk/wp-content/uploads/Attitudes_to_Guilty_Plea_Sentence_Reductions_web1.pdf.

Sentencing Council, Reduction in sentence for a guilty plea: Definitive guideline (2017). Retrieved from https://www.sentencingcouncil.org.uk/wp-content/uploads/Reduction-in-Sentence-for-Guilty-Plea-definitive-guideline-SC-Web.pdf

Tor, A., Gazal-Ayal, O., & Garcia, S. M. (2010). Fairness and the willingness to accept plea bargain offers. *Journal of Empirical Legal Studies, 7,* 97–116. doi: 10.1111/j.1740-1461.2009.01171.x.

Tversky, A., & Kahneman, D. (1986). Judgment under uncertainty: Heuristics and biases. *Judgment and decision making: An interdisciplinary reader*, 38-55.

Vamos, N. (2009). Please don't call it plea bargaining. *Criminal Law Review, 9,* 617-630.

Wilford, M. M., Wells, G. L., & Frazier, A. (2020) Plea-bargaining law: The impact of innocence, trial penalty, and conviction probability on plea outcomes. *American Journal of Criminal Justice.* doi: 10.1007/s12103-020-09564-y.

Wynbrandt, K. (2016) From false evidence ploy to false guilty plea: An unjustified path to securing convictions. *Yale Law Journal, 126,* 545-565.

Zimmerman, D. M., & Hunter, S. (2018). Factors affecting false guilty pleas in a mock plea bargaining scenario. *Legal and Criminological Psychology, 23*(1), 53-67. doi: 10.1111/lcrp.12117.

Zottoli, T., Daftary-Kapur, T., Winters, G. M., & Hogan, C. (2016). Plea discounts, time pressures, and false guilty pleas in youth and adults who pleaded guilty to felonies in New York City. *Psychology, Public Policy, and Law, 22*(3), 250-259. doi: 10.1037/law0000095.

TABLES

Table 1: *Characteristics of Discount Type experimental conditions*

Condition	Reduction in custody	Reduction in sentence length	Reduction in sentence type
SentenceLength9v6	3 years	3 years (1/3 of sentence)	No
SentenceLength3v2	1 year	1 year (1/3 of sentence)	No
SentenceLength+Type	3 years	1 year (1/3 of sentence)	Yes
SentenceType	1 year	No	Yes

Table 2

Multiple Regression Predicting Plea Decision

Variable	B	SE	β	*t*
Probability of Conviction	-.01	.01	-.05	-1.16
Guilt	-.14	.06	-.15	-2.35*
Offer Dummy 1	.03	.06	.03	.54
Offer Dummy 2	-.08	.06	.07	-1.25
Offer Dummy 3	-.04	.06	-.04	-.65
Guilt x Probability of Conviction	.06	.02	-.28	-4.01*
Offer Dummy 1 x Guilt	-.12	.089	-.08	-1.38
Offer Dummy 2 x Guilt	-.35	.090	-.24	-4.03*
Offer Dummy 3 x Guilt	-.39	.09	-.28	-4.43*
Offer Dummy 1 x Probability of Conviction	-.01	.02	-.03	-.51
Offer Dummy 2 x Probability of Conviction	-.03	.02	-.13	-2.23
Offer Dummy 3 x Probability of Conviction	-.02	.02	-.08	-1.32
Guilt x Probability of Conviction x Offer Dummy 1	.04	.02	.10	1.57
Guilt x Probability of Conviction x Offer Dummy 2	.07	.02	.17	2.91*
Guilt x Probability of Conviction x Offer Dummy 3	.07	.02	.18	2.91*

Note. Plea Decision: 0=Plea, 1=Trial; Guilt: 0=Innocent, 1=Guilty; Probability 30%, 40%, 50%, 60%, 70%, 80%; Offer Dummy 1: SentenceLength9v6 = 0, SentenceLength3v2 = 1, SentenceLength+Type = 0, SentenceType=0; Offer Dummy 2: SentenceLength9v6 = 0, SentenceLength3v2 = 0, SentenceLength+Type = 1, SentenceType=0; Offer Dummy 3: SentenceLength9v6 = 0, SentenceLength3v2 = 0, SentenceLength+Type = 0, SentenceType=1.
Regression model summary: R = .541, R^2 = .293, Adjusted R^2 = .290, *SE* = .40. *p<.05.

FIGURES

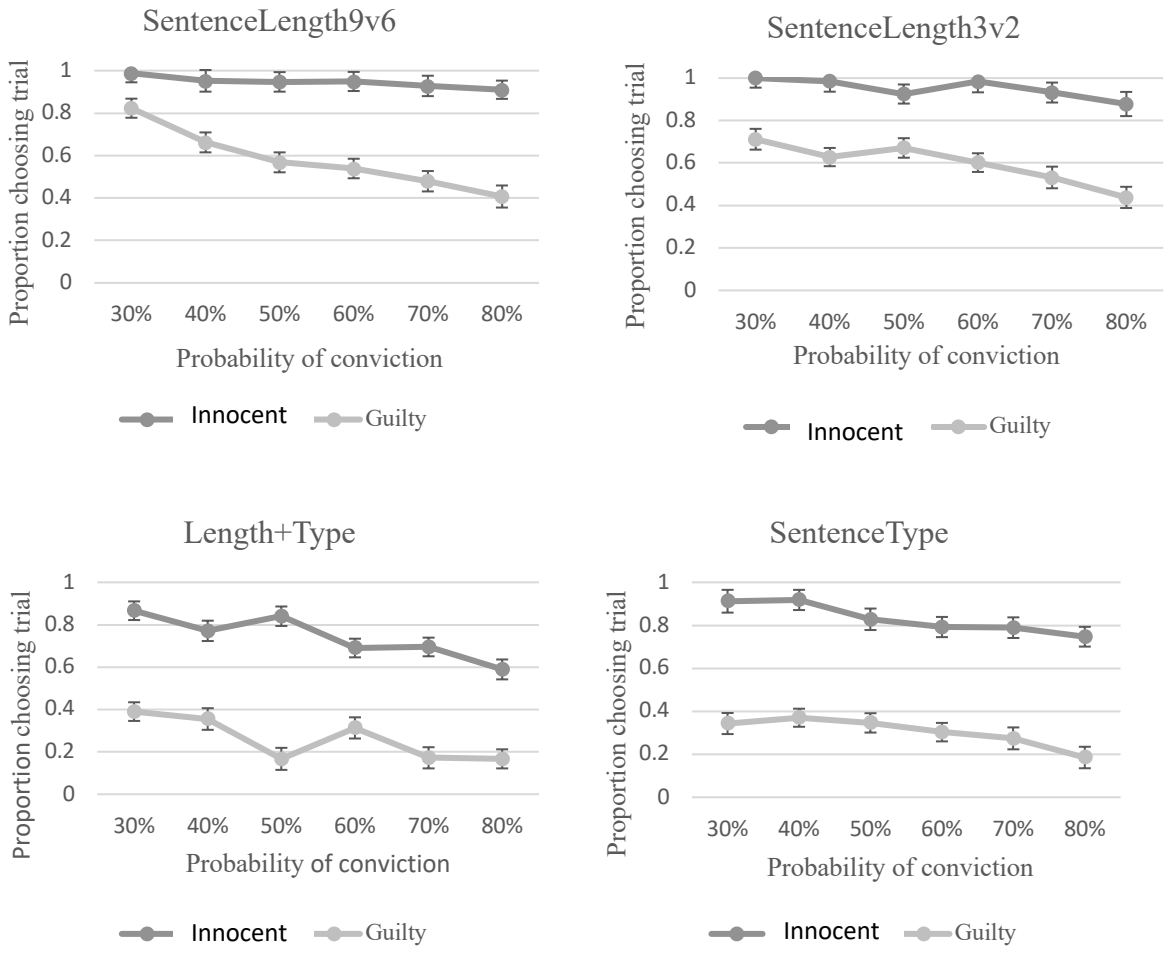

Figure 1: Interactions among probability of conviction, guilt, and discount types. Error bars +/- 1 standard error.

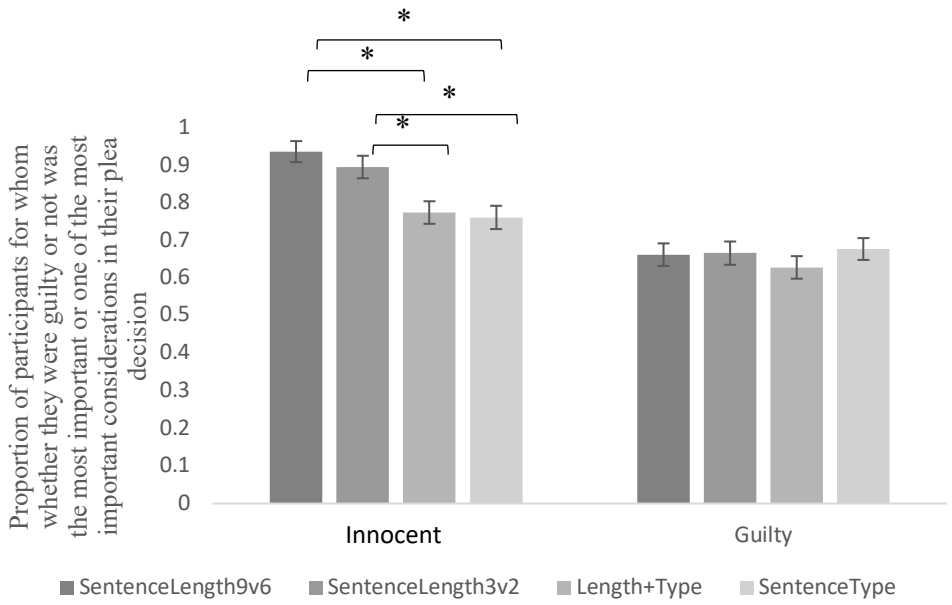

Figure 2: Significant interaction between guilt and discount type in predicting whether being guilty or not was the most important or one of the most important considerations in the participant's plea decision. Error bars +/- 1 standard error. *p<.05.

Supplemental Materials

Predictor variables in the ANOVA were guilt (innocent, guilty), probability of conviction (30%, 40%, 50%, 60%, 70%, 80%), and discount type (SentenceLength9v6, SentenceLength3v2, SentenceLength+Type, SentenceType). The dependent variable was decision to plead guilty (0) or go to trial (1).

The ANOVA revealed a significant main effect of guilt ($F(1,3327) = 974.324$, $p<.001$, $\eta_p^2 = .227$) such that participants in the guilty condition pled guilty more often than participants in the innocent condition ($M_{guilty} = .435$, $SE = .010$, $M_{innocent} = .870$, $SE = .010$). The ANOVA revealed a significant three-way interaction among guilt, probability, and offer type ($F(15,3327) = 1.726$, $p=.040$, $\eta_p^2 = .008$). To summarise, in the sentence length only conditions, only guilty participants pled guilty more at higher levels of probability of conviction (with innocent participants choosing trial regardless of the probability of conviction), but in the sentence type conditions, innocent participants and guilty participants showed a similar trend of pleading guilty more as probability of conviction increased. Detailed results are presented below.

Follow-up comparisons indicate significant main effects of probability for guilty participants in all sentence discount conditions, except for the SentenceType discount condition in which this effect just missed two-tailed significance (SentenceLength9v6 $F(5,3433)=9.642$, $p<.001$, $\eta_p^2 = .014$; SentenceLength3v2 $F(5,3433)=4.046$, $p=.001$, $\eta_p^2 = .006$; SentenceLength+Type $F(5,3433)=4.749$, $p<.001$, $\eta_p^2 = .007$; SentenceType $F(5,3433)=1.999$, $p=.076$, $\eta_p^2 = .003$). In contrast, there was no significant effect of probability for innocent participants in the two sentence length conditions ($p=.871$ and $p = .523$ respectively). As with guilty participants, in the SentenceLength+Type condition there was a significant effect of probability ($F(5,3433)=5.160$, $p<.001$, $\eta_p^2 = .007$) and in the SentenceType condition the effect

just missed two-tailed significance ($F(5,3433)=2.107$, $p=.062$, $\eta_p^2 = .003$). Pairwise comparisons are displayed in Table s1.

Other significant results are listed below for completeness. First, analyses revealed significant main effects of probability ($F(5,3327) = 18.496$, $p<.001$, $\eta_p^2 = .027$), and discount type ($F(3,3327) = 95.240$, $p<.001$, $\eta_p^2 = .079$). Participants pled guilty significantly more often when there was an 80% probability of conviction at trial than in any of the other conditions ($M_{80} = .540$, $SE = .017$; $M_{70} = .600$, $SE = .017$, $p=.014$; $M_{60} = .647$, $SE = .016$, $p <.001$; $M_{50} = .661$, $SE =.017$, $p<.001$; $M_{40} = .705$, $SE = .017$, $p<.001$; $M_{30} = .754$, $SE = .017$, $p<.001$), and participants pled guilty more often when there was a 70% probability of conviction at trial than when there was a 60%, 50%, 40%, or 30% probability of conviction ($p=.049$, $p = .010$, $p <.001$, $p<.001$ respectively). When there was a 60% probability of conviction participants pled guilty significantly more often than when there was a 40% or 30% probability of conviction ($p=.012$, $p<.001$ respectively) but there was no significant difference in plea rates at the 60% and 50% probabilities of conviction ($p=.525$). In the 50% condition, participants pled guilty significantly more often than in the 30% condition ($p<.001$) but plea rates did not differ from the 40% condition ($p=.065$). Plea rates in the 40% condition were significantly higher than those in the 30% condition ($p=.036$).

The significant main effect of discount type showed that participants pled guilty least often when they were offered a sentence length discount only (in the SentenceLength9v6 and SentenceLength3v2 conditions, $M = .764$, $SE = .014$ and $M = .777$, $SE = .014$ respectively). These two conditions were not significantly different from each other ($p=.527$) but showed significantly lower plea rates than the SentenceLength+Type condition ($M = .505$, $SE = .014$, ps $<.001$) and the SentenceType condition ($M = .571$, $SE = .014$, ps $<.001$). Participants also pled guilty significantly more often in the SentenceLength+Type than in the SentenceType condition ($p=.001$).

Results also revealed a significant interaction between discount and guilt ($F(3,327) = 9.383$, $p<.001$, $\eta_p^2 = .008$). Participants in both guilt conditions pled guilty roughly equally often in the two SentenceLength conditions (Guilty: $M_{\text{SentenceLength9v6}} = .580$, $SE = .019$, $M_{\text{SentenceLength3v2}} = .597$, $SE = .019$, $p = .520$; Innocent: $M_{\text{SentenceLength9v6}} = .946$, $SE = .019$, $M_{\text{SentenceLength3v2}} = .950$, $SE = .020$, $p = .869$). Also in both conditions, plea rates were significantly higher in the SentenceLength conditions than in the SentenceLength+Type (Guilty: $M_{length+type} = .260$, $SE = .020$, $ps<.001$; Innocent $M_{length+type} = .742$, $SE = .019$, $ps<.001$) and SentenceType conditions (Guilty: $M_{sentencetype} = .304$, $SE = .019$, $ps<.001$; Innocent $M_{sentencetype} = .831$, $SE = .020$, $ps<.001$). In the innocent condition only, participants pled guilty more often in the SentenceType condition than in the SentenceLength+Type condition ($p=.001$). This difference was not significant in the guilty condition ($p=.119$).

Table S1: *Pairwise comparisons of rates of going to trial by probability for guilty and innocent defendants in each offer condition.*

Offer Condition	Guilt	Probability with significantly higher trial rate / probabilities with significantly lower trial rate	*p*
Length9v6	Guilty	30 / 40, 50, 60, 70, 80	.013, <.001, <.001, <.001, <.001
		40 / 70, 80	.006, <.001
		50 / 80	.022
	Innocent	No significant differences	
Length3v2	Guilty	30 / 70, 80	.011, <.001.
		40 / 80	.004
		50 / 70, 80	.043, .001.
		60 / 80	.013
	Innocent	No significant differences	
Length+Type	Guilty	30 / 50, 70, 80	.001, .001, <.001
		40 / 50, 80	.010, .006
		40 / 70	.011
		50 / 60	.043
		60 / 70	.043
		60 / 80	.047
	Innocent	30 / 60, 70, 80	.005, .006, <.001
		40 / 80	.007
		50 / 60, 70, 80	.019, .024, <.001
Type	Guilty	30/ 80	.023
		40 / 80	.009
		50 / 80	.017
	Innocent	30 / 80	.019
		40 / 80	.009