# Alternative punishments:

# How laypeople and judges impose alternative non-carceral sanctions

Christian Mott[1] & Larisa Heiphetz Solomon[2]

[1] Department of Psychology, Columbia University. 1190 Amsterdam Ave., New York, NY 10027, United States. E-mail: christian.mott@columbia.edu. Phone: 212-853-1406. (Corresponding author)

[2] Department of Psychology, Columbia University. 1190 Amsterdam Ave., New York, NY 10027, United States. E-mail: lah2201@columbia.edu. Phone: 212-854-1348.

**In Press at Psychology, Public Policy, and Law**

Author Note: This publication was made possible through the support of a grant from the Oscar M. Ruebhausen Fund at Yale Law School to the first author and NSF CAREER grant #2141055 to the second author. Any opinions expressed are those of the authors alone and should not be construed as representing the opinions of any organizations that provided support for this project. Thanks to Fiona Doherty, Sarah Russell, and Niall Bolger for valuable feedback.

A partial analysis of the data prepared in Study 1 previously appeared in an unpublished manuscript posted on the first author's website. This manuscript was a shortened form of a paper that co-won Yale Law School's Benjamin Scharps Prize for best paper by a third-year student.

All data and analysis code for Study 1 and all analysis code for Study 2 are available at: https://osf.io/q6npj/?view_only=617dbafda76b49b4b49d6e66fd361922. Data for Study 2 are available at: https://www.ussc.gov/research/datafiles/commission-datafiles.

**Abstract**

Legal theorists have argued that incarceration and alternative sanctions are incommensurable – that is, beyond some crime severity threshold, replacing incarceration with alternative sanctions can never yield a sentence that people will view as appropriate (Kahan, 1996). To test whether laypeople hold this view, we elicited lay judgments about appropriate sentences for four common types of federal crimes in two different conditions: One in which participants could impose only a term of imprisonment and another in which they could impose imprisonment along with alternative sanctions. Laypeople imposed significantly less imprisonment in the latter condition and significant quantities of alternative, non-carceral sanctions. Consistent with the view that imprisonment is commensurable with other sanctions, and particularly with restraint-based sanctions, laypeople substituted supervised release almost one-for-one for imprisonment. In addition, they increased imprisonment and supervised release at similar rates as crime severity increased. Next, using individual-level sentencing data from similar cases in the federal courts, we found that judges' sentencing decisions showed similar relationships between crime severity and both imprisonment and supervised release. However, laypeople imposed dramatically larger fines and more hours of community service than did federal judges, and laypeople tied the use of these alternative sanctions more directly to crime severity. These findings suggest that federal judges do not view fines and community service as commensurable with incarceration. As a result, current criminal sentencing practices deviate from community views by placing excessive emphasis on incarceration and paying insufficient attention to alternative sanctions.

*Keywords*: punishment; psychology and law; morality; social cognition; incarceration.

# Alternative punishments: How laypeople and judges impose alternative non-carceral sanctions

Public discourse about punishment focuses on one type of punishment above others: Incarceration. When a defendant receives a sentence in a high-profile case, headlines often announce the prison sentence rather than any other component of the punishment. For example, after the actress Felicity Huffman received a sentence of 14 days in prison, 1 year of supervised release, 250 hours of community service, and a $30,000 fine, the *New York Times* headline announced, "By Turns Tearful and Stoic, Felicity Huffman Gets 14-Day Prison Sentence" (Taylor, 2019). The remaining components of the sentence did not appear until halfway through the article. Similarly, after celebrity attorney Michael Avenatti received a sentence of 4 years in prison, 3 years of supervised release, $148,750 of restitution, and $297,000 of forfeiture, the headline read, "Avenatti Gets a 4-Year Sentence for Defrauding Stormy Daniels of $300,000" (Moynihan, 2022). The body of this article did not mention the supervised release portion of the sentence at all.

The emphasis on carceral sentences when talking and thinking about punishment reflects a "widespread view [in the United States] that only imprisonment counts" (Tonry, 1996, p. 128). This view appears in domains beyond newspaper headlines, including scientific research. In psychology, researchers often assess laypeople's punishment judgments by asking them only or primarily about how long a person who committed a crime should be incarcerated (Cushman, 2008; Ginther et al., 2016; Martin & Heiphetz, 2021; Robinson et al., 2012; Robinson & Darley, 1995; Shariff et al., 2014). In economics, researchers studying the effects of various independent factors on sentences often use the amount of incarceration imposed as the primary dependent

variable reflecting punishment severity (Cohen & Yang, 2018; McConnell & Rasul, 2018; Yang, 2015).

This focus on incarceration also appears in sentencing procedures. Federal judges must engage in lengthy calculations to determine a guidelines range for the *prison* sentence in each case (USSC, 2021). Prior to the United States Supreme Court opinion in *United States v. Booker* (2005), these sentencing ranges were mandatory; since that decision, they have been advisory. Nevertheless, courts must still calculate the guidelines range in every case, any error in the guidelines calculations is still grounds for vacating a sentence (*Molina-Martinez v. United States*, 2016), and most judges sentence within or near the recommended range (Hofer, 2019). These guidelines calculations have two components: (1) determining the offense level of the specific crime(s) before the court; and (2) placing the defendant into one of six criminal history categories. To determine the offense level, the judge must follow the guideline applicable to each statute of conviction, finding facts as necessary to increase or decrease the offense level based on specific offense characteristics (USSC, ch. 2), and then the judge must group any related charges and make adjustments based on certain victim characteristics or defendant motivations, the defendant's role in the offense, and the defendant's responses to the investigation and charges (USSC, 2021, ch. 3). To determine the defendant's criminal history category, the judge must take into account the number and seriousness of the defendant's prior convictions in any United States jurisdiction (USSC, ch. 4). Finally, the judge uses these two numbers – offense level and criminal history category – to look up the defendant's recommended prison sentence on a chart in the guidelines manual (USSC, ch. 5).

By contrast, the procedures federal courts use to determine the amounts of alternative, non-prison sanctions – e.g., probation, supervised release, fines, and community service – are

much less detailed. For probation and supervised release, the availability of the sanction and its maximum length depend on the classification of the offense into one of three or four broad categories based on the statutory maximum sentence (18 U.S.C. §§ 3559(a), 3561(a), (c), 3583(b)). After this straightforward categorization task, the statutes and guidelines provide little specific guidance about what amounts of these sentences to impose, leaving them largely to the judge's discretion (18 U.S.C. §§ 3562(a); 3583(c); USSC, 2021, §§ 5B1.2, 5D1.2). For fines, statutory law and guidelines only provide broad ranges – for example, the recommended fine range based on offense level covers an entire order of magnitude (e.g., from $10,000 to $100,000 for crimes with offense levels 18 and 19; 18 U.S.C. §§ 3571(a), 3572(b), USSC, 2021, §5E1.2). And for community service, no guidance at all is provided, aside from an application note pointing out that it is burdensome on the probation office to impose community service sentences of more than 400 hours (USSC, 2021, § 5F1.3 & n. 1).

In each of these contexts – newspapers, scientific research, sentencing law and policy – people treat incarceration as a special and *primary* form of punishment. By contrast, alternative sanctions like supervision, fines, and community service appear secondary – less important to report, measure, or precisely calibrate to a defendant's culpability. This division between incarceration and alternative sanctions tracks a similar distinction legal theorists have drawn between different forms of punishment, based on the degree to which they can serve the purposes of punishment – that is, the degree to which alternative sanctions like supervision, fines, and community service can replace all or a portion of a carceral sentence and still achieve the same amount of retribution, deterrence, incapacitation, rehabilitation, condemnation, or some other proposed purpose. On some accounts, incarceration and alternative sanctions are *incommensurable* – that is, once a crime exceeds some relatively low level of severity, no

quantity of alternative sanctions can serve the same purpose as incarceration (Husak, 2020; Kahan, 1996). On other accounts, incarceration and alternative sanctions are fully commensurable: at any severity level, a punisher can craft an appropriate punishment without using incarceration (von Hirsch, 2017).

Prior research has investigated laypeople's views about the commensurability of individual alternative sanctions as complete substitutes for incarceration (Harlow et al., 1995; McFatter, 1982; Petersilia & Deschenes, 1994; Sebba, 1978). However, no prior research has investigated views on the commensurability of alternative sanctions as partial substitutes for incarceration – e.g., replacing five years of a ten-year prison sentence with supervision and a fine – and how partial substitution relates to crime severity. This latter question, however, is of immense practical importance. Every judge crafting a sentence in a criminal case must try to find a balance of incarceration and alternative sanctions that, after factoring in the seriousness of the crime, yields a sentence "sufficient, but not greater than necessary" to accomplish the goals of punishment (18 U.S.C. § 3553(a)). This balancing requires judges to consider whether some alternative sanction could replace part of any prison sentence under consideration.

This paper reports two studies that begin to fill this hole in the literature. In Study 1, we investigated how the availability of alternative sanctions affected the punishments laypeople imposed in response to vignettes describing four common federal crimes. If laypeople viewed the alternative sanctions as incommensurable with incarceration, this study would show no difference in prison sentences between conditions. Further, we investigated the relationship between perceived crime seriousness and the quantity of each type of sanction imposed, which provided insight into whether people were imposing alternative sanctions to accomplish the same purposes as imprisonment. Again, if laypeople viewed alternative sanctions as incommensurable

with incarceration, they would use those sanctions, if at all, to accomplish a punitive purpose less tightly correlated to crime severity than incarceration. In Study 2, we investigated how federal judges imposed sentences in real cases with similar facts. These data allowed an analysis of how federal judges used various forms of sanctions to respond to crimes of increasing severity, as well as a comparison of these responses to laypeople's responses. These analyses provided insight into whether the federal judges had different views than laypeople about the commensurability of incarceration and alternative sanctions.

**Punishment Commensurability in Legal Theory**

Some theories of punishment place few restraints on the particular form punishment must take. For example, standard accounts of retributivism require that the amount of hard treatment inflicted as punishment be proportional to the wrongfulness of the legal violation but do not require a particular type of hard treatment (Brooks, 2012; Husak, 2020; Robinson & Spellman, 2005). On these accounts, punishment can take any form as long as it inflicts the right amount of harm on the defendant (von Hirsch, 2017). Likewise, deterrence theory treats a punishment as justified to the extent the benefit of deterring the defendant or the public from committing future crimes outweighs the harm of the punishment itself (Brooks, 2012). Again, punishment can take any form on a deterrence theory as long as it deters; indeed, the form of punishment that most effectively deters while inflicting the *least* harm on the defendant is preferable. The theory that punishment serves a rehabilitative purpose (Brooks, 2012) is also agnostic to the precise form punishment takes. If a sanction helps bring about a change in the defendant that reduces their risk of reoffending, then it potentially has a place in rehabilitative punishment (Lipsey & Cullen, 2007).

On other theories of punishment, however, incarceration is primary, either as the only form of punishment that accomplishes its main purposes or the only one that can do so for crimes of widely varying seriousness. The theory of incapacitation – the view that punishment is justified because it prevents criminal defendants from committing additional crimes against the majority of the population during the term of their sentence (Cohen, 2013; Wilson, 2013; Zimring & Hawkins, 1995) –justifies punishments only if they restrain defendants' freedom of movement, like incarceration and (to some degree) supervised release. Similarly, some forms of expressivism – the view that punishment acts as a form of communication[1] – also give priority to carceral punishment (Brooks, 2012). Joel Feinberg, for example, distinguishes true punishment (primarily incarceration) from mere penalties (such as monetary fines) and argues that only the former expresses the "resentment," "indignation," "disapproval," and "reprobation" appropriate for felonies (Feinberg, 1965, p. 400). This theory places particular emphasis on imprisonment as a way of expressing these attitudes: in Feinberg's memorable description, "the very walls of his cell condemn" a defendant punished by incarceration (Feinberg, 1965, p. 402). A different form of expressivism claims that the imposition and service of punishment form a dialectic between the state and the defendant involving the expression of censure by the former and the acceptance of responsibility by the latter (Duff, 2001). On this view, appropriate punishment must inflict a

[1] The account of the precise message punishment communicates differs by theorist. On some views, it expresses condemnation of the defendant's law-violating behavior (Brooks, 2012; Primoratz, 1989). This condemnation may elicit shame (Husak, 2020; Kahan, 1996) or penance (Duff, 2001) from the defendant or educate the defendant and/or other members of society about the norms of a society or the costs of law-violating behavior (Hampton, 1984; Nahmias & Ahroni, 2017). On other accounts, punishment expresses a rejection of the defendant's attempt to claim some form of superiority over the victim (Hampton, 1992). And on still other accounts, punishment expresses society's hatred of the defendant (Stephen, 2014/1883). Adjudicating between these different forms of expressivism is outside the scope of this paper, though the degree to which imprisonment is the primary punishment justified by expressivism may vary based on the precise message punishment sends.

lengthy period of hard treatment so the defendant can reflect on their behavior and undergo a "secular penance" (Duff, 2001, p. 129). Other theorists take a more moderate view, emphasizing that incarceration has a special power to express censure while acknowledging that other forms of punishment may express some amount of censure in certain circumstances, at least when paired with incarceration (Husak, 2020; Kahan 1996, 1998).

These different views on the commensurability of incarceration and alternative sanctions reflect different accounts of the *meaning* of alternative sanctions. On a retributivist or deterrence account, alternative sanctions are just another way of harming defendants in order to give them their "just deserts" or increase the cost of their behavior (Brooks, 2012; Robinson & Spellman, 2005). On a rehabilitative account, alternative sanctions are another set of potential tools for reforming defendants (Lipsey & Cullen, 2007). On expressivist accounts, however, incarceration and alternative sanctions send fundamentally different messages, and those messages may or may not be appropriate given the nature of the defendant's behavior. Expressivists generally agree that incarceration sends a message that the defendants' behavior and/or the defendants themselves are unacceptable to the public, though they differ on what message incarceration sends about the reason for that unacceptability (Brooks, 2012; Duff, 2001; Kahan, 1998; Hampton, 1984, 1992; Nahmias & Ahroni, 2017; Stephen, 2014/1883). By contrast, expressivists have argued that monetary sanctions can look like prices, thereby sending the message that the harm punished is just a "cost of doing business" or that the wealthy can buy their way out of prison (Kahan, 1998, p. 698). Likewise, certain forms of community service look like the praiseworthy behavior of upstanding citizens. On an expressivist view, treating this behavior as a punishment may risk either elevating the defendant – the opposite of the intended effect – or diminishing people who undertake that service voluntarily (Kahan, 1998).

Different theories of punishment therefore disagree about the extent to which incarceration is *commensurable* with other forms of punishment (Kahan, 1998). Following Kahan, we can say that two types of punishment are commensurable when, at every level of crime seriousness, there is some amount of each punishment that could appropriately punish someone who committed a crime of that seriousness. Two types of punishment are *incommensurable* when this relationship does not hold. We can also say that two types of punishment are *partially commensurable* when an appropriate punishment consisting of one or both punishments remains appropriate after substituting some amount of one punishment for the other. The theories of retribution and deterrence imply that alternative punishments may be commensurable or partially commensurable with incarceration, to the extent they inflict the same amount of harm or prevent harmful future behavior, respectively (Brooks, 2012). Incapacitation theory implies that the only alternative punishments potentially commensurable or partially commensurable with incarceration are supervision, home confinement, and other similar restraint-based sanctions. Expressivism – at least in the forms discussed above –implies that alternative punishments are generally incommensurable with incarceration (Duff, 2001; Feinberg, 1965), at least for more serious crimes (Husak, 2020; Kahan, 1996).

**The Psychology of Alternative Sanctions**

This normative question about punishment commensurability raises a related descriptive question: To what extent do laypeople think that incarceration and alternative punishments are commensurable? One prominent position is that laypeople generally reject the replacement of incarceration with alternative punishments, particularly for serious crimes, because they think incarceration is incommensurable with standard alternative punishments like supervision, fines, and community service (Husak, 2020; Kahan, 1996). Another position is that incarceration and

alternative punishments are commensurable to a substantial degree, though there may be non-linearities that make it somewhat harder to replace imprisonment with alternative sanctions as crime seriousness increases (Robinson & Spellman, 2005).

At a high level, there are two descriptive questions one can ask about punishment commensurability. One is about how laypeople view the *complete* substitution of imprisonment with another type of punishment – that is, the whether the two types of punishment are fully commensurable. Existing research has investigated this question by eliciting judgments about the relative severity of different prison sentences and various quantities of specific alternative sanctions (Harlow et al., 1995; Petersilia & Deschenes, 1994; Sebba, 1978). These studies show non-linearities in the relative severities of different types of punishment. For example, Sebba's (1978) participants thought that a $250 fine was a harsher punishment than 1 month of imprisonment but did not think that a $50,000 fine ($250 x 200) was a harsher punishment than 16 years, 8 months (1 month x 200) of imprisonment; indeed, they believed the opposite. That study also provided evidence that people view even reasonably severe punishments as commensurable with alternative sanctions, with one sample of participants ranking a $50,000 fine as roughly equal in severity to a 10-year prison sentence. A follow-up study found similar results (Nathan & Sebba, 1984). Although more recent research has not found similar patterns (Harlow et al., 1995; Petersilia & Deschenes, 1994), these studies tested much smaller fines and quantities of imprisonment, making direct comparison impossible.

Another way of exploring laypeople's willingness to substitute alternative sanctions for imprisonment is to ask them about the appropriate sentence for a crime, while varying whether that sentence consists of incarceration or alternative sanctions. In a study using this approach, McFatter (1982) presented participants with a short description of the facts of a crime and asked

them to rate the appropriateness of each punishment on a scale containing varying amounts of pure fines, pure probation, and pure incarceration. The lay participants judged certain quantities of fines and probation as more appropriate sentences than small sentences of imprisonment for both minor (assault, car theft) and major (rape, murder) crimes. These results suggest that participants would choose to impose certain sentences of fines or probation over low levels of imprisonment. That is, laypeople at least appear to believe that certain amounts of fines and probation are preferable to a suboptimal prison sentence. Although this finding does not show that participants thought the punishments were commensurable, it does show that they did not view fines or probation as categorically incapable of punishing to the same degree as any amount of incarceration.

A final way to investigate laypeople's willingness to make complete substitutions of alternative sanctions for incarceration is to present participants with the facts of a criminal case, ask them what sentence they would impose on the defendant, and present them with a response scale with options ranging from alternative sanctions (e.g., fines, probation) to terms of incarceration. Several studies using this type of response measure have shown a general preference for carceral sentences in response to more serious crimes (Blumstein & Cohen, 1980; Rossi et al., 1997; Rossi & Berk, 1997; Silver, 2017; Silver & Berryessa, 2021; Stalans & Diamond, 1990).[2] However, these studies have generally elicited these punishment judgments in order to explore questions about punishment other than the commensurability of incarceration

[2] Another substantial line of research has elicited punishment judgments using a one-dimensional scale that either includes only terms of imprisonment or only a broad "probation" option at the lower end of the scale (Alter et al., 2007; Bailis et al., 1995; Carlsmith et al., 2002; Cushman, 2008; Darley et al., 2000; Ginther et al., 2016; Greene & Darley, 1998; Robinson & Darley, 1995; Robinson et al., 2012; Robinson & Darley, 1995; Shariff et al., 2014; Slobogin & Brinkley-Rubinstein, 2013). This type of response variable is not well-suited to assessing laypeople's willingness to substitute alternative sanctions for incarceration.

and alternative sanctions, such as the degree of alignment between lay sentences and actual or guidelines sentences (Blumstein & Cohen, 1980; Rossi et al., 1997; Rossi & Berk, 1997), the relationship between morality and different punitive orientations (Silver et al., 2017; Silver & Berryessa, 2021), and perceptions of judicial leniency (Stalans & Diamond, 1990). These studies therefore have generally treated imprisonment as intrinsically more severe than alternative sanctions (i.e.., by placing it above alternative sanctions on a one-dimensional response scale) and have not provided participants with alternative sanction options that are extremely severe (i.e., lasting more than 1 or 2 years). Thus, it is difficult to draw any conclusions about perceptions of punishment commensurability from their findings.

The second question about punishment commensurability concerns how laypeople view the *partial* substitution of imprisonment with alternative sanction(s) – e.g., replacing five years of a ten-year prison sentence with supervision and a fine. This is a question about whether people view imprisonment and alternative sanctions as partially commensurable. Much less theoretical or empirical work has investigated this question. On the theoretical side, Kahan (1996) advances arguments that cut both ways, arguing, on the one hand, that fines or community service are completely unacceptable substitutes for any amount imprisonment – because fines create the appearance that the defendant is buying their way out of some jail time and community service wrongly elevates the defendant while diminishing those who engage in service voluntarily – and, on the other hand, that adding any amount of imprisonment to an alternative sanctions sentence completely resolves any concern about commensurability. The United States Sentencing Guidelines reflect this view, to some extent, by prioritizing incarceration and placing little emphasis on fines and community service.

On the empirical side, existing evidence suggests that laypeople accept partial substitution of alternative sanctions for imprisonment for less serious crimes. McFatter's (1982) study provides evidence that laypeople think fines and probation *can* constitute at least moderately appropriate punishments, particularly for the two less serious crimes tested. Moreover, Sebba and Nathan (1984) showed that some populations think alternative sanctions can replace a portion of incarceration's severity, at least in the abstract and for relatively small aggregate sentences. In small samples of police officers and current prisoners, a $50,000 fine was viewed as *more* severe than 5 years' incarceration. In small samples of current prisoners and college students, adding a $5,000 fine to 18 months' incarceration made it more severe than 5 years' incarceration. However, this study only allowed participants to rate a limited set of punishment combinations, involving relatively small quantities of both punishment types, so it is unclear whether this finding extends to other combinations, particularly in light of the observed non-linearities in punishment severity.

These findings put pressure on Kahan's (1998) hypothesis that fines and community service have fundamentally unacceptable social meanings as punishments for serious crimes. However, McFatter's results, as well as the results of more recent research in moral psychology (Heffner & FeldmanHall, 2019), do suggest that people think alternative responses to crime, like monetary payments, become *less* appropriate as crime severity increases. One potential explanation for this severity-sensitivity is that people acting as third-party punishers think that responding more punitively to more serious crimes bolsters their moral reputation (Heffner & FeldmanHall, 2019), perhaps because punishers want to make clear to observers that they personally reject the defendant's behavior or signal their moral outrage and trustworthiness (Brown et al. 2022). Taken together, these results imply that laypeople will seek to maintain a

balance between different types of sanctions, imposing sufficient incarceration to signal clearly the unacceptability of the defendant's conduct (and their own rejection of that conduct) but also using fines and community service as partial substitutes, though to a decreasing extent as crime severity increases.

Another study that sheds light on the way laypeople think about the partial substitution of alternative sanctions for incarceration investigated the impact of *non-punitive* suffering on incarceration judgments (Robinson et al., 2012). In this work, participants read about people who had committed a crime and assigned a carceral term to each person. Then, they learned that the defendant had suffered a neck injury causing quadriplegia – that is, the defendant had suffered, but not from a punishment. Subsequently, between 45% (for the least serious offense) and 13% (for the most serious offense) of participants reduced the carceral sentence they had assigned. Additionally, the magnitude of these reductions increased as offense seriousness increased, even though the defendant's injury was the same across conditions. If participants think of alternative sanctions the same way they think as this other type of suffering, then these findings suggest that imposing alternative sanctions may lead to reduced incarceration judgments. That is, this study suggests that laypeople view incarceration and alternative sanctions as partially commensurable, even for serious crimes.

Prior research on punishment judgments bears on laypeople's views on commensurability in another way: The theories of punishment most associated with (at least) partial commensurability are retributivism and deterrence, while the theories most associated with incommensurability are incapacitation and expressivism. Prior research has found substantial evidence that ordinary people primarily use punishment to accomplish retributivist goals (Carlsmith, 2008; Carlsmith et al., 2002; Darley et al., 2000; McFatter, 1982; Nadelhoffer et al.,

2013; Warr et al., 1983) and only secondarily use punishment to accomplish the goals of expressivism (Cushman et al., 2019; Dunlea & Heiphetz, 2021; Funk et al., 2014; Nahmias & Aharoni, 2017) or incapacitation (Slobogin & Brinkley-Rubinstein, 2013). If laypeople apply these theories in the same way as legal theorists, then these two implicit goals of punishment should push in opposite directions – retributivism towards commensurability; expressivism and incapacitation towards incommensurability – with the former predominating, to some degree, over the latter.

**Federal Judges' Views on Alternative Sanctions**

The policy implications of laypeople's views on commensurability depend, in part, on the extent to which they match judges' current sentencing practices. If they do, then that may provide some grounds for maintaining those practices, as reflecting community standards (Gwin, 2010; Robinson & Darley, 1995). For the same reason, if judge's current sentencing practices deviate from laypeople's views, then that provides grounds for questioning them.

There are reasons to think that judges and laypeople will have similar views on commensurability. Some prior research has shown a reasonable fit on a variety of crimes between lay punishment judgments and the guidelines ranges (Rossi et al., 1997; Rossi & Berk, 1997; for a review, see Cullen et al., 2000). Even though judges post-*Booker* can sentence below the guidelines range, their sentences often remain close to the guidelines minimum (Gwin, 2010; Hofer, 2019). If judges and laypeople agree about how much punishment is appropriate, then it is plausible they also agree about what forms of punishment are appropriate.

However, other research suggests that laypeople may be more lenient than judges. A famous study of real judges and juries in thousands of cases found that the juries were less likely to convict than the judges would have been (Kalven & Ziesel, 1966). Although researchers

disagree on the mechanism underlying this observed leniency (Devine et al., 2004; MacCoun & Kerr, 1988), this finding seems to extend to sentencing judgments. In Judge Gwin's (2010) study, 261 real jurors in 22 criminal cases responded individually to a survey question about how many months of imprisonment would be appropriate for the defendant. In cases spanning drug possession and distribution, possession of a firearm by a prohibited person, child pornography, mail fraud, and perjury, jurors' average sentences (65 months) were roughly half the average sentences the judges imposed (125 months) and less than half the average guidelines minimum (138 months). Judges using jury polls to inform their sentencing judgments have obtained mixed results. Jurors who had convicted a defendant of distributing child pornography recommended an average of 14.5 months' incarceration, while the guidelines minimum was 262 months (*United States v. Collins*, 2016); jurors who had convicted a defendant of conspiracy to possess heroin with intent to distribute recommended an average of 19.4 years' incarceration, while the guidelines maximum was 10 years (*United States v. Obiora*, 2018). Nevertheless, this evidence taken together suggests that laypeople may be more lenient than both the guidelines and federal judges in their sentencing judgments. If so, laypeople may prefer forms of punishment that they perceive as less severe and therefore prefer alternative sanctions to incarceration.

There are also reasons to think that judges may rely on incarceration to an even greater degree than laypeople when crafting their sentencing packages, even if judges' and laypeople's punishment judgments are similar. First, the studies described above elicited laypeople's judgments on a one-dimensional scale consisting mostly of different terms of imprisonment (Gwin, 2010; Rossi et al., 1997; Rossi & Berk, 1997). This way of measuring punishment judgments smooths over any differences in the way judges and laypeople allocate punishment across modalities when given the opportunity. Second, judges must follow the guidelines, and

the guidelines manual places dramatically greater emphasis on carceral sanctions than on alternative sanctions, as described above.

Another reason to think judges and laypeople may differ in their sentencing judgments is that judges crafting sentences may take into account considerations that do not occur to laypeople. A prominent account of how judges make sentencing decision is the *focal concerns theory*, which posits that the primary concerns driving judges' sentencing decisions are "the offender's blameworthiness, protection of the community, and practical constraints and consequences" (Crow & Bales, 2006, p. 287). These concerns highlight different purposes and consequences of punishment. The first, blameworthiness, addresses the severity of the crime. This concern is likely to lead to punishments aimed at retribution and condemnation. The second, protection of the community, addresses the need to prevent the defendant from committing future crimes. This concern is likely to lead to punishments aimed at incapacitation and deterrence. The third, practical constraints and consequences, incorporates factors involved in the actual administration of a sentence, including the costs of incarceration, the costs of supervision, the defendant's ability to pay a fine, the availability of community service opportunities, and so forth.

Prior research provides reason to think that laypeople and judges may weight these concerns differently. Laypeople making punishment judgments focus primarily on the first of these concerns, retribution and condemnation, and give little or no weight to the second (e.g., Carlsmith et al., 2000; McFatter, 1982). The third concern is also likely to be less salient to experimental participants involved in a theoretical sentencing exercise than federal judges handing down a real punishment. Judges interact with the probation office, which gives them a greater appreciation of their resource constraints. Moreover, judges may see defendants who

violate terms of post-prison sanctions, like paying fines or performing community service, in court again repeatedly and conclude that those forms of punishment interfere with defendants' chance for a fresh start (Barkow, 2019).

**Overview of the Present Research**

We conducted two studies investigating how people incorporate alternative punishments into their sentencing judgments and how these alternative punishments scale with crime severity. In Study 1, we investigated how the availability of alternative sanctions affected the prison sentences laypeople imposed in response to vignettes describing four common federal crimes. If the alternative sanctions were incommensurable with incarceration, we would have expected no difference between conditions. Instead, we found a significant decrease in prison sentences across crimes of varying seriousness, suggesting participants viewed alternative sanctions as at least partially commensurable with imprisonment at all tested levels of seriousness. Participants appeared to think imprisonment and supervision accomplished similar goals, as they increased those two forms of at similar rates as crime seriousness increased. In Study 2, we investigated how federal judges imposed sentences in real cases with similar facts. If judges took cues from the emphasis placed on different punishment types in the sentencing guidelines manual, we would expect them to make minimal use of alternative sanctions, treating imprisonment as the primary way to calibrate punishment to crime severity. The judges imposed similar prison and supervision sentences but used fines and community service to a much lower degree than the experimental participants and did so in a way uncorrelated with crime severity. In sum, these results showed that laypeople viewed incarceration and alternative sanctions as fully or partially commensurable at all crime seriousness levels, while federal judges primarily relied on the sanctions that could result in incapacitation: Imprisonment and supervised release.

## Study 1

The first study compared laypeople's punishment judgments when they could just impose imprisonment and when they could impose imprisonment or a variety of alternative sanctions – including supervised release, community service, and fines – in an experimental context. If laypeople view alternative sanctions as incommensurable with incarceration, then imprisonment judgments in the two conditions should not differ. If, instead, laypeople prioritize incapacitation and therefore view imprisonment and supervised release as the only fully or partially commensurable punishment types, then their mean judgments should show some degree of substitution between those two restraint-based punishments. Finally, if laypeople view all the punishments as fully or partially commensurable (e.g., on a retributivist theory), then they should be willing to substitute multiple types of alternative punishment for imprisonment, and the amount of each type of punishment should increase as crime severity increases.

### Methods

The first study elicited participants' punishment judgments about one of four different crimes using one of two different sets of punishment types, for eight different between-subjects conditions. After providing their punishment judgments, participants rated the seriousness of all four crimes.

The four vignettes involved common federal criminal violations: drug possession with intent to distribute, wire fraud, bank robbery, and participation in a drug trafficking conspiracy while in possession of a firearm. We chose these offenses because they are federal-law analogues of common state crimes: drug possession, fraud, robbery, and drug trafficking. According to the sentencing guidelines in effect when this study took place, these crimes had offense levels 3, 18,

21, and 27,[3] respectively, which were associated with guidelines sentencing ranges of 0-6 months, 27-33 months, 60-60 months, and 180-180 months, respectively, for an individual in the lowest criminal history category (United States Sentencing Commission (USSC), 2014). The ranges for bank robbery and the drug conspiracy only contain a single value because the top of the guidelines range is lower than the statutory mandatory minimum. The guidelines provide that the bottom of the guidelines range cannot be lower than an applicable mandatory minimum – here, 15 years or 180 months (USSC, 2014).[4] If the guidelines in the bank robbery and drug conspiracy cases are calculated without positing a charge under the law prohibiting the possession of a gun during a crime of violence or drug trafficking offense (18 U.S.C. § 924(c)), then the offense levels would be 26 and 29, respectively, and the sentencing ranges would be 63-78 months and 87-108 months (USSC §§ 2D1.1, 2B3.1, 2K2.4 & n. 2, 2014). Thus, if participants' judgments tracked the offense levels and recommended sentences, we would observe substantial separation between the punishment judgments for these four offenses.

***Participants***

Prior literature suggested an effect size of at least $d = 0.4$ between the imprisonment-only group and the alternative sanctions group (Robinson et al., 2012). Achieving power of 0.8 with $\alpha = 0.05$, given that effect size, required 200 participants. Two hundred fifty-five adult participants completed this study via Amazon's Mechanical Turk.[5] We excluded 18 participants for failing a

[3] The offense levels of 21 for the bank robbery and 27 for the drug trafficking offense do not reflect the mandatory 5-year consecutive sentence required due to the possession of a gun during a crime of violence or drug trafficking offense (18 U.S.C. § 924(c)(1)(A)).
[4] We calculated the offense levels and guidelines ranges based on the version of the United States Sentencing Guidelines Manual in effect at the time we conducted this study.
[5] Samples recruited on Amazon's Mechanical Turk (mTurk) are generally more diverse than in-person samples recruited at universities, which improves the quality of the data (e.g., Buhrmester et al., 2011). Since Summer 2018, however, researchers have detected a decline in quality of mTurk data, either due to an increase in bots or a decrease in the sincerity of responses (Ahler et

comprehension check described below, leaving 237 in the analysis ($M_{age}$ = 32.84 years, $SD_{age}$ = 10.41 years; 42% female, 55% male, 3% did not respond; 73% White; 9% Black, 7% Hispanic, 7% Asian, 1% American Indian or Alaska Native, and 1% who responded that their race or ethnicity was not listed).[6] Of these 237 participants, sixty completed the drug conspiracy vignette, sixty-one completed the drug possession with intent to distribute vignette, fifty-six completed the wire fraud vignette, and sixty completed the bank robbery vignette.[7]

***Procedure***

First, each participant provided informed consent to take part in this study.[8] Then, we randomly assigned each participant to one of the two punishment-type conditions – that is, to impose imprisonment alone as a punishment or five different types of punishment including imprisonment. Participants in the imprisonment-only condition received the following instructions:

> You are going to read a story about an individual who has committed a crime under Federal law and pled guilty. Once you have read this story, you are going to be asked what sentence that individual should receive.
>
> When you are deciding on a sentence, you will be able to order that the individual spend a certain number of years in prison.

al., 2020; Chmielewski & Kucker, 2020), though it remains possible to draw meaningful conclusions from these responses despite this increase in noise (Johnson & Ryan, 2020). Regardless, we conducted this study before this decline in quality and used a comprehension check to screen out participants who did not remember a basic fact about the vignette.

[6] The patterns of results in the full sample, without these exclusions, remain similar to those reported in the main text.

[7] Five of these participants completed the sentencing portion of the survey but declined to answer the questions rating the seriousness of all four vignettes. We therefore excluded these five participants from any analyses that include seriousness scores.

[8] This study was reviewed and deemed exempt by the Yale University IRB.

Participants in the alternative sanctions condition received a different version of the instructions that described the forms of punishment available to participants: Imprisonment, forfeiture, fine, supervised release, and community service:

> You are going to read a story about an individual who has committed a crime under Federal law and pled guilty. Once you have read this story, you are going to be asked what sentence that individual should receive.
>
> When you are deciding on a sentence, you will be able to order a number of different punishments in addition to prison. These are:
>
> - "Forfeiture": Forfeiture of all property involved in commission of the crime, all money earned from commission of the crime, and all property purchased with money earned from commission of the crime.
> - "Fine": Payment of a fine to the government.
> - "Supervised release": A program in which, after being released from prison, an individual registers with a probation officer. The individual is required to meet with this probation officer regularly, receive permission from the probation officer before traveling out of state, submit to regular drug testing, and maintain stable employment.
> - "Community service": Performance of a set number of hours of community service.

Participants received these five options because they are the primary types of punishment available under federal law: The statutes governing federal punishment contain specific sections dedicated to imprisonment, fines, supervised release, and forfeiture (18 U.S.C. §§ 3554, 3556, 3571, 3581, 3583), and the United States Sentencing Guidelines make community service the default alternative to fines and restitution for defendants unable to pay, as well as an available condition of both probation and supervised release (USSC, 2021).

We then randomly assigned each participant to receive one of four vignettes. One example vignette appears below, and the other three appear in the Appendix:

*Drug Conspiracy*

> John is 26 years old. He is part of an organization that imports cocaine into a major American city. His organization imports the drugs through Central America and then uses a large number of street-level dealers to sell the drugs to individuals in the United States. John never sells drugs to individuals himself. Instead, John helps organize both the importation of drugs into the country and the protection of the drugs once they are in the country. He always carries a handgun, although he has never had to use it. John has received his high school diploma and has no criminal record.
>
> One day, after he has been involved in importing cocaine for just over a year, he is arrested while standing in a stash house containing more than 5kg of cocaine. At the time he is arrested, he is carrying a handgun in his pocket. He is charged with and pleads guilty to conspiracy to distribute over 5kg of cocaine and carrying a handgun during the commission of a drug offense.

Then, in the imprisonment-only condition, participants answered the following question:

> Imagine that you are the judge deciding on John's sentence.
>
> Please let us know how many years you think he should be sentenced to spend in prison: ___

Participants entered their responses as numbers in a text-entry field that only accepted numerical responses.

In the alternative sanctions condition, participants answered the following questions in counterbalanced order. Participants entered their responses to all but the forfeiture question in text-entry fields that only accepted numerical responses. Participants used radio buttons to give a "yes" or "no" answer to the forfeiture question.

> Imagine that you are the judge deciding on John's sentence.
>
> Please let us know how many years you think John should be sentenced to spend in prison: ___
>
> Please let us know how many years you think John should be sentenced to spend in supervised release after he is released from prison: ___
>
> Please let us know how many hours of community service John should be required to perform: ___

> Please let us know how large a fine John should be required to pay to the government: ___
>
> Do you think John should be sentenced to forfeiture? [yes/no]

Participants then rated their confidence in their sentences on scale ranging from 1 (Not at all confident) to 4 (In between) to 7 (Completely confident) and answered a comprehension question about a detail from the story that may play a role in their punishment judgments – i.e., a yes-or-no question about whether or not John had a criminal record.

On the next page, participants received the text of all the vignettes, including the vignette for which they had already imposed a punishment, in counterbalanced order. Participants were told to read the stories and use the sliders below each story to assign a score from 0 (Not at all serious) to 100 (The most serious) that reflects that seriousness of the offense described in the story.

Finally, participants answered optional demographic questions. In addition to standard questions about age, gender, and race or ethnicity, participants responded to a binary measure asking whether or not they were lawyers, judges, parole officers, or law students.[9]

***Transparency and openness***

We report how we determined our sample size, all data exclusions (if any), all manipulations, and all measures in the study. All data and analysis code are publicly available at: https://osf.io/q6npj/?view_only=617dbafda76b49b4b49d6e66fd361922. This study was not pre-registered.

[9] Only 23 participants had any legal experience. Including an indicator variable for legal experience did not affect the pattern of results reported in the main text.

## Results

### *Imprisonment Judgments*

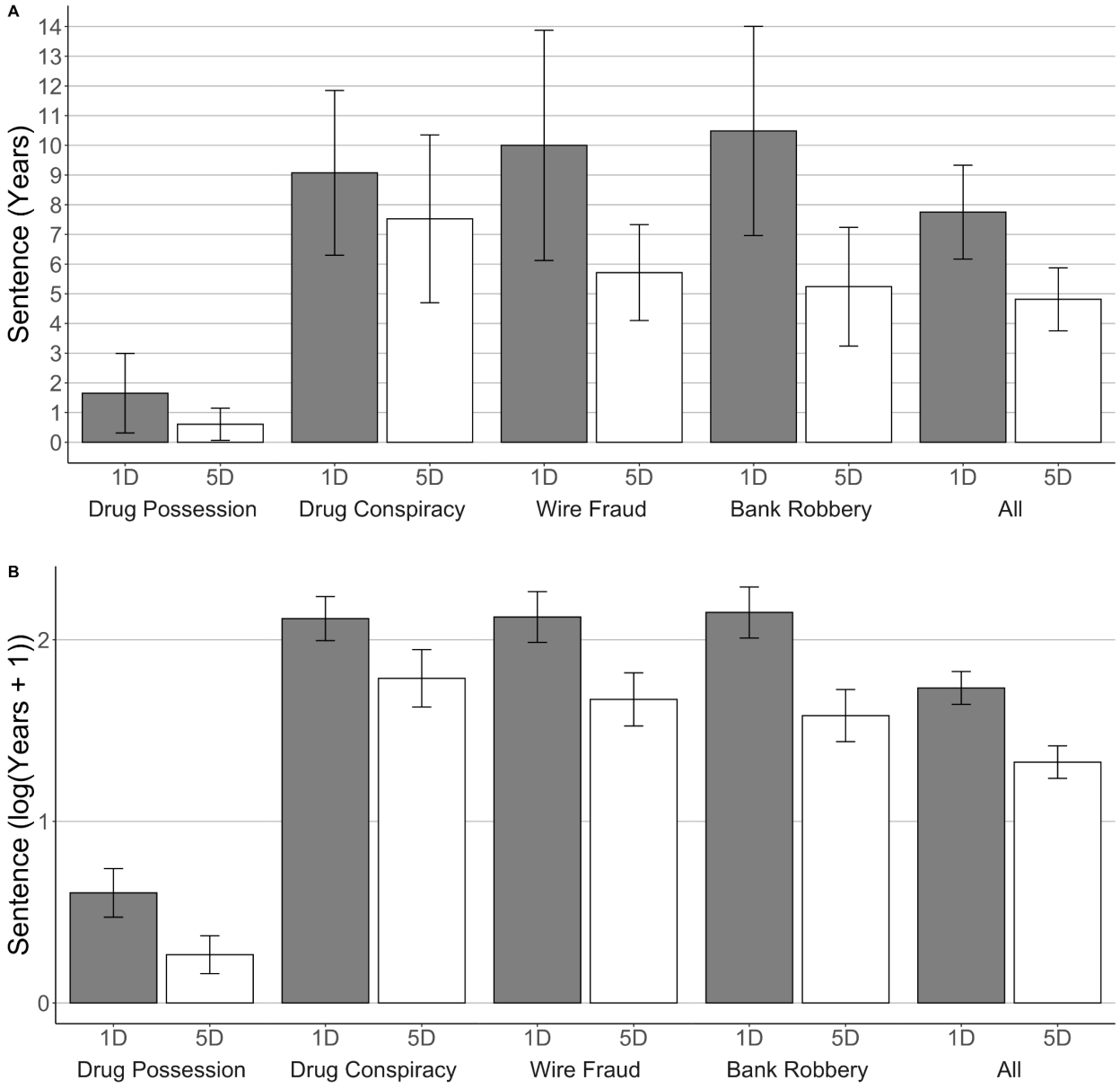
First, we investigated the effect of punishment-type condition (imprisonment-only versus alternative sanctions) on imprisonment judgments. The goal of this analysis was to establish whether participants did, in fact, spontaneously reduce imprisonment judgments when they could impose other types of punishment.

To start, we tested whether the punishment-type condition affected punishment judgments, collapsing across vignette. Regressing imprisonment judgments on punishment-type condition revealed a significant effect of punishment types available, such that participants who answered questions about multiple punishments (alternative sanctions condition; coded as 1) assigned fewer years of imprisonment than participants who just answered a question about imprisonment (imprisonment-only condition; coded as 0), $\beta$ = -2.94 [95% CI: -4.82 to -1.05], $t(235) = -3.06$, $p = 0.002$. Looking at each vignette separately, punishment-type condition significantly predicted imprisonment judgments within the wire fraud ($t(54) = -2.132$, $p = 0.038$) and bank robbery ($t(58) = -2.529$, $p = 0.014$) vignettes but not within the other two vignettes (drug possession: $t(59) = -1.485$, $p = 0.143$; drug conspiracy: $t(58) = -0.804$, $p = 0.425$). The effect of punishment types available on imprisonment judgments across all vignettes appears in **Figure 1**. Actual prison sentences are log-normally distributed (Anderson & Spohn, 2010), and these results show the same pattern. Re-running the analysis using the natural logarithm of the imprisonment judgments also showed a significant effect of punishment-type condition, $\beta$ = -0.41 [95% CI: -0.66 to -0.16], $t(235) = -3.22$, $p = 0.001$.[10]

[10] Participants could and did impose fractional sentences, so these data are not counts that can be modeled with a Poisson or Negative Binomial distribution.

**Figure 1**

*Prison sentences by crime and by punishment-type condition*.




*Note*. Mean prison sentence (A) and log prison sentence (B) by vignette and punishment-type condition (1D = imprisonment only condition; 5D = alternative sanctions condition). Error bars represent 95% confidence intervals.

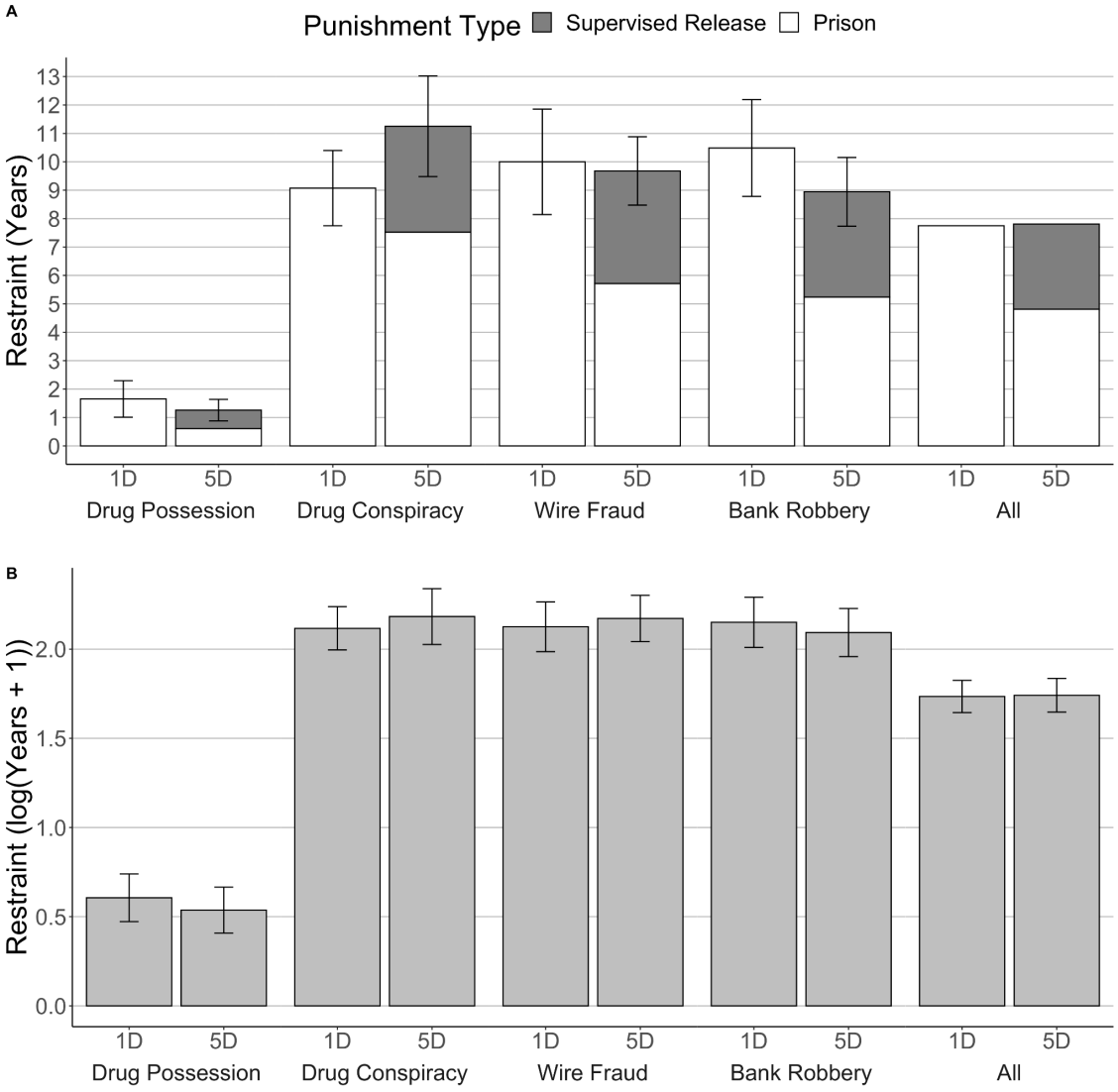
This analysis suggests that type of crime, treated as a categorical variable, significantly affected imprisonment judgments. Regressing imprisonment judgments on dimension condition, vignette (effect-coded), and the interaction again showed a significant effect of dimension, $\beta$ = -3.03 [95% CI: -4.76 to -1.30], $t(229) = -3.45$, $p < 0.001$, as well as a significant effects of each vignette (all $p$s < 0.039), but no significant interactions (all $p$s > 0.146). These results indicate that participants spontaneously reduced imprisonment when other types of punishment were available. Again, re-running the analysis using the natural logarithm of the imprisonment judgments showed the same thing: A significant effect of punishment-type condition, $\beta$ = -0.42 [95% CI: -0.61 to -0.23], $t(229) = -4.40$, $p < 0.001$, significant effects of each vignette (all $p$s < 0.001), but no significant interactions (all $p$s > 0.382). These results suggest that participants thought the alternative sanctions were at least partially commensurable with imprisonment.

***Substitution of Other Modalities for Imprisonment***

Next, we investigated which other types of punishment, if any, participants substituted for imprisonment. The potential substitute most consistent with the incapacitation theory of punishment is supervised release, as it is the next most severe restraint on liberty amongst the modalities tested. Thus, we defined a new variable called “restraint,” which was equal to imprisonment judgments in the unidimensional condition and the sum of imprisonment and supervised release in the multidimensional condition. **Figure 2** shows restraint judgments by vignette and condition in years.

**Figure 2**

*Restraint sentences by crime and by punishment-type condition.*




*Note.* Mean prison sentence plus supervised release both raw (A) and logged (B) by vignette and punishment-type condition (1D = imprisonment only; 5D = multiple punishment types). Intermediate shade of gray in panel B indicates that the log transformation was performed on the summed values of prison and supervised release (i.e., the full bar shown in A). Error bars reflect 95% confidence intervals calculated using restraint (prison + supervised release).

We did not observe a significant effect of punishment-type condition on restraint judgments, $\beta = 0.060$ [95% CI: -0.25 to 0.26], $t(235) = 0.055$, $p = 0.956$. We also did not observe a significant effect of punishment-type condition on restraint within any of the four vignettes (all $p$s > 0.347). Thus, the significant effect of condition on imprisonment judgments did not appear when restraint judgments, rather than imprisonment judgments, served as the dependent variable. We also did not observe a significant effect of punishment-type condition on log-transformed restraint judgments, $\beta = 0.007$ [95% CI: -0.25 to 0.26], $t(235) = 0.051$, $p = 0.959$.

Regressing restraint judgments on punishment-type condition, vignette (effect-coded), and the interaction also showed no significant effect of punishment types available, $\beta = -0.020$ [95% CI: -1.94 to 1.90], $t(229) = -0.020$, $p = 0.984$, significant effects of each vignette (all $p$s < 0.024), and no significant interactions (all $p$s > 0.193). Re-running the analysis using the natural logarithm of the restraint judgments showed the same thing: No significant effect of dimension, $\beta = -0.003$ [95% CI: -0.19 to 0.19], $t(229) = -0.036$, $p = 0.971$, significant effects of each vignette (all $p$s < 0.001), and no significant interactions (all $p$s > 0.675).

The substitution of alternative sanctions for imprisonment is also apparent from inspecting the means of the log-transformed sanction quantities, which appear in **Table 1**.

**Table 1**

*Log-transformed restraint, imprisonment, supervised release, fine, and community service imposed by experimental participants in Study 1 and by federal judges in Study 2, by crime type and experimental condition.*

| | | Mean Amount Imposed (SD) | | |
|---|---|---|---|---|
| | | Experimental Participants | | |
| **Crime** | **Sanction Type** | **1D** | **5D** | **Judges** |
| Bank Robbery | Restraint | 2.15 (0.79) | 2.09 (0.69) | 2.44 (0.17) |
| | Prison | 2.15 (0.79) | 1.58 (0.73) | 2.03 (0.21) |
| | Supervised Release | – | 1.35 (0.67) | 1.56 (0.20) |
| | Community Service | – | 4.97 (2.44) | 0 (0) |
| | Fine | – | 5.90 (4.38) | 0.7 (2.24) |
| Drug Conspiracy | Restraint | 2.12 (0.62) | 2.18 (0.89) | 2.66 (0.32) |
| | Prison | 2.12 (0.62) | 1.79 (0.89) | 2.20 (0.52) |
| | Supervised Release | – | 1.37 (0.64) | 1.77 (0.21) |
| | Community Service | – | 5.07 (2.06) | 0.14 (0.84) |
| | Fine | – | 6.57 (5.19) | 0.76 (2.31) |
| Drug Possession | Restraint | 0.61 (0.73) | 0.54 (0.69) | 1.30 (0.28) |
| | Prison | 0.61 (0.73) | 0.27 (0.56) | 0.11 (0.22) |
| | Supervised Release | – | 0.39 (0.46) | 1.26 (0.27) |
| | Community Service | – | 3.52 (1.66) | 0.90 (1.86) |
| | Fine | – | 5.07 (3.38) | 1.34 (2.77) |
| Wire Fraud | Restraint | 2.13 (0.73) | 2.17 (0.67) | 2.07 (0.45) |
| | Prison | 2.13 (0.73) | 1.67 (0.76) | 1.60 (0.61) |
| | Supervised Release | – | 1.39 (0.71) | 1.31 (0.35) |
| | Community Service | – | 5.37 (2.29) | 0 (0) |
| | Fine | – | 7.76 (5.12) | 0.30 (1.62) |

*Note*. Mean and standard deviation of the log-transformed sanction amounts imposed by experimental participants in Study 1 and federal judges in Study 2. Participants were in one of two sanctions conditions, the imprisonment-only ("1D") condition and the alternative sanctions ("5D") condition. The "Restraint" quantity is the natural logarithm of the sum of the amount of imprisonment and supervised release imposed, with the amount of supervised release treated as 0 for the participants in the imprisonment-only condition. Note that Restraint is not equal to the sum of Prison and Supervised Release in the columns of this table because the sum was taken before the natural logarithm.

***Severity and Punishment Type***

Next, we investigated the degree to which crime severity affected the imposition of each type of punishment. To perform this analysis, we created a severity variable that reflected each participant's severity rating for the crime they sentenced.

First, we checked whether punishment-type condition affected severity in order to ensure that any association between these variables did not bias the results. An independent-samples t-test found no significant difference in severity between the two conditions, $t(230) = 1.58$, $p = 0.116$.

Second, we regressed log-transformed imprisonment judgments on punishment-type condition, severity ratings, their interaction, and vignette (effect-coded). We found a significant effect of punishment type condition, $\beta = -0.503$ [95% CI: -0.670 to -0.336], $t(225) = -5.94$, $p < 0.001$, a significant effect of crime severity, $\beta = 0.017$ [95% CI: 0.013 to 0.021], $t(225) = 8.57$, $p < 0.001$, significant effects of each vignette (all $ps < 0.048$), and no significant interaction, , $\beta = -0.003$ [95% CI: -0.009 to 0.003], $t(225) = -1.09$, $p = 0.275$.

Third, following an approach to analyzing similar data previously used to compare punishment and compensation judgments (Heffner & FeldmanHall, 2019), we analyzed the relationship between crime severity and imprisonment, supervised release, fine, and community service judgments. To perform this analysis, we restricted the sample just to the 118 participants in the alternative sanctions condition. To enable meaningful comparison of judgments across punishment modalities with three different metrics (years, dollars, and hours), we then z-scored the log-transformed imprisonment, supervised release, fine, and community service judgments, as well as the raw severity judgments. We then analyzed the data using a hierarchical linear model, fitted with the *lme4* package (Bates et al., 2015) for R (version 4.2.2), with the standardized amount of punishment imposed (across all four punishment types) as the dependent variable and standardized severity ratings, indicator variables for the type of punishment (imprisonment, supervised release, fine, and community service), their interactions, random intercepts for each participant, and random intercepts for each vignette as independent variables.

Using imprisonment as the reference punishment type, this analysis showed a significant effect of crime severity, $\beta = 0.57$ [95% CI: 0.41 to 0.76], $p < 0.001$, a significant interaction between crime severity and the fine punishment type indicator, $\beta = -0.35$ [95% CI: -0.54 to -0.15], $p < 0.001$, and a significant interaction between crime severity and the community service punishment type indicator, $\beta = -0.36$ [95% CI: -0.55 to -0.17], $p < 0.001$, but no significant interaction between crime severity and the supervised release punishment type indicator, $\beta = -0.03$ [95% CI: -0.22 to 0.16], $p = 0.751$. The analysis also showed no significant effect of the indicator variables for the different punishment types when they were not interacted with severity (all $p$s > 0.739), likely because each punishment type was independently z-scored and

the severity that predicted average imprisonment also predicted averages of the other three punishment types.

***Zero Responses***

Finally, we quantified the percentage of "0" responses participants provided for each sanction in the two conditions. We conducted this analysis to confirm that participants did not interpret the questions eliciting the sanction amounts to require a non-zero answer. The percentage of zero responses for each question in the two Study 1 conditions appear in **Table 2**.

**Table 2**

*Percentage of participants or federal judges who imposed 0 imprisonment, supervised release, restraint, fine, and community service, by crime type.*

| Source | Condition | Crime | Prison Sentences (% 0) | Supervised Release (% 0) | Community Service (% 0) | Fine (%0) |
|---|---|---|---|---|---|---|
| Experimental Participants | 1D | Bank Robbery | 3.03 | - | - | - |
| | | Drug Conspiracy | 0 | - | - | - |
| | | Drug Possession | 41.94 | - | - | - |
| | | Wire Fraud | 0 | - | - | - |
| | 5D | Bank Robbery | 3.70 | 7.41 | 14.81 | 33.33 |
| | | Drug Conspiracy | 6.06 | 6.06 | 9.09 | 33.33 |
| | | Drug Possession | 76.67 | 53.33 | 10.00 | 26.67 |
| | | Wire Fraud | 7.14 | 10.71 | 7.14 | 28.57 |

*Note*. Percentage of experimental participants imposing a zero quantity of each sanction. Experimental participants were in one of two sanctions conditions, the imprisonment-only

("1D") condition and the alternative sanctions ("5D") condition, and one of four crime type conditions.

**Discussion**

The first set of analyses compared prison, supervised release, and restraint sentences between the imprisonment-only and alternative sanctions conditions. These analyses showed that laypeople imposed significantly shorter prison sentences when they had other means of punishing criminal defendants. Strikingly, the lay participants appeared, on average, to replace the forgone prison time with supervised release, such that the restraint variable was almost exactly equal between conditions for all vignettes. The near-perfect equality of the log-transformed restraint judgments between punishment-types conditions for all vignettes, see Table 1, is particularly surprising given the between-subjects design. Two entirely different groups of participants produced near-identical judgments about how long to restrain the defendants in these vignettes, despite the difference in punishment types available.

These results support the view that people think of incarceration and alternative sanctions as at least partially commensurable. Participants reduced their incarceration imposed significantly when they could also impose alternative sanctions and replaced it, nearly one-for-one, with supervised release. This pattern of judgments suggests that participants were sensitive to the ability of different types of punishment to achieve incapacitation, which both forms of restraint-based punishment accomplish to some degree. To the extent participants viewed imprisonment as necessary communicate censure, they apparently thought that carceral sentences roughly one-third shorter than those imposed in the imprisonment-only condition communicated sufficient censure.

Moreover, when mapping between crime severity and sanction amount, participants in the alternative sanctions condition sorted the four different types of punishment into two groups. First, they treated log-transformed imprisonment and log-transformed supervised release similarly, increasing each by roughly half a standard deviation as severity increased by one standard deviation (at average levels of the other punishments). Second, they treated log-transformed fines and log-transformed community service similarly, increasing each by roughly one-fifth a standard deviation as severity increased by one standard deviation (at average levels of the other punishments). The separate treatment of these two groups of sanctions also suggests that participants viewed the two restraint-based punishments as at least partially commensurable, since the imposition of those punishments was similarly responsive to increases in crime severity. By contrast, participants did not respond as directly to increases in crime severity with increases in fines or community service, suggesting that these alternative sanctions may have served a purpose less tightly correlated with crime severity than did the restraint-based punishments. The different slopes of these two groups bears a resemblance to different slopes of punishment and compensation as severity increased shown by Heffner and FeldmanHall (2019).

Finally, participants appeared to understand the task. A potential limitation to this study is that the questions eliciting the sanction quantities asked participants to "[p]lease let us know how many" years, hours, or dollars of the sanction "John should be" sentenced or required to undergo but did not explicitly state that participants could enter "0" as a response. Nevertheless, participants' responses indicate that they understood this fact. Table 2 reflects a significant percentage of zero answers across all sanction types. Unsurprisingly, fewer zero answers are provided for more serious crimes.

## Study 2

To investigate the extent to which laypeople's use of alternative sanctions tracked the use of alternative sanctions by federal judges, Study 2 applied the analyses used in Study 1 to real sentences in comparable cases. We used anonymized individual-level data from initial sentencing hearings in United States District Courts from the United States Sentencing Commission website (United States Sentencing Commission (USSC), 2021) for the twenty-year period from 2002 to 2021.[11] We filtered these cases using criteria derived from the vignettes to find the most analogous cases in this database, as set out in detail below. Then, we compared the sentences imposed in the resulting cases to the sentences the Study 1 participants imposed.

### Method

We used the following procedure to select the cases for comparison from the sentencing data: First, we restricted the data to people with convictions and sentences under relevant statutes and guidelines. We required at least one count of conviction under the following statutes: 21 U.S.C. § 841 (including drug trafficking and possession), 21 U.S.C. § 2113(a),(d) (including bank robbery with and without a dangerous weapon), 18 U.S.C. § 1343 (including wire fraud), and 18 U.S.C. § 924(c) (including possession of a firearm during a drug trafficking offense or a crime of violence, which is consistent with the facts of the drug trafficking and bank robbery vignettes). We also required that the primary guideline calculation have occurred under United States Sentencing Guidelines §§ 2B1.1 (including fraud offenses), 2D1.1 (including drug offenses), or 2B3.1 (including robbery offenses).[12]

---

[11] These are the years for which the United States Sentencing Commission made anonymized individual-level sentencing data available on its website.

[12] Any conviction under 18 U.S.C. § 924(c) would not make the guideline for that offense, USSC § 2K2.4(b), the primary guideline, as that guideline simply adds a 5-year mandatory consecutive sentence onto the guideline for the primary offense.

Second, we filtered to defendants in criminal history category I, which includes both defendants with no prior convictions and defendants with a single prior 1-point offense – i.e., a conviction leading to an incarceration of less than 60 days. All the defendants described in the vignettes had no criminal history at all, so this category is slightly overbroad. However, narrowing the sample to include defendants only if they had exactly 0 criminal history points did not meaningfully affect the point estimates reported below and left relatively small samples for two of the crimes (bank robbery and drug possession).

Third, we filtered to defendants who received a 2- or 3-point reduction to their offense level for acceptance of responsibility (USSC § 3E1.1, 2014). All the defendants described in the vignettes pled guilty to their crimes and did not go to trial. Although pleading guilty does not guarantee a 2- or 3-point reduction under this section, it is "significant evidence" for acceptance of responsibility (USSC § 3E1.1 n.3, 2014). Failure to receive the reduction after pleading guilty is strong evidence that the defendant engaged in other behavior surrounding the investigation or trial that the judge viewed negatively.

Fourth, we filtered to defendants for whom the sentencing court conducted just a single sentencing computation. This criterion selected cases with just a single group of offenses and did not exclude defendants with an 18 U.S.C. § 924(c) count (possession of a firearm during a crime of violence or drug trafficking offense), as that statute does not require a guidelines computation and therefore does not contribute to the number of computations. Under the guidelines, offenses may be grouped when – as for the drug and fraud guidelines – the offense level is determined primarily by some aggregate quantity, which can be aggregated across counts (USSC § 3D1.2, 2014). Using this approach instead of restricting just to cases with a single count avoids excluding cases where, for example, a person engaged in wire fraud has been charged with a

separate count for each victim. However, restricting just to defendants with 1 or 2 counts (to ensure inclusion of cases with 18 U.S.C. § 924(c) charges) does not change the analyses presented below.

Finally, we used a series of crime-specific filters to narrow down each category of cases to those most similar to the vignettes. The exact filter imposed depended on both the crime and the precise edition of the United States Sentencing Guidelines Manual used to sentence the defendant, as the base offense levels associated with each particular guideline and the level adjustments associated with the specific offense characteristics has differed over the years. The defendants in the filtered data were sentenced between 2002 and 2020 using the Guidelines Manuals from 1993 through 2018.[13] More specific descriptions of these restrictions are described in the Appendix.[14]

This procedure yielded 630 total cases, including 76 drug possession cases, 388 drug conspiracy cases, 111 wire fraud causes, and 55 bank robbery cases. The defendants were 66% White or Caucasian, 30% Black or African American, 2% American Indian or Alaska Native, 2% Asian or Pacific Islander, and 1% other or missing; 40% Hispanic, 52% non-Hispanic, and 8% with missing ethnicity information; and 89% male and 11% female. Although the judge who imposed the sentence is more analogous to the participants in Study 1, the sentencing data do not report that information. These data report just the court in which a judge sentenced the defendant, and all United States District Courts contain more than one judge.

---

[13] Defendants' guidelines ranges are determined using either the edition of the Guidelines Manual in effect when they committed the crime or the edition of the Guidelines Manual in effect at the time of sentencing, whichever yields the lowest recommended sentencing range, because applying a more punitive later manual to earlier conduct would violate the *ex post facto* clause (USSC § 1B1.11; *Peugh v. United States*, 569 U.S. 530 (2013)).

[14] The precise selection criteria are reflected in the analysis code available at the link in the Author Note.

To conduct our analyses, we identified variables in the sentencing data most similar to the outcome variables in Study 1. For prison sentence, we combined two variables – one reporting all non-zero prison sentences including fractions for months, time served, and other credits, and the other including zeros for defendants who did not receive any prison sentence – to create one variable that equaled 0 when no prison sentence was imposed and equaled a fraction for sentences including day or month increments. For supervised release, we summed the variable reporting supervised release and the variable reporting probation, the latter of which is similar to supervised release except that it can be imposed only when no prison sentence is imposed. We allowed participants in Study 1 to impose a sentence of pure supervised release, so the most comparable variable is the sum of these two quantities. For fines, we used the difference between a variable reporting the total monetary penalties imposed—summing fines, restitution, and supervision costs—and the amount of restitution. In Study 1, we did not give participants the option of requiring the defendant to pay supervision costs, but we concluded that those fees fit naturally under fines, since they are both payments to the state imposed as part of a punishment and the sentencing data sometimes include supervision costs in the fine variable (USSC, 2022). The sentencing data did not include a variable reflecting the amount of forfeiture. As participants in Study 1 responded to a binary question about forfeiture, we excluded forfeiture from our analysis.

***Transparency and openness***

We report how we determined our sample size, all data exclusions (if any), all manipulations, and all measures in the study. The analysis code for this study is publicly available at https://osf.io/q6npj/?view_only=617dbafda76b49b4b49d6e66fd361922. The data are

available at https://www.ussc.gov/research/datafiles/commission-datafiles. This study was not pre-registered.

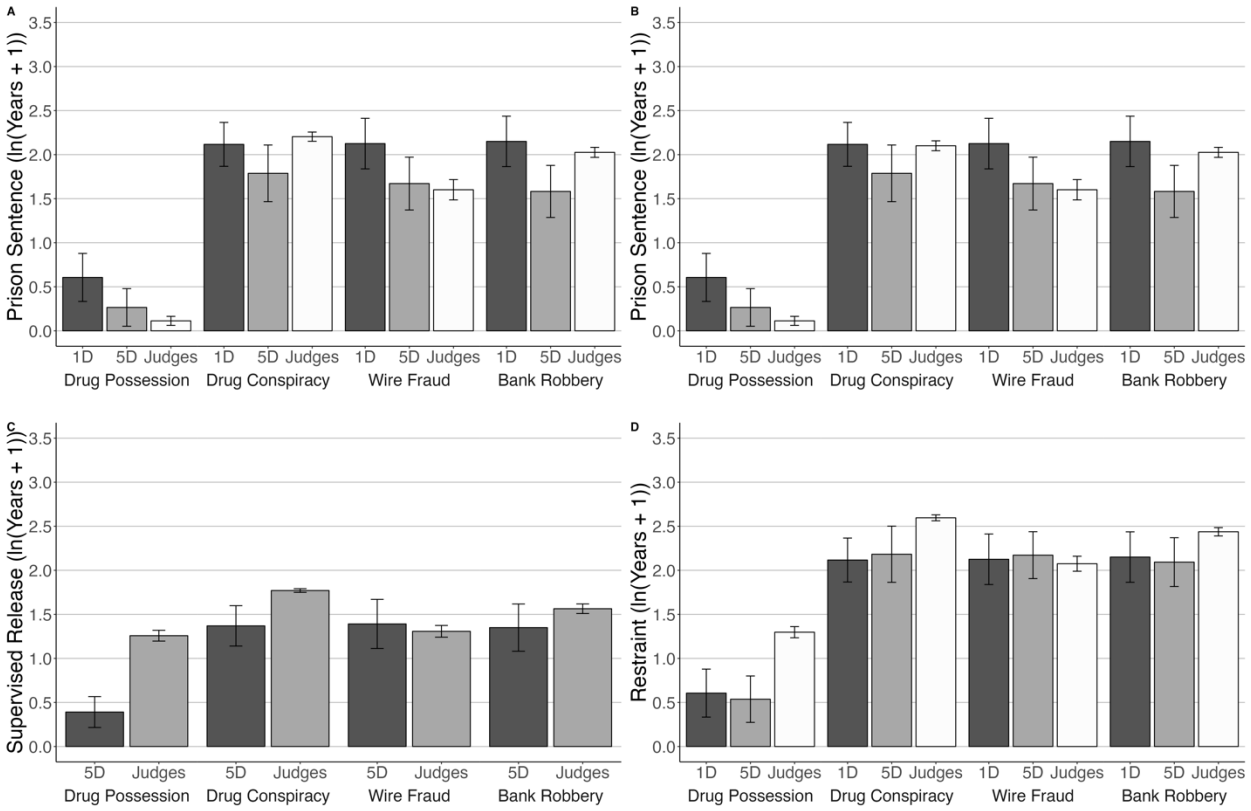
## Results

### *Prison and Restraint Sentences*

We have the sentencing judgements of federal judges only in a situation similar to the alternative sanctions condition, where they had multiple different forms of punishment at their disposal. Although we therefore cannot compare the judges' sentences across those two conditions, we can compare them by crime type. The mean sentences federal judges imposed by crime appear in **Figure 3** and **Table 1**, along with the mean sentences imposed by the experimental participants.

**Figure 3**

*Prison, supervised release, and restraint sentences imposed by experimental participants and federal judges, separated by crime.*




*Note.* Mean restraint sentences imposed by experimental participants in the imprisonment-only ("1D") condition, the alternative sanctions ("5D") condition and by federal judges ("Judges") across all four vignettes. Error bars reflect 95% confidence intervals. The sanctions visualized are (A) log-transformed years of prison imposed; (B) log-transformed years of prison imposed after excluding defendants in the real drug conspiracy cases who were subject to a 15-year mandatory minimum; (C) log-transformed years of supervised release imposed; and (D) log-transformed years of restraint imposed, where restraint is years of prison plus years of supervised release for participants in the alternative sanctions ("5D") condition and federal judges ("Judges") and years of prison imposed for participants in the imprisonment-only condition ("1D").

This graph suggests that judges' sentences, overall, resemble the experimental participants' sentences quite closely, with some exceptions investigated in more detail below. In addition, as discussed below, comparing the "Judges" bars for the drug conspiracy cases in panels A and B, the former of which includes defendants subject to a statutory mandatory minimum for possessing a firearm under 18 U.S.C. § 924(c) and the latter of which does not, shows that the presence of this mandatory minimum did not meaningfully constrain judges' sentencing discretion. Statistical tests of the comparison between judges and experimental participants appear in **Table 3**. Before that comparison, however, we analyzed the effect of severity on judges' sentences.

***Severity and Punishment Type***

Next, we investigated the degree to which crime severity affected judges' use of each type of punishment. To perform this analysis, we used the standardized offense level calculated by the judge before sentencing as a standardized crime severity rating.

Again, following the approach to analyzing similar date in the literature (Heffner & FeldmanHall, 2019), we tested the relationship between crime severity and imprisonment, supervised release, community service, fine, and restitution judgments. To enable meaningful comparison of judgments across punishment modalities with three different metrics (years, dollars, and hours), we then z-scored the log-transformed imprisonment, supervised release, community service, fine, and restitution judgments. We then analyzed the data using a hierarchical linear model, fitted with the *lme4* package (Bates et al., 2015) for R (version 4.2.2), with the standardized amount of punishment imposed (across all five punishment types) as the dependent variable and standardized severity ratings, indicator variables for the type of punishment (imprisonment, supervised release, fine, and community service), their interactions,

random intercepts for each participant, and random intercepts for each vignette as independent variables.

Using imprisonment as the reference punishment type, this analysis showed a significant main effect of crime severity, $\beta = 0.78$ [95% CI: 0.69 to 0.87], $p < 0.001$, a significant interaction between crime severity and the supervised release punishment type, $\beta = -0.28$ [95% CI: -0.38 to -0.18], $p < 0.001$, a significant interaction between crime severity and the community service punishment type, $\beta = -0.99$ [95% CI: -1.09 to -0.89], $p < 0.001$, a significant interaction between crime severity and the fine punishment type, $\beta = -0.83$ [95% CI: -0.93 to -0.73], $p < 0.001$, and a significant interaction between crime severity and the restitution punishment type, $\beta = -0.75$ [95% CI: -0.85 to -0.65], $p < 0.001$. The analysis also showed no significant effect for the indicator variables for the different punishment types when they were not interacted with severity (all $p$s > 0.851), likely because each punishment type was independently z-scored and the severity that predicted average imprisonment also predicted averages of the other three punishment types.

***Comparison to Study 1***

Because the experimental participants in Study 1 and the federal judges in Study 2 imposed sentences using overlapping punishment types, we were also able to compare the amount of each sanction that the two sets of actors imposed. To do so, we regressed each sanction amount on an indicator variable for the sentence source (i.e., experimental participants, coded as 0, or judges, coded as 1) in a dataset restricted to just one type of offense. **Table 3** shows the effect size of the difference between the two sets of judgments by crime and outcome variable.

**Table 3**

*Difference between federal judges and experimental participants in imprisonment, supervised release, restraint, fine, and community service judgments, by crime type.*

| | **Crime** | | | |
|---|---|---|---|---|
| **Dependent Variable** | **Drug possession** | **Drug conspiracy** | **Wire fraud** | **Bank robbery** |
| ln(Prison + 1) **(1D only)** | -0.913*** [-1.252, -0.574] | 0.152 [-0.209, 0.515] | -0.781*** [-1.177, -0.386] | -0.215 [-0.601, 0.17] |
| ln(Prison + 1) **(5D only)** | -0.356* [-0.709, -0.005] | 0.568*** [0.295, 0.842] | -0.102 [-0.492, 0.288] | 0.823*** [0.431, 1.215] |
| ln(Restraint + 1) **(1D + 5D)** | 1.35*** [1.024, 1.676] | 0.866*** [0.677, 1.056] | -0.126 [-0.427, 0.174] | 0.58** [0.202, 0.957] |
| ln(Supervised Release + 1) **(5D only)** | 2.301*** [1.924, 2.678] | 0.847*** [0.647, 1.048] | -0.151 [-0.484, 0.183] | 0.436* [0.043, 0.831] |
| ln(Fine + 1) **(5D only)** | -1.208*** [-1.616, -0.799] | -1.446*** [-1.68, -1.211] | -1.964*** [-2.261, -1.667] | -1.496*** [-1.913, -1.079] |
| ln(Community Service + 1) **(5D only)** | -1.484*** [-1.923, -1.046] | -3.13*** [-3.353, -2.905] | -3.323*** [-3.585, -3.06] | -2.88*** [-3.257, -2.503] |

*Note.* Effect size, in Cohen's *d* value, of switching from a sentence imposed by lay study participants in the condition indicated under the DV name to a sentence imposed by a federal judge. Positive values reflect federal judges imposing higher sentences than Study 1 participants; negative values reflect federal judges imposing lower sentences than Study 1 participants. Asterisks reflect significance of mean difference used to compute the *d* value according to a t-test: *** < 0.001; ** < 0.01; * < 0.05.

We conducted additional analysis on three additional points: First, we investigated whether the judges' sentences differed pre- and post- *Booker*, in order to determine whether it would be necessary to make separate comparisons between judicial sentences in these two time periods and laypeople's sentences. Second, we investigated whether the comparison between judges' and laypeople's sentences in the drug conspiracy cases differed depending on whether a mandatory minimum applied. Third, we explored the reasons for the dramatic differences in the use of fines and community service by judges and lay participants.

As reviewed above, *United States v. Booker* (2005) fundamentally altered the nature of the sentencing guidelines, changing them from mandatory to advisory. After *Booker*, federal judges have been empowered to impose sentences outside the guidelines sentence range and have regularly imposed sentences below that range (Hofer, 2019). Nevertheless, we find no effect of *Booker* on any sanction type for any crime type in the judge's sentences. Using just the database of judge's sentences, we estimated four models regressing the log-transformed quantity of each sanction (imprisonment, supervised release, community service, and fines) on a dummy variable for whether the sentence was pre- or post-*Booker* and dummy variables for the crime type. The *Booker* dummy variable was not significant in any model (all $p$s > 0.116).

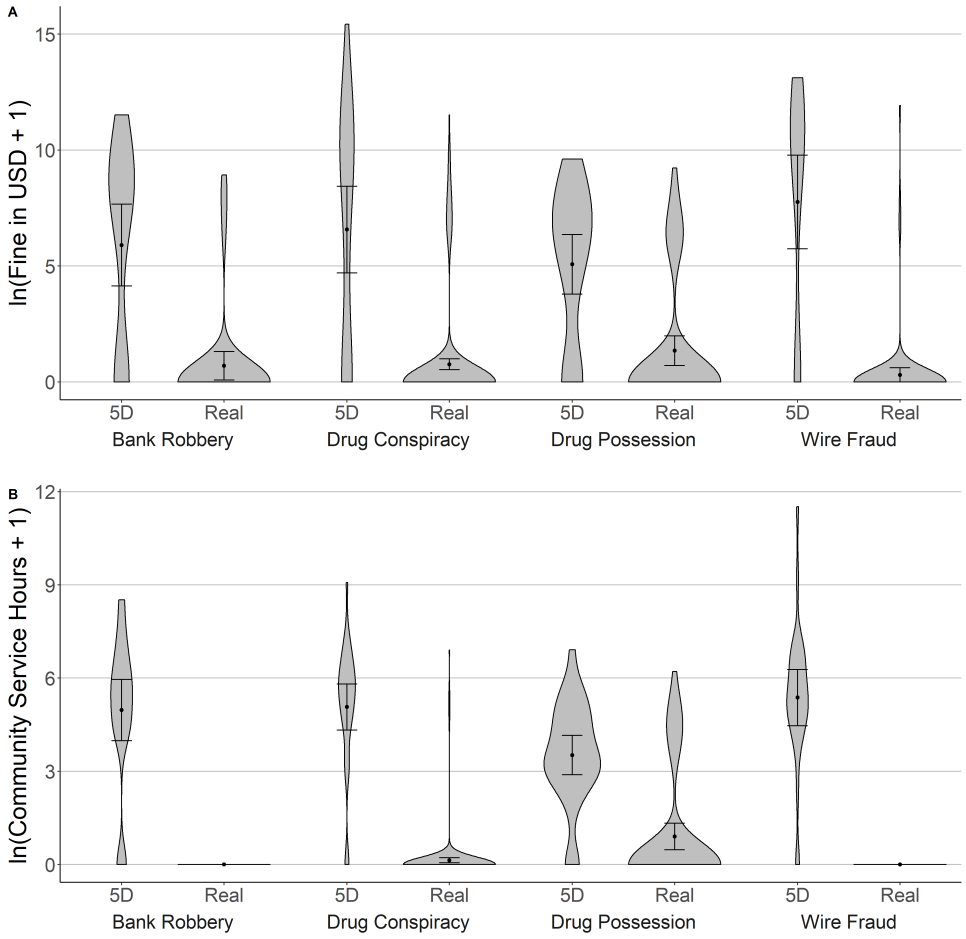
As discussed above, both the drug conspiracy and bank robbery cases were potentially subject to a 5-year mandatory minimum prison sentence under 18 U.S.C. § 924(c) if the prosecutor chose to charge that offense. The application of a mandatory minimum could have placed the judges in a different position than the lay participants, who had no such restrictions on the sentences they were allowed to impose. Although every single bank robbery case in our sample was subject to the five-year mandatory minimum, the mandatory minimum applied in just 115 of the 388 drug conspiracy cases. Therefore, using that crime as a case study, we

conducted two sets of analyses to assess the effect of the presence of a statutory minimum sentence on judges' sentencing judgments. First, using just the data on judges' sentences in the drug conspiracy case, we regressed log-transformed quantities of each sanction type on an indicator variable for whether the statutory minimum applied. The application of the statutory minimum made no significant difference to the amount of any sanction (all *p*s > 0.290). Second, we regressed the amount of each sanction on an indicator variable for the source of the sentences (judges vs. lay participants) in a subsample comprising only drug conspiracy cases in which no mandatory minimum applied. The effect size estimates were all within one one-hundredth of those in the drug conspiracy column of Table 3, indicating that the mandatory minimum sentence did not drive the effects described there. The absence of any difference when taking this factor into account is visually apparent when comparing the "Judges" bar for the drug conspiracy vignette in panels A and B of Figure 3.

Turning to the difference in the quantity of fines and community service imposed: The reason judges and laypeople differed so dramatically in these quantities is that the fines and community service imposed by judges clustered near zero for all the crimes. For the drug possession crime, there is a cluster of cases at around $500 in fines and around 90 hours of community service. But for the other, more serious crimes, nearly all the fine and community service sentences in real cases were at or around zero, as the violin plots in **Figure 4** demonstrate.

**Figure 4**

*Violin plots of fines and amounts of community service imposed by experimental participants and federal judges, by crime type.*




*Note.* Violin plot of log-transformed fine in dollars and community service hours imposed by experimental participants in the alternative sanctions ("5D") condition and by federal judges ("Real") across all four vignettes. Points reflect mean values and error bars reflect 95% confidence intervals.

**Discussion**

Federal judges showed significant differences in imprisonment, supervised release, and restraint judgments by crime type. Using standardized punishment amounts, federal judges also

showed a strong relationship between crime severity and the amounts of imprisonment and supervised release, with a one-standard deviation increase in severity predicting a 0.78 standard deviation increase in imprisonment and a 0.50 standard deviation increase in supervised release. Federal judges showed almost no association between crime severity and quantity of community service, fines, or restitution. A one standard deviation increase in severity predicted a 0.21 standard deviation *decrease* in community service, a 0.05 standard deviation decrease in fine, and a 0.03 standard deviation increase in restitution. These relationships between crime severity and different types of sanctions are similar to the relationships shown in the judgments of experimental participants, though federal judges make much less use of alternative sanctions such as community service and fines to respond to increasingly severe crimes.

Due to the use of multiple punishment types in the measures provided to experimental participants, we were also able to directly compare the punishment judgments of federal judges to those of participants. A striking pattern emerged. Compared to participants in the imprisonment-only condition, federal judges' prison sentences were significantly lower with a large effect size for drug possession and wire fraud offenses – the two least serious offenses by offense level – but did not differ significantly for drug conspiracy and bank robbery offenses – the two most serious offenses by offense level. Conversely, compared to participants in the alternative sanctions condition, their sentences were significantly higher with a moderate-to-large effect size for drug conspiracy and bank robbery offenses but did not differ significantly for wire fraud and were significantly lower but with a small effect size for drug possession. In addition, because judges imposed significantly more supervised release for the two drug offenses and the bank robbery offense, judges ultimately imposed significantly more restraint than the experimental participants for both drug offenses and bank robbery, all with large effect size.

In contrast to the restraint sentences, for which judges' and participants' responses were closely comparable, the amount of fines and community service differed dramatically between the experimental sentences and the real sentences. Experimental participants imposed substantially larger fines and more hours of community service than federal judges. Indeed, federal judges generally imposed zero fines and zero community service, except in the least serious crime considered (drug possession). This finding suggests that federal judges view fines and community service as alternative sanctions to impose in less serious cases where the amount of prison is low, while laypeople see a greater role for fines and community service across a wide spectrum of cases. In other words, federal judges do not view fines and community service as commensurable with incarceration to the same degree as lay participants.

## General Discussion

Theorists writing about punishment have argued that real-world punishment practices do, and should, to a large degree reflect the views of laypeople (Kahan, 1998; Robinson & Darley, 2007; Robinson & Kurzban, 2007). Experimental studies using one-dimensional punishment scales – i.e., asking participants just about how much time a particular person should spend in a carceral facility – have found evidence that laypeople's judgments generally follow a *retributivist* theory of punishment (Carlsmith et al., 2002; Robinson & Darley, 1995; Robinson et al., 2012). However, other researchers have found evidence that laypeople also pursue other ends for which they are sometimes willing to sacrifice retributivist goals, such as incapacitation (Slobogin & Brinkley-Rubinstein, 2013) and expressing censure (Funk et al., 2014; Nahmias & Aharoni, 2017). This prior research has also suggested that laypeople view alternative sanctions as at least partially commensurable with incarceration, contrary to the arguments of some legal theorists (Husak, 2017; Kahan, 1998). First, laypeople appear to prioritize theories of punishment

(like retributivism) consistent with commensurability (Robinson & Spellman, 2005). Second, prior work studying the perceived severity of different types of punishment has not shown complete sorting by punishment type (Serra, 1978).

The present studies further investigated the degree to which laypeople view alternative sanctions as commensurable with incarceration and how their judgments compare to the sentences imposed by federal judges. Study 1 tested how both the availability of different types of sanctions and crime severity affect laypeople's punishment judgments, in order to explore the degree to which people make trade-offs between different types of punishment. Study 2 then identified federal criminal cases that were analogous to the vignettes used in Study 1 and examined whether federal judges showed patterns of judgment similar to those of laypeople.

Five main findings emerged. First, when given more punishment options, laypeople significantly reduced the amount of imprisonment they imposed. Second, laypeople substituted supervised release for imprisonment in a one-to-one ratio on average when they had the option. Third, experimental participants increased the amount of imprisonment and supervised release they imposed as crime severity increased at rates that did not differ significantly, while they increased community service and fines at much lower rates. Fourth, federal judges showed a similar but more extreme pattern, increasing imprisonment and supervised release at a moderate to high rate as severity increased but not increasing fines or restitution at all and reducing community service. Fifth, experimental participants imposed dramatically larger fines and more hours of community service than federal judges, who imposed effectively none.

The first two findings suggest that laypeople see supervised release and incarceration as partially commensurable – one can freely substitute for the other up to a point. Participants, on average, chose to reduce the amount of imprisonment imposed by a least a third when they had

the option to impose supervised release, and they did so for all the crimes tested. This pattern of responses is not consistent with the view that laypeople think incarceration and alternative sanctions are incommensurable past some severity level (Kahan, 1998). This pattern of responses also provides further evidence about the way laypeople think about incapacitation. Prior research suggests that factors related to incapacitation, such as likelihood of recidivism, do not drive lay sentencing judgments to the same degree as factors related to retribution, like crime seriousness (Darley et al., 2000). However, factors related to incapacitation may play at least a secondary a role in lay punishment judgments (Slobogin & Brinkley-Rubinstein, 2013). In the present study, the near equality between the imprisonment sentences imposed by participants in the imprisonment-only condition of Study 1 and the sum of imprisonment and supervised release (restraint) imposed by participants in the alternative sanctions condition (see Table 1) suggests that laypeople viewed these two incapacitative punishments as closely comparable. That is, even if incapacitation is not a primary driver of punishment quantity (Darley et al., 2000), laypeople appear sensitive to the category of incapacitative punishments when trading off between different punishment types.

The third finding, that experimental participants increase imprisonment and supervised release at a significantly higher rate than fines and community service as crime severity increases, provides evidence that they think the two restraint-based punishments serve a similar function, such as retribution or incapacitation. This finding is consistent with prior work showing that crime seriousness is the primary factor in laypeople's judgments about appropriate punishment (Warr et al., 1983). The quantity of these two punishments that the lay participants imposed increased in response to increases in crime severity at rates that did not differ significantly from one another. However, the smaller but still positive correlations between crime

severity and both fines and community service suggest that participants used alternative sanctions to pursue other goals that were less tightly correlated with crime severity. For example, participants may have viewed community service and fines as serving the goals of *restoration* or *educative communication* – that is, teaching defendants how to behave normatively through building habits of normative behavior (Ho et al., 2019; Sarin et al., 2021), such as helping others (community service) or giving up ill-gotten benefits (fines).

Fourth, as compared with lay participants in Study 1, federal judges in Study 2 showed a similar but more extreme relationship between crime severity and the amount of each type of punishment imposed. Imprisonment quantity showed a strong positive correlation with crime severity – an unsurprising association, given that the crime severity metric used was offense level, which directly determines the sentencing range for imprisonment. More surprising was the similar association between supervised release imposed and crime severity. Under the sentencing guidelines, judges determine imprisonment and supervised release separately, adding supervised release on top of the prison sentence in an amount that does not depend directly on the offense level. Instead, it depends primarily on the class of the offense, which is based on the statutory maximum prison sentence for the crime of conviction (USSC § 5D1.2(a); 18 U.S.C. § 3559(a), 2018). These results suggest that judges may use the offense level to guide their supervised release decisions even though the guidelines do not call for it. Another striking finding was the null or negative association between crime severity and community service, fines, and restitution in federal judges' sentencing decisions. This result arises from the fact that federal judges rarely used any of these alternative sanctions, imposing substantial amounts of community service only for drug possession, the least serious of the three offenses. This finding suggests that the primary concerns of judges when imposing sentences differ meaningfully from the concerns of lay

participants. One possible source of this difference is that judges may attend to practical constraints on the imposition of each punishment type, such as the defendant's ability to pay, the government's available supervision resources, or the consequences of imposing on-going post-release obligations that defendants will likely find difficult to meet, leading to violations and possible additional prison time (Crow & Bales, 2006).

Finally, the sentences imposed by laypeople and federal judges were broadly similar, which is consistent with some older research but inconsistent with more recent studies. Rossi and colleagues showed that laypeople's judgments about prison sentences largely followed the sentencing guidelines in place at the time (Rossi et al., 1997; Rossi & Berk, 1997). However, in 2005, judges were freed from the constraints of the mandatory guidelines (*United States v. Booker*, 2005) and began regularly varying below them (Hofer, 2019). Moreover, a later study with real jurors showed that their sentencing judgments were significantly less harsh than the guidelines (Gwin, 2010). The present research, finding reasonably close alignment between laypeople's and federal judges' sentencing judgments, is more in line with the older findings. One possible explanation for this difference could be that Gwin's (2010) study used real jurors, who had just seen the defendant in person in the courtroom, while the present research and the older research used experimental participants, who provided sentencing judgments in response to written vignettes (Rossi et al., 1997; Rossi & Berk, 1997). Laypeople may be less lenient when the defendant receiving the sentence is less of an abstraction.

The differences that did appear between laypeople's and judge's sentences are of interest, however, and underscore the value of eliciting laypeople's punishment judgments both with and without alternative sanctions. Federal judges' imprisonment decisions most closely resembled laypeople's (longer) imprisonment decisions in the imprisonment-only condition – and differed

significantly from laypeople's (shorter) imprisonment decisions in the alternative sanctions condition – for the two offenses with the highest offense levels: Bank robbery and drug conspiracy. Conversely, federal judges' imprisonment decisions most closely resembled laypeople's (shorter) imprisonment decisions in the alternative sanctions condition – and differed significantly from laypeople's (longer) imprisonment decisions in the imprisonment-only conditions – for the two offenses with the lowest offense levels: Drug possession and wire fraud. These results suggest that federal judges prioritize imprisonment more for the crimes the law treats as most serious. In other words, for the most serious crimes, the view that the unique censuring function of imprisonment places a floor on carceral punishment (Husak, 2020; Kahan, 1998), and that therefore incarceration and alternative sanctions are incommensurable for these offenses, is more consistent with the sentencing decisions of federal judges than with the sentencing decisions of laypeople.

In addition to shedding light on how laypeople think about punishment, these results have normative implications for legal punishment practices. Part of the justification for legal punishment is that it imposes and enforces community values– that is, laypeople's beliefs about how bad certain behaviors are and what response they deserve (Gwin, 2010; Robinson & Darley, 1995; Robinson & Kurzban, 2007) – and experimental results can provide information about those values (Sommers, 2021). Indeed, theorists have argued that the criminal law, at least in part, implements a community's moral beliefs (Dan-Cohen, 1984; Hart, 1967),[15] which implies

[15] One could argue that, in some contexts, the law should not respond to public opinion but should help lead the public to some more normatively desirable view. That argument is more plausible for highly visible public policies – like the change in public opinion after the United States Supreme Court protected the right to same-sex marriage nationwide in *Obergefell v. Hodges* (2015; Ofosu et al, 2019) – than it is for the intricacies of the federal sentencing guidelines.

that the law should not deviate too dramatically from those beliefs without good reason. Apart from these normative claims, bringing punishment practices into line with laypeople's views may have implications for legal compliance. Laypeople may view punishment practices that do not correspond to those standards as illegitimate (Robinson & Darley, 2007; Tyler & Jackson, 2014), leading to reduced compliance with the law (Tyler, 2006) or less willingness to cooperate with authorities or actively engage with their communities (Tyler & Jackson, 2014).

These results therefore suggest reforms to current federal sentencing practices to bring them into line with community views, though additional research will be necessary to confirm these findings before using them to justify a change in policy. Under the federal sentencing guidelines, judges engage in detailed factfinding and calculations to ascertain the recommended range for a defendant's prison sentence (e.g., USSC ch. 5, pt. A, 2018). The recommended amount of supervised release, by contrast, is based on the class of the offense as determined by the statutory maximum rather the specifics of the offense. The present results, which show laypeople treating imprisonment and supervised release as partially interchangeable and highly sensitive to offense seriousness, suggest that the more detailed procedure currently used for determining a recommended imprisonment range should instead recommend a restraint range. Judges could then determine what proportion of that restraint term should be served in prison and what proportion as supervised release, with the latter proportion defaulting to roughly 1/3. Based on the current results, this procedure would more closely match the way laypeople think about these different types of sanctions, at least for the crimes tested.

Similarly, these results suggest that alternative sanctions like fines or community service should play a greater role in sentencing than they currently do, particularly for more serious crimes. Laypeople showed substantial uncertainty about the precise amount of these sanctions to

impose, as reflected in Figure 4, with values stretching many orders of magnitude. This finding is consistent with research into jury damage awards for emotional damages, which show substantial variance unless guided by anchors and instructions (Hans et al., 2022). However, lay participants broadly agreed that the value should be above zero, which is where the quantities imposed in real sentences cluster. This implication is, however, subject to an important caveat: The practical concerns and consequences of these alternative sanctions. It is well-documented that court-imposed fines and supervision can trap people who have served their prison sentences in a cycle of violations and court hearings, as these obligations can make it difficult to live on the wages paid by jobs defendants are able to obtain post-release (Barkow, 2019). These practical concerns provide reasons to deviate from community standards. The practical feasibility of imposing these alternative sanctions is therefore an important topic to study before implementing any reforms.

**Limitations and Future Directions**

The present study assessed punishment judgments for four common federal crimes: possession of a controlled substance with intent to distribute, participation in a drug distribution conspiracy while in possession of a firearm, bank robbery, and wire fraud. We chose these offenses because they are federal-law analogues of common state crimes (drug possession, fraud, robbery, and drug trafficking), which makes these findings potentially relevant outside the context of federal criminal law, while permitting a comparison of participants' sentences to judicial sentences in publicly available data. These four crimes yielded a wide range of seriousness judgments, which allowed us to assess the effect of changing the available punishment modalities at many different levels of crime seriousness. However, future research using a wider variety of crimes and state sentencing data could confirm the generalizability of these findings and provide a firm foundation for changes in sentencing policy.

In addition, experimental participants deciding on sentences were confronted with a 1-2 paragraph description of the crime and the defendant for each crime. Federal judges, by contrast, receive a Presentence Investigation Report (PSR) from the probation office that includes a lengthy description of the facts of the crime, details on the defendant's criminal, social, educational, employment, family, and medical histories, and a full calculation of the applicable guidelines range (18 U.S.C. § 3552). Differences in the facts or the characteristics of the offenders not used in the case selection procedure could, potentially, explain the differences in sentences imposed by laypeople and federal judges. The significance of this limitation depends on whether the assumptions participants drew about the characteristics of the offenders described in the vignettes match the mean characteristics of the real defendants. However, this concern should not be overstated. Although some PSRs are long and detailed, many are fairly bare-bones, containing just a few sentences of straightforward description in each category of information.[16]

Finally, using a between-subjects design in this study avoided the possibility that participants' punishment judgments would interfere with one another, yielding comparative rather than absolute punishment judgments. However, there may be heterogeneity in how laypeople conceptualize different forms of punishment, which could lead to individual-level differences in their patterns of responses. Future research applying a within-subjects design could examine potential individual differences in the degrees to which people are willing to substitute alternative sanctions for imprisonment. This research could provide additional insight

[16] PSRs are not subject to the right of public access that applies to most court documents (*United States v. Corbitt*, 1989), so there are not many examples available to support this point. However, this example PSR distributed at a workshop by the Inter-university Consortium for Political and Social Research at the University of Michigan is illustrative: https://www.icpsr.umich.edu/summerprog/2009/nijworkshop/PSRDrugScenario.pdf

into laypeople's views of the functions of different forms of punishment, which could aid in developing sentencing procedures that more closely reflect community norms.

## Conclusion

Judges in criminal cases must use multiple different types of punishment to accomplish the purposes of punishment – e.g., reflecting the seriousness of the crime, deterring future offenses, incapacitating the person who committed the crime (18 § U.S.C. 3553(a)). The present research shows that when laypeople engage in the same task, they substitute a less serious restraint-based punishment – supervised release – one-for-one for around a third of the carceral sentence they would otherwise impose, and they engage in more substitution as the seriousness of the crime increases. These patterns suggest that laypeople do not agree with the view that incarceration and alternative sanctions are incommensurable. Instead, laypeople treat incarceration and alternative sanctions as partially commensurable. Laypeople also appear to use different types of punishment to accomplish different penological goals, using restraint-based sanctions in a way that is more directly responsive to crime severity than their use of fines and community service. Ultimately, when given the chance to employ multiple different punishment modalities, laypeople do not place the same emphasis on imprisonment that characterizes media reports about crime, scientific research on the topic, and federal sentencing procedures.

# Appendix

## Study 1 Vignette Text

The three additional vignettes used in Study 1 were:

*Drug Possession with Intent to Distribute*
John is 26 years old. He just started selling marijuana on the street in a major American city. He buys the drugs he sells from a single supplier, but doesn't know anything about the supplier's business. When he is selling to others, John never carries more than a half a pound of marijuana at a time. When he sells the drugs on the street, he sells small plastic bags containing an eighth of an ounce of marijuana. Sometimes, he will sell several of these at a time. John has received his high school diploma and has no criminal record.

One day, after he has been selling marijuana for just over a month, he is arrested while carrying a pound of marijuana. He is charged with and pleads guilty to possession with intent to distribute a controlled substance.

*Wire Fraud*
John is 26 years old. He just started running a website that is directed at homeowners going through a mortgage foreclosure. On the website, he claims that he can help homeowners modify their loans and stay in their homes if they pay him a flat fee of $1,000. John has programmed his computer to keeps track of all the foreclosure cases that are filed in several different states and to automatically generate letters to homeowners facing foreclosure once cases against them have been filed. He never provides borrowers with any help. Instead, he collects their money, loan information, and financial information, and then stops communicating with them. To date, borrowers have paid him over $100,000. John has received his high school diploma and has no criminal record. One day, while he is working on his computer, John is arrested. He is charged with and pleads guilty to wire fraud.

*Bank Robbery*
John is 26 years old. For several months, he has been planning on robbing a large bank in the city where he lives. One day, he enters the bank and goes up to one of the tellers. He slides the teller a note saying that he has a gun and will use it unless she hands over $10,000 and does not in any way indicate that something is wrong. The teller activates a silent alarm and hands over $10,000. As soon as John leaves the bank, he is arrested by the police. They find a small revolver in the pocket of his coat. John has received his high school diploma and has no criminal record.

He is charged with and pleads guilty to bank robbery.

**Study 2 Case Selection Criteria**

*Guideline 2B1.1* (Fraud Offenses). For the wire fraud offense, we restricted to defendants with a statutory minimum sentence of 0, a base offense level of 4, 6, 7 (depending on the guidelines year), a 4-level adjustment for number or number and effect on victims (depending on the guidelines year), and a 2-level adjustment for either sophisticated means or more than minimal planning (depending on the guidelines year). We also restricted based on the specific offense characteristic related to loss, but this restriction is complex. This specific offense characteristic increases the offense level based on the amount of money that the defendant either actually caused the victims to lose or *intended* the victims to lose, whichever is greater (USSC § 2B1.1 n.3, 2014). The vignette described a defendant who mailed a fraudulent letter to everyone his computer detected had a foreclosure filed against them but only succeeded in obtaining $1,000 each from 100 people. Thus, the actual loss was $100,000 but the intended loss was unknown and possibly extremely high. For the purpose of this filter, therefore, we included the specific offense level increases corresponding to a loss of between $100,000 and $10,000,000, with the exact size of the adjustment depending on the year of the guidelines. We also filtered to ensure that the only guideline provisions affecting the sentence were the base offense levels and specific offense characteristics mentioned above, as well as acceptance of responsibility.

*Guideline 2D1.1* (Drug Offenses). For the marijuana possession offense, we restricted to defendants with a substance code corresponding to marijuana, with a statutory minimum sentence of 0, a base offense level of 6 or 8 (reflecting that the defendant possessed 1lb of marijuana, with the exact level depending on the year of the guidelines used for the calculation), and a final offense level affected by the base offense level and acceptance of responsibility adjustments alone (i.e., not affected by any other specific offense characteristics).

For the cocaine conspiracy and firearm offense, we restricted to defendants with a substance code corresponding to cocaine, with a statutory minimum sentence of 120 or 180 (allowing the former only when no 18 U.S.C. § 924(c) was charged but the firearm specific offense characteristic was added and allowing the latter only when it included 60 months for a § 924(c) violation), a base offense level of 28, 30, or 32 (reflecting that 5kg of cocaine is on the border between two offense levels, though which two depends on the guidelines year), and a final offense level affected by just the base offense level, acceptance of responsibility adjustments, and the firearm specific offense characteristic.

*Guideline 2B3.1* (Bank Robbery). For the bank robbery offense, we restricted to defendants who had a base offense level of 20 and a 2-level increase for targeting a financial institution. We also had to impose a restriction related to the possession of a firearm. The vignette describes a bank robbery in which the defendant robs a bank while in possession of a firearm but without using or brandishing it. This behavior could increase the defendant's offense level in one of two ways: If the prosecutor included a count charging a violation of 18 U.S.C. § 924(c) and the court agreed that armed bank robbery while merely in possession of a weapon met the statutory definition of a crime of violence, then the defendant would receive a mandatory 60-month consecutive sentence on top of the sentence for robbery. If no such charge were brought, then the defendant would have their offense level increased by 5 under the relevant specific offense characteristic. We therefore filtered to cases where the defendant either had a 60-month mandatory minimum for an 18 § 924(c) violation or a 5-level increase for possession of a firearm. Finally, we filtered to ensure that the only guideline provisions affecting the sentence were the base offense levels and specific offense characteristics mentioned above, as well as acceptance of responsibility.

## References

Ahler, D. J., Roush, C. E., & Sood, G. (2019). The micro-task market for lemons: Data quality on Amazon's Mechanical Turk. *Political Science Research and Methods*, 1-20, https://doi.org/10.1017/psrm.2021.57

Alter, A. L., Kernochan, J., & Darley, J. M. (2007). Transgression wrongfulness outweighs its harmfulness as a determinant of sentence severity. *Law and Human Behavior, 31*, 319-335, https://doi.org/10.1007/s10979-006-9060-x

Anderson, A. L., & Spohn, C. (2010). Lawlessness in the federal sentencing process: A test for uniformity and consistency in sentence outcomes. *Justice Quarterly*, 27, 362-393, https://doi.org/10.1080/07418820902936683

Bailis, D.S., Darley, J.M., Waxman, T.L., & Robinson, P.H. (1995). Community standards of criminal liability and the insanity defense. *Law and Human Behavior, 19*, 425-446, https://doi.org/10.1007/BF01499336

Bates, D., Mächler, M., Bolker, B., Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*, 1–48. doi:10.18637/jss.v067.i01.

Blumstein, A. & Cohen, J. (1980). Sentencing of convicted offenders: An analysis of the public's view. *Law & Society Review, 14*, 223-261. https://www.jstor.org/stable/3053313

*Booker v. United States*, 543 U.S. 220 (2005)

Bregant, J., Shaw, A., & Kinzler, K. D. (2016). Intuitive jurisprudence: Early reasoning about the functions of punishment. *Journal of Empirical Legal Studies, 13*, 693-717, https://doi.org/10.1111/jels.12130

Brooks, T. (2012). *Punishment*. Routledge.

Brown, M., Keefer, L. A., Sacco, D. F., & Brown, F. L. (2022). Demonstrate values: Behavioral displays of moral outrage as a cue to long-term mate potential. *Emotion, 22*, 1239-1254, https://doi.org/10.1037/emo0000955

Buhrmester, M., Kwang, T., & Gosling, S. D. (2011). Amazon's Mechanical Turk: A new source of inexpensive, yet high-quality, data?. *Perspectives on Psychological Science, 6*, 3-5, https://doi.org/10.1177/1745691610393980

Carlsmith, K. M., Darley, J. M., & Robinson, P. H. (2002). Why do we punish? Deterrence and just deserts as motives for punishment. *Journal of Personality and Social Psychology, 83*, 284-299, https://doi.org/10.1037/0022-3514.83.2.284

Carlsmith, K. M. (2006). The roles of retribution and utility in determining punishment. *Journal of Experimental Social Psychology, 42*, 437-451, https://psycnet.apa.org/doi/10.1016/j.jesp.2005.06.007

Chmielewski, M., & Kucker, S. C. (2020). An MTurk crisis? Shifts in data quality and the impact on study results. *Social Psychological and Personality Science, 11*, 464-473, https://doi.org/10.1177/1948550619875149

Cohen, J. (1983). Incapacitation as a strategy for crime control: Possibilities and pitfalls. *Crime and Justice, 5*, 1-84.

Cohen, A., & Yang, C. S. (2019). Judicial politics and sentencing decisions. *American Economic Journal: Economic Policy*, 11, 160-91, https://doi.org/ 10.1257/pol.20170329

Crow, M. S., & Bales, W. (2006). Sentencing guidelines and focal concerns: The effect of sentencing policy as a practical constraint on sentencing decisions. *American Journal of Criminal Justice, 30*, 285-304.

Cullen, F. T., Fisher, B. S., & Applegate, B. K. (2000). Public opinion about punishment and corrections. *Crime and Justice, 27*, 1-79.

Cushman, F. (2008). Crime and punishment: Distinguishing the roles of causal and intentional analyses in moral judgment. *Cognition*, 108, 353-380, https://doi.org/10.1016/j.cognition.2008.03.006

Cushman, F. A., Sarin, A., & Ho, M. (2019). Punishment as communication. In J. Doris & M. Vargas (Eds.), *The Oxford Handbook of Moral Psychology*. Oxford University Press.

Dan-Cohen, M. (1984). Decision rules and conduct rules: On acoustic separation in criminal law. *Harvard Law Review, 97*, 625-677, https://www.jstor.org/stable/1340892

Darley, J. M., Carlsmith, K. M., & Robinson, P. H. (2000). Incapacitation and just deserts as motives for punishment. *Law and Human behavior, 24*, 659-683, https://doi.org/10.1023/A:1005552203727

Devine, D. J., Olafson, K. M., Jarvis, L. L., Bott, J. P., Clayton, L. D., & Wolfe, J. M. (2004). Explaining jury verdicts: Is leniency bias for real?. *Journal of Applied Social Psychology, 34*, 2069-2098, https://doi.org/10.1111/j.1559-1816.2004.tb02691.x

Duff, A. (2001). *Punishment, communication, and community*. Oxford University Press.

Dunlea, J. P., & Heiphetz, L. (2021). Moral psychology as a necessary bridge between social cognition and law. *Social Cognition, 39*, 183-199, https://doi.org/10.1521/soco.2021.39.1.183

Feinberg, J. (1965). The expressive function of punishment. *The Monist, 49*, 397-423, https://www.jstor.org/stable/27901603

Funk, F., McGeer, V., & Gollwitzer, M. (2014). Get the message: Punishment is satisfying if the transgressor responds to its communicative intent. *Personality and Social Psychology Bulletin, 40*, 986-997, https://doi.org/10.1177/0146167214533130

Ginther, M. R., Bonnie, R. J., Hoffman, M. B., Shen, F. X., Simons, K. W., Jones, O. D., & Marois, R. (2016). Parsing the behavioral and brain mechanisms of third-party punishment. *Journal of Neuroscience, 36,* 9420-9434, https://doi.org/10.1523/JNEUROSCI.4499-15.2016

Greene, E.J. & Darley, J.M. (1998). Effects of necessary, sufficient, and indirect causation on judgments of criminal liability. *Law and Human Behavior, 22*, 429-451, https://psycnet.apa.org/doi/10.1023/A:1025723010408

Gwin, J. S. (2010). Juror sentiment on just punishment: Do the federal sentencing guidelines reflect community values. *Harvard Law & Policy Review, 4*, 173-200.

Hampton, J. (1984). The moral education theory of punishment. *Philosophy & Public Affairs, 13*, 208-238.

Hampton, J. (1992). Correcting harms versus righting wrongs. *UCLA Law Review, 39*, 1659-1702.

Hans, V. P., Reed, K., Reyna, V. F., Garavito, D., & Helm, R. K. (2022). Guiding jurors' damage award decisions: Experimental investigations of approaches based on theory and practice. *Psychology, Public Policy, and Law, 28*, 188-212, https://doi.org/10.1037/law0000342

Harlow, R.E., Darley, J.M., & Robinson, P.H. (1995). The severity of intermediate penal sanctions: A psychophysical scaling approach for obtaining community perceptions.

Journal of Quantitative Criminology, 11, 71-95, https://psycnet.apa.org/doi/10.1007/BF02221301

Hart, H. L. (1967). Social solidarity and the enforcement of morality. The University of Chicago Law Review, 35, 1-13.

Heffner, J., & FeldmanHall, O. (2019). Why we don't always punish: Preferences for non-punitive responses to moral violations. *Scientific Reports, 9*, 1-13, https://doi.org/10.1038/s41598-019-49680-2

Hofer, P. J. (2019). Federal sentencing after *Booker*. *Crime and Justice, 48*, 137-186, https://doi.org/10.1086/701712

Husak, D. (2020), The metric of punishment severity: A puzzle about the principle of proportionality. In M. Tonry (Ed.), O*f One-Eyed and Toothless Miscreants: Making the Punishment Fit the Crime?*, pp. 97–126. Oxford University Press.

Johnson, D., & Ryan, J. B. (2020). Amazon Mechanical Turk workers can provide consistent and economically meaningful data. *Southern Economic Journal, 87*, 369-385, https://doi.org/10.1002/soej.12451

Kahan, D. M. (1996). What do alternative sanctions mean?. *The University of Chicago Law Review*, 63, 591-653, https://doi.org/10.2307/1600237

Kahan, D. M. (1998). Punishment incommensurability. *Buff. Crim. L. Rev., 1*, 691-708, https://doi.org/10.1525/nclr.1998.1.2.691

Kalven Jr, H., & Zeisel, H. (1966). *The American jury*. Little, Brown.

Lipsey, M. W., & Cullen, F. T. (2007). The effectiveness of correctional rehabilitation: A review of systematic reviews. *Annual Review of Law and Social Science, 3*, 297-320.

MacCoun, R. J., & Kerr, N. L. (1988). Asymmetric influence in mock jury deliberation: Jurors' bias for leniency. *Journal of Personality and Social Psychology, 54*, 21-33, https://psycnet.apa.org/doi/10.1037/0022-3514.54.1.21

Martin, J. W., & Heiphetz, L. (2021). "Internally wicked": Investigating how and why essentialism influences punitiveness and moral condemnation. *Cognitive Science*, 45, e12991, https://doi.org/10.1111/cogs.12991

McConnell, B. & Rasul, I. (2018). Hispanic-white sentencing differentials in the federal criminal justice system. Working Paper.

McFatter, R. M. (1982). Purposes of punishment: Effects of utilities of criminal sanctions on perceived appropriateness. *Journal of Applied Psychology, 67*, 255.

*Molina-Martinez v. United States*, 136 S. Ct. 1338 (2016)

Moynihan, C. (2022). Avenatti gets a 4-year sentence for defrauding Stormy Daniels of $300,000, *New York Times*, https://www.nytimes.com/2022/06/02/nyregion/michael-avenatti-sentenced-stormy-daniels.html

Nadelhoffer, T., Heshmati, S., Kaplan, D., & Nichols, S. (2013). Folk retributivism and the communication confound. *Economics & Philosophy, 29*, 235-261, https://doi.org/10.1017/S0266267113000217

Nahmias, E., & Aharoni, E. (2017). Communicative theories of punishment and the impact of apology. In C.W. Surprenant (Ed.) *Rethinking punishment in the era of mass incarceration* (pp. 144-161). Routledge.

*Obergefell v. Hodges*, 576 U.S. 644 (2015)

Ofosu, E. K., Chambers, M. K., Chen, J. M., & Hehman, E. (2019). Same-sex marriage legalization associated with reduced implicit and explicit antigay bias. *Proceedings of the National Academy of Sciences, 116*, 8846-8851.

*Peugh v. United States*, 569 U.S. 530 (2013)

Petersilia, J. & Deschenes, E.P (1994). Perceptions of punishment: Inmates and staff rank the severity of prison versus intermediate sanctions. *The Prison Journal, 74*, 306-328, https://doi.org/10.1177/0032855594074003003

Primoratz, I. (1989). Punishment as language. *Philosophy, 64*, 187-205, https://www.jstor.org/stable/3751407

Robinson, P.H. & Darley, J.M. (1995). *Justice, liability, and blame*. Westview Press.

Robinson, P.H. (2000). Testing lay intuitions of justice: How and why? *Hofstra Law Review, 28*, 611-634.

Robinson, P.H. & Spellman, B.A. (2005). Sentencing decisions: Matching the decisionmaker to the decision nature. *Columbia Law Review, 105*, 1124-1161.

Robinson, P.H. & Darley, J.M. (2007). Intuitions of justice: Implications for criminal law and justice policy. *Southern California Law Review, 81*, 1-68.

Robinson, P.H. & Kurzban, R. (2007). Concordance and conflict in intuitions of justice. *Minnesota Law Review*, 91, 1829-1907.

Robinson, P.H., Jackowitz, S., & Bartels, D.M. (2012). Extralegal punishment factors: A study of forgiveness, hardship, good-deeds, apology, remorse, and other such discretionary factors in assessing criminal punishment. *Vanderbilt Law Review, 65*, 737-826.

Rossi, P. H., Berk, R. A., & Campbell, A. (1997). Just punishments: Guideline sentences and normative consensus. *Journal of Quantitative Criminology, 13*, 267-290, https://doi.org/10.1007/BF02221093

Sebba, L. (1978). Some explorations in the scaling of penalties. *Journal of Research in Crime and Delinquency, 15*, 247-265.

Sebba, L., & Nathan, G. (1984). Further explorations in the scaling of penalties. *The British Journal of Criminology, 24*, 221-249.

Shariff, A. F., Greene, J. D., Karremans, J. C., Luguri, J. B., Clark, C. J., Schooler, J. W., Baumeister, R.F., & Vohs, K. D. (2014). Free will and punishment: A mechanistic view of human nature reduces retribution. *Psychological Science, 25,* 1563-1570, https://doi.org/10.1177/0956797614534693

Silver, J. R. (2017). Moral foundations, intuitions of justice, and the intricacies of punitive sentiment. *Law & Society Review, 51*, 413-450, https://www.jstor.org/stable/45093847

Silver, J. R., & Berryessa, C. M. (2021). Remorse, perceived offender immorality, and lay sentencing preferences. *Journal of Experimental Criminology*, 1-39, https://doi.org/10.1007/s11292-021-09488-5

Slobogin, C. & Brinkley-Rubinstein, L. (2013). Putting desert in its place. *Stanford Law Review, 65*, 77-136.

Stalans, L. J., & Diamond, S. S. (1990). Formation and change in lay evaluations of criminal sentencing. *Law and Human Behavior*, 14, 199-214, https://psycnet.apa.org/doi/10.1007/BF01352749

Stephen, J.F. (2014/1883). *A history of the criminal law of England* (Vol. 2). Cambridge University Press.

Taylor, K. (2019). By turns tearful and stoic, Felicity Huffman gets 14-day prison sentence. *New York Times*, https://www.nytimes.com/2019/09/13/us/felicity-huffman-sentencing.html

Tonry, M. (1996). *Sentencing matters*. Oxford University Press.

Tyler, T. (2006). *Why people obey the law*. Princeton University Press.

Tyler, T. R., & Jackson, J. (2014). Popular legitimacy and the exercise of legal authority: Motivating compliance, cooperation, and engagement. *Psychology, Public Policy, and Law, 20,* 78-95, https://doi.org/10.1037/a0034514

United States Sentencing Commission (2014). *2014 federal sentencing guidelines manual*. https://www.ussc.gov/guidelines/archive/2014-federal-sentencing-guidelines-manual

United States Sentencing Commission (2018). *2018 guidelines manual annotated*. https://www.ussc.gov/guidelines/2018-guidelines-manual-annotated

United States Sentencing Commission. (2021). Commission datafiles, 2002-2021 [Data set]. https://www.ussc.gov/research/datafiles/commission-datafiles

*United States v. Corbitt*, 879 F.2d 244 (7th Cir., 1989)

*United States v. Collins*, 828 F. 3d 386 (6th Cir. 2016)

*United States v. Obiora*, 910 F. 3d 555 (1st Cir. 2018)

Von Hirsch, A. (2017). *Deserved criminal sentences*. Bloomsbury Publishing.

Warr, M., Meier, R. F., & Erickson, M. L. (1983). Norms, theories of punishment, and publicly preferred penalties for crimes. *Sociological Quarterly, 24*, 75-91, https://doi.org/10.1111/j.1533-8525.1983.tb02229.x

Wilson, J. Q. (1983). Dealing with the High-Rate Offender. *The Public Interest, 72*, 52-71.

Yang, C.S. (2015). Free at last? Judicial discretion and racial disparities in federal sentencing. *Journal of Legal Studies*, 44, 75-111, https://www.jstor.org/stable/10.1086/680989

Zimring, F. E., & Hawkins, G. (1995). *Incapacitation: Penal confinement and the restraint of crime*. Oxford University Press.

18 U.S.C. § 924(c)

18 U.S.C. § 1343

21 U.S.C. § 841

21 U.S.C. § 2113

18 U.S.C. § 3553