

## TESTING THE TRANSITIVITY OF REPORTED RISK PERCEPTIONS: EVIDENCE OF COHERENT ARBITRARINESS\*

KYLE J. THOMAS,<sup>1</sup> BENJAMIN C. HAMILTON,<sup>1</sup>  
and THOMAS A. LOUGHRAN<sup>2</sup>

<sup>1</sup>Department of Criminology and Criminal Justice, University of Missouri—St. Louis

<sup>2</sup>Department of Criminology and Criminal Justice, University of Maryland—College Park

**KEYWORDS:** deterrence, risk perceptions, coherent arbitrariness

*An often implicit assumption of perceptual deterrence tests is that the elicited values pertaining to arrest risk reflect stable underlying beliefs. But researchers in other disciplines have found that reported expectations are highly susceptible to exogenous factors (e.g., anchors and question ordering), indicating that such values are somewhat arbitrary responses to probabilistic questions. At the same time, reported expectations are coherent within persons, such that respondents rank order them rationally. For deterrence, then, absolute values reported on arrest risks are likely not stable but individuals still rank order specific crimes in meaningful ways. We examine the interpretability of reported arrest risk for three possibilities: 1) Reported risks are stable probabilistic values; 2) reported risks are arbitrary and uninformative for deterrence research; or 3) reported risks display “coherent arbitrariness” with unstable values between individuals but stable rank ordering of crimes within individuals. Through the use of three random experiments of college students, our results indicate that elicited risk perceptions are arbitrary in that they are influenced by the presentation of anchors and question ordering. Nevertheless, the rank ordering of crimes within and across conditions is unaffected by the presentation of anchors, suggesting that reported risks are locally coherent within persons.*

According to deterrence theory, variation in individual offending is related to variation in perceived arrest risk, so that as the subjective probability of getting caught for offending increases, the likelihood or rate of offending should decrease. This simple idea is at the heart of criminal justice policy in the United States (Kennedy, 2009; Sherman, 1990), and deterrence remains one of the preeminent perspectives in criminology (Paternoster,

---

\* Additional supporting information can be found in the listing for this article in the Wiley Online Library at <http://onlinelibrary.wiley.com/doi/10.1111/crim.2018.56.issue-1/issuetoc>.

Greg Pogarsky, Jody Miller, and the anonymous reviewers provided helpful comments on earlier versions of this article. We would also like to acknowledge the Emperor of Wyoming, Ray Paternoster, both for his comments on this project as well as for his incredible support and influence during our careers. We will miss you, Ray.

Direct correspondence to Kyle J. Thomas, Department of Criminology and Criminal Justice, University of Missouri—St. Louis, 331 Lucas Hall, St. Louis, MO 63121 (e-mail: [thomaskj@umsl.edu](mailto:thomaskj@umsl.edu)).

2010). Traditionally, tests of deterrence have relied on between-individual analyses that compare offending levels (or intentions) of individuals reporting high subjective arrest risk to those who report the probability of arrest as lower (Paternoster, 1987), with the results typically indicating that there is a statistically significant, although substantively weak, effect (Pratt et al., 2006). These findings have led many scholars to question whether sanction threats have a meaningful impact on behavior (Pratt, 2008). Even supporters of deterrence theory have concluded that the statistically significant deterrent effect “must be swallowed with a hefty dose of caution and skepticism” (Paternoster, 2010: 765; see also Nagin, 1998).

One reason for skepticism centers around the way deterrence theory has been tested as scholars have traditionally used regression techniques and concluded that *any* statistically significant association is evidence of deterrence, regardless of the effect size or the model’s overall explanatory power (Paternoster, 2010). Beyond the concerns of weak design and misspecification presented with this empirical strategy (Paternoster, 1987), a more fundamental issue with testing deterrence centers on the *measurement* of risk perceptions. In most studies in which deterrence is tested, researchers have relied on survey questions that ask participants’ perceived risk of arrest on scales ranging from 0 percent to 100 percent (Horney and Marshall, 1992; Jensen, Erickson, and Gibbs, 1978; Kleck et al., 2005). For comparison purposes, researchers, often implicitly, assume that reported probabilities reflect stable beliefs and can be interpreted as literal values with ratio properties. That is, an individual who reports a 60 percent likelihood of arrest on a survey for a certain crime type genuinely believes that this is his or her perceived arrest risk, and that it is 20 percent higher than another individual who reports a 40 percent chance of arrest for the same crime. The implications of this interpretation are that 1) individuals hold stable subjective beliefs on the likelihood of arrest that are reflected in responses to probabilistic expectation questions and 2) the probability scale is directly comparable for all individuals. Numerous researchers of perceptual deterrence subscribe to this interpretation as it is necessary for studies in which the deterrent effect across individuals is examined. Nagin (1998: 17), however, openly questioned whether reported risk perceptions elicited in surveys are “anchored in reality,” and other scholars have called for tests of the assumption that individuals report probabilistic expectations in ways that allow for relevant between-person comparisons (Pogarsky and Loughran, 2016).

The elicitation of subjective and probabilistic expectations in surveys is a complicated process that has been studied extensively by scholars in other disciplines (Manski, 2004). The basic yet inescapable finding of this research is that considerable *instability* likely exists in the reporting of subjective probabilities. Research outcomes in psychology and economics demonstrate that individuals often do not fully understand probabilities (Alberini et al., 2004), frequently round and report probabilistic expectations in terms of focal responses<sup>1</sup> (e.g., 0 percent, 50 percent, 100 percent; Hudomiet and Willis, 2012; Hurd, 2009; Manski and Molinari, 2010), and anchor responses to meaningless information in the study design (Ariely, Loewenstein, and Prelec, 2003; Tversky and Kahneman, 1974). Put simply, the absolute value responses to probabilistic questions can be impacted by exogenous factors (e.g., anchors and question framing) that are wholly unrelated to the

---

1. For instance, an individual who reports 50 percent may be simply signaling he or she has maximum uncertainty about the true probability, not that he or she believes there to be a 50/50 chance of detection.

respondents' true subjective beliefs. Indeed, the reported values in subjective expectation surveys are often *arbitrary* responses to questions of probabilistic expectations, and the assumption that they can be interpreted and compared across persons may be tenuous. These concerns have only recently attracted the interest of criminologists (Loughran, Paternoster, and Thomas, 2014; Pickett, Loughran, and Bushway, 2015), but an important implication is that the arrest risk perceptions used in prior work as predictors in regressions may simply not be precise measures of individuals' subjective beliefs about arrest.

Nevertheless, between these two polar opposite interpretations—that elicited risk perceptions reflect stable subjective beliefs or instead are completely arbitrary and, in turn, worthless—there is a third plausible interpretation. Research results in judgment and decision-making have found that responses to subjective probability items, although not stable in a numerically absolute sense, are not entirely random, either. Once an initial subjective probability is selected—or “imprinted”—reported probabilities tend to become locally coherent within individuals, so that the *relative rank order of probabilistic responses is static and predictable*. For example, although individuals may not hold (and report) a stable absolute value on the probability of dying before 80 years of age, they consistently rank this probability as higher than the probability of dying before 50 years of age. Thus, subjective probabilities “display a combination of arbitrariness and coherence” referred to as “coherent arbitrariness” (Ariely, Loewenstein, and Prelec, 2003: 74). Kahneman and colleagues (1993) argued that a less-than-literal interpretation of subjective beliefs “does not deny the existence of *relatively stable* attitudes towards some issues and objects” but simply acknowledges that values elicited in surveys are “[susceptible] to framing effects” (pp. 310–1, emphasis added).<sup>2</sup> In this way, the strong assumption that the absolute values of reported probabilities reflect stable beliefs that can be compared across persons may not be a necessary condition for theories purporting a relationship between subjective expectations and behavior because internal coherence within persons can still allow for powerful tests of theories.

If subjective arrest risk perceptions demonstrate coherent arbitrariness, then this would have important implications for deterrence research. It would suggest individuals' reported arrest risk perceptions may be susceptible to anchoring and framing effects that impact one's subjective probability distribution, and thus, reported risk perceptions are unstable in an absolute sense. As a result, a comparison of reported arrest risks—and its relationship to offending—*between* individuals may be misleading because a reported value of 40 percent might mean something different to different individuals (i.e., the scale is not absolute).<sup>3</sup> Also, the rank order of crime types *within* an individual likely remains mostly unaffected by the arbitrary nature of probabilistic reporting. An individual may not hold and report a stable absolute probability on the risk of arrest for drunk driving, but he or she may believe that this risk is less than the risk of arrest for armed robbery, and order the crime types accordingly. In other words, exogenous factors may shift one's arrest probability distribution toward 0 or 1, but how one rank orders the arrest risk of crimes remains unchanged. Therefore, although treating any two individuals' risk

- 
2. Both Ariely, Loewenstein, and Prelec (2003) and Kahneman et al. (1993) considered the willingness to pay rather than subjective risk, although we stress that the distinction is unimportant as both can essentially be thought of as a price.
  3. That is, individuals who report a 40 percent chance of arrest may have *greater* risk perceptions than those who report a 60 percent chance of arrest.

perceptions as comparable may be problematic, a “perceived risk of arrest” construct is not entirely inconsequential. Indeed, Pogarsky and Loughran (2016) argued that meaningful relative values within persons—that is, coherently arbitrary risk perceptions—would indicate that perceived risks are still “anchored in reality” and is a sufficient condition for deterrence to operate as hypothesized.

We examine elicited subjective risk perceptions for three possible interpretations: 1) They are stable probabilistic values, 2) they are completely arbitrary and hold no value in deterrence research, or 3) they follow a pattern of coherent arbitrariness. We explore these competing hypotheses by looking at three experimental studies where individuals are 1) randomly provided an “informative” anchor (i.e., ostensibly related to the issue of interest) on the probability of being arrested for stealing from a store; 2) randomly provided a “noninformative” anchor on the likelihood of contracting a sexually transmitted infection (STI); and 3) provided surveys that randomly vary the first crime type in a risk perception scale. If reported risks reflect literal values, then they should be unaffected by anchoring or design effects (i.e., stable values and stable ordering). Alternatively, if they are completely arbitrary, then we would expect to observe design and anchoring effects that make comparisons both between and within individuals nonsensical (unstable values and unstable ordering). If, however, reported risks are characterized by coherent arbitrariness, then we would expect to see anchoring and design effects on the distribution of responses between experimental conditions but consistent rank ordering of crime types within conditions (unstable values but stable ordering).

## SANCTION RISK PERCEPTIONS AND DETERRENCE

Deterrence theory is based on the standard economic assumption that humans are rational beings and refrain from engaging in criminal behavior when the marginal cost exceeds the marginal benefit (Becker, 1968). Deterrent effects are most commonly estimated by regressing (self-reported) offending on the perceived risk of arrest ( $p$ ). If the coefficient on  $p$  is negative and statistically significant, it is interpreted as evidence that individuals consider and weigh the risk of formal sanctions before offending in a manner consistent with deterrence. Although some scholars have suggested that many tests of this linkage have lacked appropriate methodological techniques to establish causality (Paternoster et al., 1983; Saltzman et al., 1982), the results of extant research in which deterrence has been tested offer only minimal support for this hypothesis overall. Nagin (1998), Paternoster (1987, 2010), and Pratt et al. (2006) all provided similar conclusions: The effect of perceived risk of arrest on crime is, at best, weak and of low explanatory power.

Paternoster (2010) offered two explanations for this weak relationship. The first is that the “decision” to offend may not be that rational—that is, individuals do not (or only minimally) consider the costs and benefits of crime before offending. For example, high impulsivity—a strong correlate of crime—is defined as acting without consideration of the long-term consequences of behavior (Gottfredson and Hirschi, 1990; Piquero and Tibbetts, 1996; Wright et al., 2004). In more rigorous inquiries, this idea has been squarely challenged, finding that individuals are responsive to rational concerns when offending (Anwar and Loughran, 2011; Loughran et al., 2016; Matsueda, Kreager, and Huizinga, 2006; Piliavin et al., 1986). In the second explanation, it is acknowledged that individuals are rational but that *formal* sanction threats are of little import in this calculus. From this view, individuals weigh the benefits and costs of crime, but the weight placed on

arrest is small relative to considerations such as disappointing significant others, negative peer reactions, or potential monetary gain (Paternoster, 1987). Neither explanation is favorable to the deterrence perspective as they both raise serious questions about core tenets of the theory.

Before accepting such conclusions on the validity of deterrence theory, it is worthwhile to consider the assumptions that researchers have made when examining between-individual differences in the certainty–crime relationship. An inherent—yet often implicit—assumption made by researchers in these studies is that the reported risk perception  $p$  is internally accurate and has ratio meaning (i.e., the quantity can be interpreted literally). The former assumption maintains that individuals hold stable beliefs about arrest risk and accurately report their subjective expectations in a self-report survey, as well as that responses to a probabilistic question are determined only by underlying subjective beliefs. In the corollary ratio interpretation assumption, meaningful comparisons between individuals are possible. If individual A reports a 50 percent likelihood of arrest and individual B reports a 30 percent likelihood, then one assumes not only that individual A has a greater perception of risk but also that this risk is 20 percent higher than individual B's. In the assumption of ratio properties, a meaningful interpretation of the estimated regression coefficient of risk perceptions is possible (i.e., a 10 percent increase in risk perceptions is associated with a  $\beta$  decrease in crime). This set of conditions, which Pogarsky and Loughran (2016) referred to as the “strong form” interpretation of Nagin's (1998) idea regarding perceptions being “anchored in reality,” is a necessary condition required to estimate and interpret deterrent effects in between-individual analyses (Paternoster, 1987). But researchers in psychology and behavioral economics often find that these strong assumptions are untenable (Hurd, 2009; Manski, 2004; see also Pogarsky, Roche, and Pickett, 2017).

### ARBITRARINESS OF SUBJECTIVE PROBABILITY REPORTING

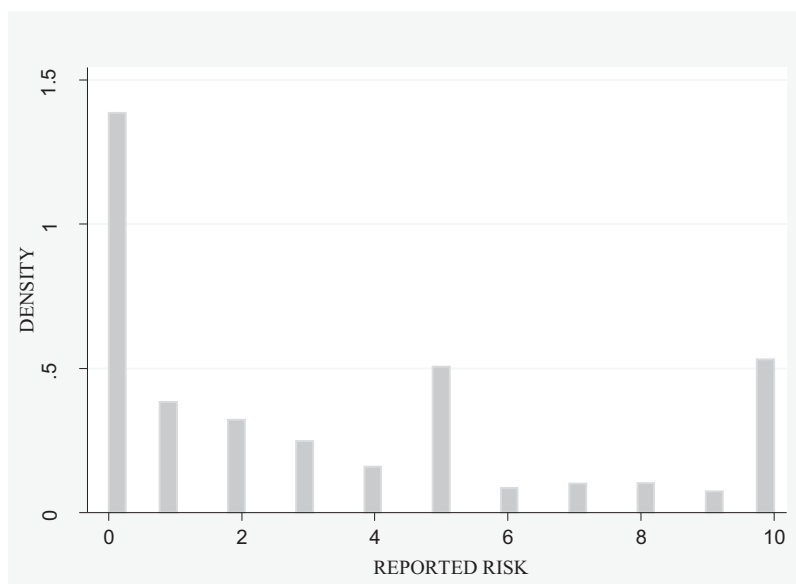
The results of survey research indicate that elicited probabilistic expectations are not as continuously distributed as one would expect if they represented the stable beliefs of respondents; rather, reported probabilities tend to cluster around focal response categories such as 0, 50, and 100 (Hurd, 2009; Manski and Molinari, 2010). Manski (2004) posited that the 50 percent option is extensively used because it incorporates both individuals who truly believe there is a .50 probability as well as individuals who are completely uncertain about the likelihood of the outcome (Bruine de Bruin et al., 2002). After all, 50 percent is the optimal response for individuals who only have confidence that the value is somewhere between 0 and 100. Hurd (2009) found that almost two thirds of respondents who reported a 50 percent chance of dying before 75 years of age reported they used that value because they were unsure of the actual probability. In cases of uncertainty, Tversky and Kahneman (1974) showed that even when reported probabilities do not fall directly on 50 percent, they tend to fall close to it. A consequence of this reporting habit is that true probabilities greater than 50 percent tend to get underestimated, whereas true probabilities less than 50 percent tend to get overestimated (Hurd, 2009). The 0 and 100 response options are also not inherently flawed, but many individuals report these categories for outcomes where such extreme certainties cannot literally be true (e.g., how likely is it that you will live to age 75; Hurd, 2009). In this case, individuals hold reasonably low or high subjective beliefs but

round toward the extreme ends when reporting (Manski and Molinari, 2010). There is growing consensus, then, that the clustering of responses around focal categories may not reflect the distribution of true subjective beliefs but are indicative of the precarious reporting habits of respondents (Hurd, 2009; Pickett, Loughran, and Bushway, 2015).

Additional concerns over the reporting of subjective expectations can be found in research findings indicating that subjective expectations are highly susceptible to *anchoring* effects (Wright and Anderson, 1989). Tversky and Kahneman (1974) showed that subjects anchor reported probabilities based on the presentation of both informative and noninformative values prior to answering probabilistic questions. Anchoring to “noninformative” values is most troubling because it suggests that reporting patterns can be affected by the presentation of information that is entirely unrelated to the question of interest. For example, Tversky and Kahneman (1974) were able to affect estimates of the percentage of African countries in the United Nations simply by spinning a wheel of fortune in the respondents’ presence. Similarly, Ariely, Loewenstein, and Prelec (2003) found that priming respondents with the last two digits of their Social Security number affected the reporting of subjective perceptions. Individuals whose last two digits were in the upper quintile (high anchor) were willing to pay a greater amount for material goods than were individuals with lower digits (low anchor). Even when presented with uninformative numbers, respondents often use these values as starting points to judge the probabilities of outcomes (however, see Fudenberg, Levine, and Maniadis, 2012, who question the existence of anchoring effects).

Researchers have also found that individuals anchor reported probabilities based simply on the order in which questions appear, such that the characteristics of the initial question affect reported values on subsequent questions (Andersson and Svensson, 2014; O’Hagan et al., 2006; Wright and Anderson, 1989). Question order effects have been observed in a wide range of subjective valuation and expectation studies, including the willingness to pay for health-care programs (Stewart et al., 2002) and perceived risk of developing cancer (Taylor et al., 2002). Importantly, Siminski (2008) observed that question ordering tends to bias response patterns in systematic ways. This may be particularly true for studies in which probabilistic expectations questions are relied on because, as noted, in cases of uncertainty, individuals tend to place initial probability responses closer to the 50th percentile (Hurd, 2009), and this initial value then acts as an anchor for subsequent questions. As a result, surveys beginning with items typically considered to be a low probability may upwardly bias a scale (when compared with surveys beginning with items that have a higher probability of occurrence) because individuals select a response option at or near the .50 mark, and this sets the lower bound when responding to the rest of the probabilistic questions. Taken together, these findings have led scholars to conclude that there is considerable *instability* in the reporting of subjective probabilities and that elicited values may be highly influenced by purely exogenous factors (e.g., anchors and question ordering).

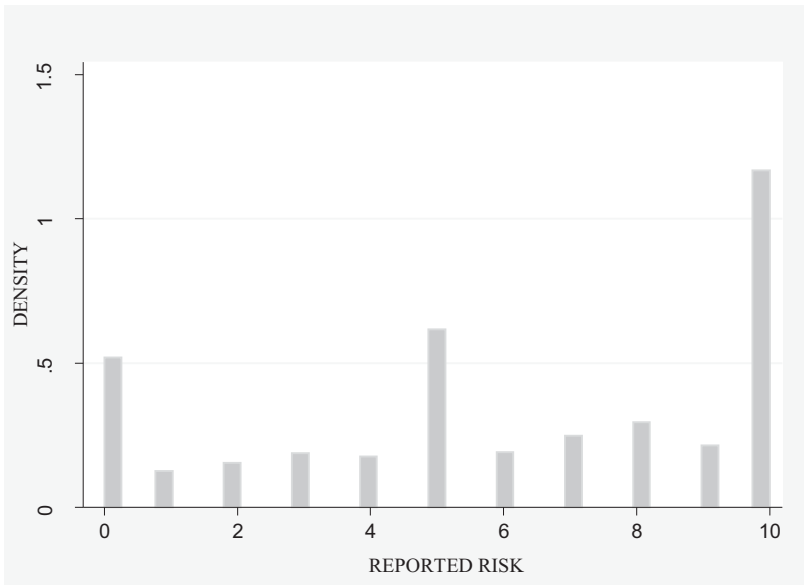
Although rarely tested directly, several concerns indicative of reporting instability have been observed in criminological studies of sanction risk perceptions. For example, the distribution of reported probabilistic expectations of arrest tends to cluster around focal responses. Figures 1 through 3 show the reported risks of arrest for fighting, breaking and entering, and auto theft among respondents at the baseline interview of the Pathways to

**Figure 1. Reported Risk of Arrest for Fighting**

Desistance Study (Mulvey et al., 2004). For each crime, the three most common responses correspond to the 0 percent, 50 percent, and 100 percent categories. Nevertheless, many criminologists continue to treat reported risk perceptions as stable beliefs that can be compared across individuals. In one of the few studies in which researchers have challenged this, Loughran, Paternoster, and Thomas (2014) tested the validity of reported sanction risk perceptions by using Bayesian truth serum (BTS; Prelec, 2009), an application of scoring rules in which respondents are incentivized to be honest and thoughtful in survey responses and problematic effects such as rounding, satisficing, and focal responses are supposed to be reduced (see Prelec, 2009).<sup>4</sup> Loughran, Paternoster, and Thomas (2014) found that the distribution of formal sanction risk perceptions in the group incentivized to be honest were significantly *lower* for driving drunk and stealing.<sup>5</sup> The conclusions offered by Loughran, Paternoster, and Thomas (2014) were grim, cautioning researchers from placing much stock in the utility of reported subjective probabilities in self-report surveys. Nevertheless, these conclusions were based on a strong form assumption of reported risk, which requires that elicited probabilities reflect actual values with ratio interpretation. Despite the considerable *arbitrariness* in reported subjective

- 
4. In BTS scoring, individuals are asked questions that require categorical (or yes/no) answers and are asked to provide a personal answer, as well as their prediction of the how they believe everyone else will respond. The individual's information score is then computed, based on giving answers that are "surprisingly common," that is, where the actual frequency exceeds their predicted frequency. Prelec (2004) showed that telling the truth represents the Bayesian Nash equilibrium strategy to maximize one's score and, hence, payoff.
  5. The authors reasoned that, when pressed to think about the risk of detection, respondents considered it to be much lower than individuals typically report.

**Figure 2. Reported Risk of Arrest for Breaking in**



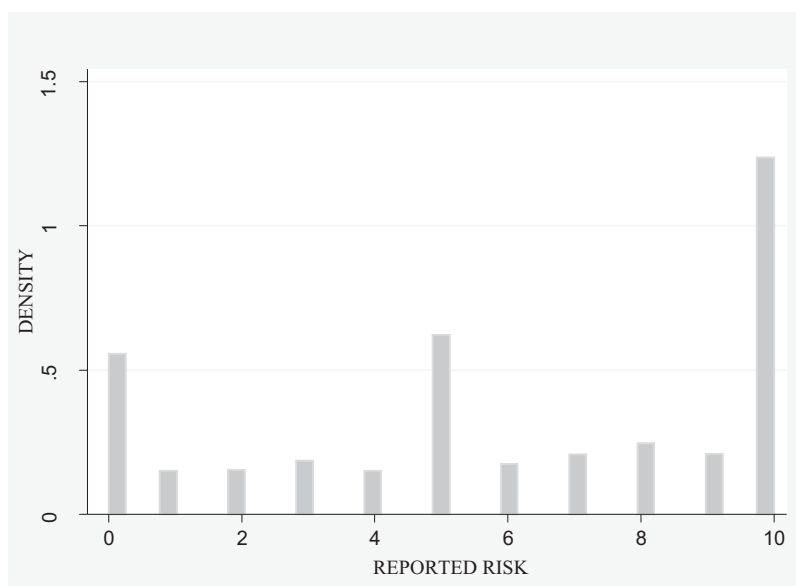
probabilities, there is evidence that survey answers are not entirely random—rather, there is predictable consistency in reported subjective probabilities.

COHERENCE IN SUBJECTIVE PROBABILITY REPORTING

Ariely, Loewenstein, and Prelec (2003) made an important observation when analyzing responses to subjective expectations. Although individuals are sensitive to anchoring effects in initial reported expectations, they are surprisingly rational in responses to subsequent questions—individuals tend to *rank order* subjective expectations in predictable ways. Specifically, Ariely, Loewenstein, and Prelec found that the absolute stated amount that individuals were willing to pay for products was affected by the presentation of anchors but that individuals within and across conditions consistently valued computer keyboards greater than Belgian chocolates or average wine. Similar results were found by Kahneman, Ritov, and Schkade (1999) in their study of dollar amounts issued in jury awards. Based on these observations, Ariely, Loewenstein, and Prelec (2003) concluded that the reporting of subjective expectations is simultaneously characterized by both randomness and internal consistency.

The apparent paradox of between-individual instability but within-individual stability is known as “coherent arbitrariness” (Ariely, Loewenstein, and Prelec, 2003), and it has important implications for understanding how individuals respond to subjective probability questions. Specifically, the rank ordering of responses as a means for determining the relativity of estimated probabilities can be applied to perceptual deterrence research in instances where perceived arrest risk is measured for multiple crime types in a single questionnaire. Although individuals may not be certain or report unstable estimates (or both) about the risk of arrest for using marijuana, they may nonetheless believe that using



**Figure 3. Reported Risk of Arrest for Auto Theft**

marijuana is less likely to result in an arrest than is committing a robbery, and subsequently, they will rate the risk associated with robbery as greater. This can be thought of as a weaker assumption on the interpretability of reported risk perceptions than is typically implicit in tests of deterrence theory—that is, that subjective probability distributions are absolute and directly comparable across persons. Pogarksy and Loughran (2016) referred to this type of internal consistency within the individual, but lack of absolute meaning between individuals, as the “weak form” of Nagin’s (1998) proposition about anchoring in reality.

Here we argue that the strong form assumption is *not* necessary for the key input in deterrence theory as it gives considerable confidence to the respondents’ reporting habits when responding to probabilistic questions. This does not mean that individuals do not hold underlying perceptions of arrest risk, only that these risks are difficult to elicit reliably in surveys (Manski, 2004). Whereas in the strong form assumption of the interpretability of risk perceptions within-individual stability in rank ordering is a necessary but not sufficient condition for valuing subjective probabilities, under the weak form assumption, rank ordering is a necessary *and* sufficient condition. Thus, our faith in respondents lay just in the ability to rank order crimes in a coherent way.

Deterrence, at its core, is a theory of information (Kennedy, 2009; Sherman, 1990), which is acquired through a communication process and subsequently affects behavior (Zimring and Hawkins, 1973). The weak form condition we outlined is informative and leads to falsifiable predictions about the risk–crime relationship in that it predicts that changes in perceived risk are ultimately linked to changes in offending behavior. In other words, risk perceptions in isolation should be less important than relative rankings *within persons*, be it across crime types or within-individual changes in risk perceptions of the

same crime over time (Ariely, Loewenstein, and Prelec, 2003).<sup>6</sup> Moreover, as we return to shortly, we believe that a study of the within-person ordinal properties of subjective risk beliefs holds great promise for the study of deterrence theory beyond merely concepts of measurement.

## CURRENT STUDY

In this study, we examine the assumption implicit in most deterrence tests that the elicited probabilities of arrest reflect stable underlying beliefs. By using three experimental designs, we test the effects of 1) informative anchors, 2) noninformative anchors, and 3) question ordering on responses to probabilistic assessments of arrest risk. If reported probabilities reflect the respondents' true perceptions of risk and if these values are stable, then anchoring should not impact responses. We hypothesize that responses to arrest risk questions are arbitrarily reported and will be impacted by the presentation of both informative and noninformative anchors. We also suspect that individuals will anchor responses based on the ordering of the crime types in which they are asked to report the risk of arrest. Specifically, we predict that beginning a survey with a crime that is typically considered to be "low" in the probability of arrest (e.g., smoking marijuana) will result in all crimes being systematically overestimated relative to surveys that begin with a crime that is often considered to be "high" in the probability of arrest (e.g., homicide). The rationale is that individuals have a tendency to report initial probabilistic expectations on or around the 50th percentile, so beginning a survey with a "low"-risk crime sets the lower bound of the risk of arrest closer to 50 percent (Hurd, 2009). Thus, we hypothesize as follows:

*Hypothesis 1:* Responses to probabilistic expectations of arrest will be affected by anchoring. Individuals exposed to a high anchor will report higher subjective probabilities compared with those exposed to a low anchor. Moreover, individuals whose survey begins with a crime often considered to be a "low" risk of arrest will report higher risk perceptions when compared with those whose survey begins with a perceived "high" risk crime.

We also predict that there is within-individual consistency in the rank ordering of crime types. Although the means and distributions of reported probabilities for crimes may differ across conditions as a result of anchoring and framing, we expect that the rank order of crimes within and across conditions will not differ. So, for instance, marijuana use will be perceived as having the lowest probability of arrest regardless of the presented anchor, and that the rank position of marijuana use will be stable across conditions. This leads to our second hypothesis:

*Hypothesis 2:* Reported risk of arrest will demonstrate "coherent arbitrariness" in that there will be within-individual stability in the rank ordering of crime types across conditions.

---

6. In Sherman's (1990) study, in which he found a deterrent effect of rotating police crackdowns—where the overall mean level of policing is constant but changes across time and location—a clear personification of this concept is in practice.

This second hypothesis can be considered to be an alternative to a third, and more problematic, possibility: That reported probabilities on the risk of arrest are simply arbitrary. This possibility would be supported if there is evidence of anchoring effects between conditions, but no evidence of rank order coherence (support for hypothesis 1 but not for hypothesis 2). If this is the case, then the reported subjective probabilities reflecting individuals' risk of arrests would essentially be random nonsense (i.e., unstable values and unstable ordering) and of little use for empirical studies. Nevertheless, if reported risk perceptions display coherent arbitrariness, then this would indicate that reported risk perceptions are "anchored in reality" and thus potentially meaningful for empirical tests of deterrence.

## DATA AND METHOD

### EXPERIMENT 1: INFORMATIVE ANCHOR

We begin by providing respondents with what we are calling an "informative anchor"—that is, an anchor related to the risk of arrest after committing a single crime: stealing clothes from a store. It is prudent to begin with an informative anchor for several reasons. First, it provides an initial examination of the effect that anchoring has on responses to subjective probability questions. Second, it allows us to examine how presenting a likelihood of arrest for a single crime affects survey responses to other crime types. The third and primary reason for using an informative anchor is that the results of prior research have shown that individuals are surprisingly arbitrary in survey responses *even when under full information*. That is, when individuals are told the objective value of a product, they still tend to report their subjective value arbitrarily (Ariely, Loewenstein, and Prelec, 2003). The findings from this research reveal that arbitrariness under full information is contrasted with internal coherence in the rank ordering of responses. In other words, individuals seem to be more concerned with maintaining rank order than with using the informative anchor to report a single item accurately. For the current study, using the informative anchor allows us to examine the portion of people who report a subjective likelihood of arrest for stealing clothes from a store that directly corresponds to the provided anchor, and we contrast this with the degree at which rank ordering of offense types affects reported probabilities, regardless of the information provided.

The informative anchor experiment relies on data from undergraduate students at a public university in the Midwest. The researchers visited five large introductory classes and explained that they were interested in college students' perceptions of the risk of arrest after engaging in criminal behavior. It was explained that all of the surveys would be anonymous and that students who were 18 years of age or older and had not already completed the survey in another class were eligible to participate. Respondents did not receive any incentive for participating and were given the first 10 minutes of class to complete the paper survey. Each survey began with a brief description of a fictional study said to have been conducted at the University of Maryland. Three versions of the fictional study make up the conditions of the experiment. Specifically, individuals received one of the following paragraphs at the top of the survey:

*A recent study by researchers at the University of Maryland showed that 32% of people who steal clothes from a store are caught and arrested for their crime. The authors raised questions about how risky people think it is to commit crime.*

*A recent study by researchers at the University of Maryland showed that 71% of people who steal clothes from a store are caught and arrested for their crime. The authors raised questions about how risky people think it is to commit crime.*

*A recent study by researchers at the University of Maryland examined the likelihood of arrest after stealing clothes from a store. The authors raised questions about how risky people think it is to commit crime.*

The surveys were ordered with use of a random number generator to eliminate concerns of selection bias and were then distributed to students in the classes. Respondents were separated into one of three conditions: 1) low anchor (32 percent), 2) high anchor (71 percent), and no anchor. One hundred and fifty-five students completed the survey. On average, the sample was 22 years of age and 48 percent male. Furthermore, the sample was 62 percent White, 21 percent African American, 11 percent Hispanic, 2 percent Asian, and 5 percent self-identified as "Other." There were 54 students who completed the low-anchor survey, 51 students completed the high-anchor survey, and 50 students completed the no-anchor survey. Descriptive statistics are presented in table 1.

After reading the description of the fictional study, participants were asked to report their perceived risk of arrest for six different crimes: robbery, marijuana use, breaking into someone's home, stealing clothes from a store, driving drunk, and auto theft.<sup>7</sup> The response options were on a 12-point scale where the first option corresponded to 0 percent, the last option to 100 percent, and the middle options all increased in 10 percent increments (1–10 percent, 11–20 percent, 21–30 percent, etc.). We converted these values into probabilities using midpoints for the middle options, such that 0 = .00, 1–10 percent = .05, 11–20 percent = .15, ..., 100 percent = 1.00. The survey contained measures intended to capture the background characteristics of respondents (race, gender, age, etc.), impulsivity (do participants tend to think before they act, plan for the future, etc.), and prior offending (cheating on exams, using marijuana, driving drunk), which were included to assess the effectiveness of the randomization process. We assessed balance between conditions by regressing treatment status on the observed covariates by using a multinomial logit and examining the overall model fit. The null hypothesis is that the observed coefficients of the covariates are jointly equal to zero, which would indicate that the treatment conditions do not significantly differ on the covariates that we observe. Indeed, the results (presented in table 1) indicated that we fail to reject the null hypothesis, suggesting that there is balance between treatment conditions ( $\chi^2 = 8.860$ , degrees of freedom [d.f.] = 14,  $p = .840$ ).<sup>8</sup>

7. The ordering of crime-types was the same across all conditions.

8. We also examined whether there was equality between groups for each covariate individually by using an ANOVA. These results are presented in the appendix in the online supporting information. (Additional supporting information can be found in the listing for this article in the Wiley Online Library at <http://onlinelibrary.wiley.com/doi/10.1111/crim.2018.56.issue-1/issuetoc>.) Although not significantly different ( $p = .14$ ), the "no-anchor" condition had a slightly lesser tendency to use marijuana in the past compared with the other conditions. As a sensitivity check, we examined anchoring effects while adjusting for prior marijuana use and the findings are identical to those presented in the text. These results are available by request.

**Table 1. Summary Statistics for Each Experimental Study**

Variable	Experiment 1 (Informative Anchor)	Experiment 2 (Noninformative Anchor)	Experiment 3 (Question Ordering)
Age	21.680 (4.673)	21.741 (4.630)	23.239 (7.152)
Male	.481 (—)	.426 (—)	.452 (—)
White	.618 (—)	.707 (—)	.587 (—)
Black	.211 (—)	.153 (—)	.258 (—)
Hispanic	.105 (—)	.076 (—)	.071 (—)
Asian	.020 (—)	.006 (—)	.019 (—)
Other Race	.046 (—)	.057 (—)	.065 (—)
Impulsivity	1.912 (.593)	1.851 (.576)	1.827 (.570)
Prior Cheating	.176 (—)	.178 (—)	.167 (—)
Prior Marijuana Use	.307 (—)	.310 (—)	.335 (—)
Prior Drunk Driving	.211 (—)	.184 (—)	.258 (—)
Joint-significance $\chi^2$ Test (d.f. = 14) $p > \chi^2$	$\chi^2 = 8.860$ .840	$\chi^2 = 14.310$ .427	$\chi^2 = 15.360$ .354
N	155	158	156

NOTES: Standard deviations are reported in parentheses; joint-significance  $\chi^2$  test reflects whether the observed covariates predict treatment assignment for each experiment using a multinomial logit, with the null hypothesis being that the coefficients in the model are jointly equal to zero.

**Table 2. Average Reported Risk Perceptions by Treatment Group for Clearance Rate Study (N = 155)**

Variable	No Anchor	Low Anchor	High Anchor	Kruskal–Wallis Test
Robbery	.643	.504	.660	$p = .002$
Marijuana Use	.249	.138	.252	$p = .014$
Break In	.586	.472	.598	$p = .020$
Steal Clothes	.470	.372	.576	$p = .000$
Drive Drunk	.478	.375	.505	$p = .055$
Steal Car	.717	.634	.800	$p = .000$
Average Score	.523	.416	.568	$p = .000$
N	50	54	51	—

RESULTS

Table 2 presents the mean elicited subjective probability values across conditions for each crime, as well as the average composite score of the items. The final column presents the results of Kruskal–Wallis tests—a nonparametric analog of an analysis of variance (ANOVA) for noncontinuous data—testing whether the differences across the groups

are statistically significant. The mean elicited responses in the high-anchor condition were .660 for robbery, .252 for marijuana, .598 for breaking into a building, .576 for stealing clothes from a store, .505 for driving drunk, and .800 for stealing a car. The corresponding elicited probabilities in the low-anchor condition are .504, .138, .472, .372, .375, and .634, respectively. For the no-anchor condition, the corresponding probabilities are .643, .249, .586, .470, .478, and .717, respectively. For five of the six behaviors, and the average score, the results of the Kruskal–Wallis test indicate that the differences across conditions in the reported probability of arrest are statistically significant at a 5 percent level. The sixth behavior (drunk driving) is not quite significant at  $\alpha = .05$  ( $p = .055$ ). These differences are systematic and in the direction one would expect if there are anchoring effects in the reporting of subjective probabilities: The high-anchor group reports a higher probability of arrest than does the low-anchor group for all crime types. On average and for most crimes, individuals in the high-anchor condition report the risk of arrest as about 15 percentage points higher than the low anchor group.

Given that the provided anchor describes the likelihood of arrest for stealing clothes from a store, we cannot rule out the possibility that the differences are a result of individuals rationally updating subjective beliefs using Bayes rule (Anwar and Loughran, 2011). Nevertheless, there are reasons to suspect that the observed differences are not—at least entirely—a result of updating. First, the differences across conditions are consistent for every crime type and not limited to stealing clothes from a store. Prior study findings have revealed that risk updating is crime specific (Anwar and Loughran, 2011), which challenges the notion that the effects are simply a result of updating. Second, if individuals are rationally updating, we would expect that the reported risk of arrest for stealing clothes would correspond to the provided anchor (e.g., individuals given the 32 percent anchor would endorse the category where 32 percent falls). The data indicate that this is not the case. Just around a quarter of respondents in both the 32 percent anchor condition (24 percent) and the 71 percent anchor conditions (27 percent) selected the response category corresponding to those anchors. Furthermore, slightly more than 50 percent in either condition even reported the risk of arrest for stealing clothes as being within one adjacent category of the provided anchor (51 percent in the high-anchor condition and 54 percent in the low-anchor condition). In other words, even under “full information,” nearly half of the respondents report a risk of arrest that is at least 20 percentage points away from the information they were provided. Taken together, the results indicate that individuals are arbitrary in reporting the risk of arrest for various different crime types.

Next, we consider if, despite the observed anchoring, individuals still rank order subjective beliefs about arrest risk in an internally coherent manner. To examine the internal consistency of crime ordering, we create a rank score of each crime type for respondents using the *rowranks* command in Stata. For each subject, this command uses the reported arrest risks for the six crimes to determine each individual’s rank ordering of crime from lowest perceived risk (rank = 1) to highest perceived risk (rank = 6). For instance, consider a hypothetical individual who reports the subjective risks of arrest as marijuana = .05 (rank = 1), drunk driving = .15 (rank = 2), stealing clothes = .25 (rank = 3), breaking and entering = .35 (rank = 4), robbery = .45 (rank = 5), and auto theft = .55 (rank = 6). The rank score indicates this individual perceived marijuana use to be the least risky crime and

**Table 3. Average Rank of Crimes by Treatment Group for Clearance Rate Study ( $N = 155$ )**

Variable	No Anchor	Low Anchor	High Anchor	Kruskal–Wallis Test
Marijuana	1.260	1.407	1.451	$p = .905$
Drive Drunk	2.720	2.851	2.667	$p = .912$
Steal Clothes	2.840	2.944	3.137	$p = .583$
Break In	3.460	3.463	3.235	$p = .524$
Robbery	3.840	3.852	3.843	$p = .996$
Steal Car	4.720	4.870	4.824	$p = .812$
$N$	50	54	51	—

auto theft to be the most risky.<sup>9</sup> Now consider another hypothetical individual who reports the subjective arrest risks as marijuana = .45 (rank = 1), drunk driving = .55 (rank = 2), stealing clothes = .65 (rank = 3), breaking and entering = .75 (rank = 4), robbery = .85 (rank = 5), and auto theft = .95 (rank = 6). Notice that when comparing these two individuals that, whereas the absolute reported probabilities differ, the rank values of each crime type are the same. Assigning rank values allows us to examine whether the relative position of arrest risk for each crime type differs between experimental conditions. By using a series of Kruskal–Wallis tests, we test the null hypothesis that, in the example of using marijuana:

$$H_0 : \text{Rank}_{\text{low anchor}}(\text{Marijuana}) = \text{Rank}_{\text{high anchor}}(\text{Marijuana}) = \text{Rank}_{\text{no anchor}}(\text{Marijuana})$$

Rejecting this null hypothesis would provide evidence that reported risk perceptions are unstable in the order of crime types. If, however, we fail to reject the null hypotheses, it would suggest that crime types are ordered similarly across conditions—i.e., coherent arbitrariness.

Table 3 presents the average rank position for each crime type, as well as the results of Kruskal–Wallis tests used to examine whether the rank positions of crime types differ between conditions. The results indicate that the average rank order of the crime types displays similar patterns across all conditions, with marijuana, on average, being ranked the least risky of the crimes, and auto theft the most risky of the crimes. In other words, *individuals tended to rank marijuana the least risky crime and auto theft the most risky regardless of the stated point estimates of risks and any information provided to them.* Moreover, the average rank scores across conditions for each crime type are remarkably similar, which demonstrates no observable difference in rank for the same crime between conditions (e.g., for marijuana no anchor = 1.26, low anchor = 1.41, and high anchor = 1.45). The Kruskal–Wallis tests indicate that we fail to reject the null hypotheses that the rank orders of the crime types do not vary across conditions, with the  $p$  values ranging from .52 and .99. That we can retain the null hypothesis suggests to us that rank order of crime arrest risk is consistent across conditions, and that although the absolute values

9. When individuals report the same arrest risk for two crimes, the crimes are given the same rank  $k$ , and the next lowest crime receives a rank score of  $k + 2$ . For example, an individual reporting the risks of arrest as marijuana = .05, drunk driving = .15, stealing clothes = .15, breaking and entering = .35, robbery = .45, and auto theft = .55 would receive the respective ranks of 1, 2, 2, 4, 5, and 6. Similar logic applies when three or more crimes are reported as having the same arrest risk. The nonparametric test we employ allows for such ties.

reported by respondents are affected by the presentation of anchors, the rank order of arrest risk seems mostly unaffected.

The results of the informative anchor experiment provide evidence that reported risk perceptions display “coherent arbitrariness.” Reported perceptions of risk are arbitrary in the sense that the absolute values reported by respondents are susceptible to the presentation of an anchor. Even when individuals were provided “full information” on the risk of arrest for stealing clothes from a store (the anchor), most did not select the corresponding probability when responding to the question asking the risk of arrest for that crime. The finding that individuals’ reported beliefs are arbitrary even under complete information can be explained by the fact that individuals maintain internal consistency in the rank ordering of crimes. Individuals seem arbitrary in their reported risk of arrest for stealing clothes from a store when presented with complete information because they are more concerned with maintaining internal consistency in their subjective rank ordering of crime risk than with accurately eliciting a single probability.

## EXPERIMENT 2: NONINFORMATIVE ANCHOR

Experiment 2 is aimed at examining anchoring effects using a “risky” behavior that is unrelated to crime and the criminal justice system. If the results of the first experiment were simply a result of individuals updating their subjective beliefs of arrest risk, then the presentation of irrelevant information should not impact survey responses. Respondents were presented with a study similar to the one presented in experiment 1, but rather than presenting the risk of arrest after stealing clothes from a store, respondents were presented an estimate of the percentage of college students who contract a sexually transmitted infection (STI) prior to graduating. Each respondent received one of the following paragraphs at the top of their questionnaire:

*A recent study by researchers at the University of Maryland showed that 32% of college students contract a sexually transmitted disease before graduating. The findings raised questions about how risky people think certain behaviors are.*

*A recent study by researchers at the University of Maryland showed that 71% of college students contract a sexually transmitted disease before graduating. The findings raised questions about how risky people think certain behaviors are.*

*A recent study by researchers at the University of Maryland examined the percentage of college students who contracted a sexually transmitted disease. The findings raised questions about how risky people think certain behaviors are.*

A total of 158 undergraduates attending five large classes at the same public university in the Midwest took part in the noninformative anchor experiment. The individuals who took part in this experiment are from different classes and are unique individuals from those in the informative anchor experiment. The description of the study verbalized by the researchers and the consent form given to students were identical to the informative anchor study, and students were informed to check “I have already had the opportunity to take the survey” on the consent form and not fill it out if they had been in a class where a survey was already administered. The survey procedures were identical to the informative anchor experiment (e.g., the randomization process and no incentives). Moreover, the surveys in the noninformative experiment—the crime types, measures, and coding of



**Table 4. Average Reported Risk Perceptions by Treatment Group for STI Study (*N* = 158)**

Variable	No Anchor	Low Anchor	High Anchor	Kruskal–Wallis Test
Robbery	.621	.565	.744	<i>p</i> = .000
Marijuana Use	.203	.140	.255	<i>p</i> = .024
Break In	.614	.528	.665	<i>p</i> = .018
Steal Clothes	.483	.450	.578	<i>p</i> = .033
Drive Drunk	.469	.381	.494	<i>p</i> = .087
Steal Car	.725	.643	.819	<i>p</i> = .000
Average Score	.519	.451	.594	<i>p</i> = .002
<i>N</i>	56	51	51	—

perceived arrest risk—were the same as in the informative experiment with the sole exception that in the fictional study portrayed in the former, the risk of contracting a STI rather than the risk of arrest for stealing clothes is described. There were 56 participants in the no-anchor, 51 in the low-anchor, and 51 in the high-anchor condition. The sample was 71 percent White, 15 percent Black, 8 percent Hispanic, .6 percent Asian, and 6 percent identified as “Other” (table 1). We examined the effectiveness of the randomization process by regressing the treatment condition on the observed covariates (table 1), and we were unable to reject the null hypothesis that the coefficients of the covariates were jointly equal to zero ( $\chi^2 = 14.310$ , d.f. = 14, *p* = .427), which suggests that the randomization process effectively achieved balance.<sup>10</sup>

## RESULTS

Table 4 presents the mean responses for six crime types and the average composite score for each condition, and the last column presents the results of a Kruskal–Wallis test examining whether observed differences are statistically significant. The results are highly congruent to those in the previous experiment. In the high-anchor condition, reported probabilities of arrest are .744 for robbery, .255 for marijuana, .665 for breaking into a building, .578 for stealing clothes from a store, .494 for driving drunk, and .819 for stealing a car. The corresponding reported probabilities in the low-anchor condition are .565, .140, .528, .450, .381, and .643, respectively. In the no-anchor condition, the reported probabilities are .621, .203, .614, .483, .469, and .725, respectively. The results of a series of Kruskal–Wallis tests indicate that these differences in the reported risk of arrest across conditions are significant at a 5 percent level for five of the six crime types, as well as for the composite average risk perception scale. As with the informative anchor experiment, drunk driving is not significantly different across groups at  $\alpha = .05$  (*p* = .087). Respondents in the high-anchor condition report a higher risk of arrest for all crimes when compared with those in the low-anchor condition, with the differences in the reported risk of arrest between these two conditions being approximately 15 to 20 percentage points

10. The appendix in the online supporting information includes the results of ANOVAs examining the equality of the covariates across conditions independently. None of the covariates differed across conditions at a *p* < .05 level, but subjects in the low-anchor condition were slightly older (*p* = .09). To assess the robustness further, we reexamined all of the results controlling for age, and the findings were substantively the same. These results are available by request.

**Table 5. Average Rank of Crimes by Treatment Group for STI Study (N = 158)**

Variable	No Anchor	Low Anchor	High Anchor	Kruskal–Wallis Test
Marijuana	1.250	1.235	1.255	$p = .956$
Drive Drunk	2.803	2.549	2.411	$p = .272$
Steal Clothes	3.017	3.117	3.039	$p = .879$
Break In	3.696	3.431	3.333	$p = .382$
Robbery	3.730	3.922	3.961	$p = .651$
Steal Car	4.339	4.927	4.627	$p = .392$
N	56	51	51	—

for each crime, and for the average scale. For example, whereas individuals in the high-anchor condition report the risk of arrest at 82 percent for stealing a car, individuals in the low-anchor condition report the risk of arrest for auto theft at 64 percent. These anchoring effects are large given that the information provided to respondents dealt with the probability of contracting a STI, had nothing to do with the crime or the criminal justice system, and are difficult to account for from any rational choice perspective that considers elicited probabilities to be meaningful values.<sup>11</sup>

We tested for rank ordering in the reporting of subjective risks of arrest for crimes across conditions in a similar manner as the previous experiment. The results are reported in table 5. The average rank order of crime types is the same across all conditions with marijuana use rated as having the lowest risk of arrest and auto theft rated as having the highest risk of arrest. This pattern is the same as what we observed in the experiment with the informative anchor. Moreover, table 5 indicates that for each crime type, there is considerable similarity in the average rank positions. For example, the average rank position of robbery arrest in the no-anchor, low-anchor, and high-anchor conditions are 3.73, 3.92, and 3.96, respectively. Table 5 also provides the results of Kruskal–Wallis tests used to examine whether the rank positions of each crime type significantly differ across experimental conditions. For each comparison, we fail to reject the null hypothesis that the rank position of a crime in one condition does not significantly differ from the rank position of that same crime in the other conditions, with the  $p$  values ranging from .27 to .96. In other words, each crime type displays a similar rank position in the risk of arrest regardless of the presentation of an anchor.

The findings of this experiment are arguably more revealing than those in experiment 1 as the provided anchor had nothing to do with crime, the criminal justice system, or even the likelihood of informal apprehension. This challenges the assumption that reported risk perceptions represent stable subjective beliefs about the likelihood of arrest. Nevertheless, there was also evidence of consistent coherence in the rank ordering of crime types across conditions. Put simply, although reported risks of arrest are responsive to noninformative anchors, they also demonstrate impressive internal consistency in the rank ordering of the crime types.

11. Interestingly, the size of the discrepancy between the low and high anchors is almost identical to those observed in experiment 1. If what we observed in experiment 1 was merely an updating, rather than an anchoring, effect, then we would expect to see a difference in the results between these two studies.

## EXPERIMENT 3: ORDERING OF SURVEY ITEMS

The last experiment was used to examine the degree to which the ordering of survey items affects the reported risks of arrest, and the extent to which responses to initial items can serve as an “anchor” to subsequent items. Prior research findings have shown that when presented with questions eliciting probabilistic expectations, individuals gravitate toward the 50th percentile (or “middle point” on an interval/ratio scale) for the initial item, and then rank order the subsequent items. When risk perception scales begin with a crime type in which respondents usually report a high risk of arrest (e.g., homicide), the reported probabilities for that crime may be lower than would be the case if that item came later in the scale—that is, in surveys where the risk perception scale begins with homicide, individuals will systematically report the risk of arrest closer to the 50th percentile, whereas in surveys where it comes later, they are more likely to report it closer to 100 percent. Conversely, when a scale begins with a crime that is usually lower in the reported risk of arrest (e.g., marijuana use), the elicited risk perception may be higher than if that item came later. By extension, initially reported probabilities serve to “anchor” responses to subsequent survey items, which consequently impact overall reporting (toward either the 100 or 0 categories). In this instance, item distributions that begin with behaviors that are typically reported as “more risky” will serve to anchor responses toward the lower end (the 0 percent category), whereas initial items pertaining to typically reported “less risky” behaviors will anchor overall reporting to the higher end (the 100 percent category). Furthermore, if individuals coherently rank crimes, then this tendency would affect the other items in the scale as well as the initial reported probability for the high/low-risk crimes set a bound at which all other crimes must be greater or less than.

We use three experimental conditions. Each condition asks respondents their perceived risk of arrest for the same seven crimes, but the conditions differ in the order in which the crimes were presented. In the first condition, the survey begins with a crime often observed as having the highest perceived risk of arrest: homicide. Specifically, the order of the crimes when assessing risk perceptions in the first condition is homicide, robbery, driving drunk, stealing clothes from a store, breaking into someone’s home, auto theft, and using marijuana. The second condition begins with a common crime where individuals generally believe the risk of arrest is low, with the order of risk perception questions being marijuana, robbery, driving drunk, stealing clothes from a store, breaking into someone’s home, auto theft, and homicide. Thus, in the first two conditions, the location of the first and last perceived risk of arrest questions is switched but the remaining crimes are in the same order. This switch allows for a direct assessment of how beginning surveys of probabilistic expectations with low-risk versus high-risk crimes affects the pattern of survey responses. In the third and final condition, the crimes are more randomly ordered and begin with a crime in which the reported risk of arrest does not tend to be so extremely high or low, as is the case for homicide and marijuana use. The order of the risk perception questions by crime types for this condition is stealing clothes from a store, breaking into someone’s home, driving drunk, robbing someone with a gun, stealing a car, marijuana use, and homicide. No numerical anchors are presented in this experiment; rather, the purpose of the study is to examine whether reported probabilities can be affected simply by the order of the questions. The coding of the risk perception items is identical to the first two experiments.

**Table 6. Average Reported Risk Perceptions by Treatment Group for Crime Order Study (N = 156)**

Variable	Marijuana-First	Homicide-First	Theft-First	Kruskal–Wallis Test
Robbery	.699	.591	.620	$p = .080$
Marijuana Use	.257	.153	.163	$p = .001$
Break In	.616	.481	.525	$p = .012$
Steal Clothes	.450	.395	.454	$p = .474$
Drive Drunk	.427	.366	.425	$p = .228$
Steal Car	.723	.561	.633	$p = .010$
Homicide	.865	.684	.777	$p = .000$
Average Score	.577	.461	.514	$p = .002$
N	51	54	51	—

Undergraduate students attending six classes at a public university in the Midwest were recruited to partake in a survey. The procedures were identical to the previous experiments (e.g., the randomization process and no incentives), and respondents were told not to participate if they had taken a survey from the researchers in another class. Overall, 156 participants completed a survey, with the sample being 59 percent White, 26 percent Black, 7 percent Hispanic, 2 percent Asian, and 6.5 percent self-identifying as “Other.” On average, the sample is 23 years of age and 45 percent male. A total of 54 participants received the survey beginning with the risk of arrest for homicide, 51 received the survey beginning with the risk of arrest for marijuana use, and 51 received the survey beginning with stealing clothes from a store. When we regressed treatment status on the observed covariates, we were unable to reject the null hypothesis that the model parameters were jointly equal to zero ( $\chi^2 = 15.360$ , d.f. = 14,  $p = .354$ ), indicating that there was balance across conditions in demographics, prior offending, and impulsivity (see table 1).<sup>12</sup>

RESULTS

The results examining the effect that differentially ordering the crime types has on reported risk of arrest are presented in table 6, which shows the mean elicited probabilities across conditions and the results of the Kruskal–Wallis tests assessing the significance of these differences. The findings indicate that simply altering the order of the crime types affects survey responses. In the conditions that begin with the risk of arrest for marijuana use, the reported probabilities of arrest are .699 for robbery, .257 for using marijuana, .616 for breaking into a building, .450 for stealing clothes from a store, .427 for driving drunk, .723 for stealing a car, and .865 for homicide. The corresponding reported probabilities of arrest when beginning with homicide are .591, .153, .481, .395, .366, .561, and .684. Recall that the ordering of the “marijuana first” and “homicide first” conditions were the same with the exception of the first and last crime types. In the condition in which the

12. The differences across conditions in the covariates independently are presented in the appendix in the online supporting information. The marijuana first condition contained slightly fewer males when compared with the other conditions, although the differences were not significant at a  $p < .05$  level ( $p = .09$ ). As a robustness check, we examined differences in the reported arrest risk across conditions while controlling for gender, and the findings were identical to those presented in text. These results are available by request.

**Table 7. Average Rank of Crimes by Treatment Group for Crime Order Study ( $N = 156$ )**

Variable	Marijuana-First	Homicide-First	Theft-First	Kruskal-Wallis Test
Marijuana	1.451	1.260	1.255	$p = .361$
Drive Drunk	2.431	2.722	3.000	$p = .166$
Steal Clothes	2.706	2.944	3.033	$p = .705$
Break In	3.941	3.518	3.588	$p = .131$
Robbery	4.392	4.574	4.255	$p = .496$
Steal Car	4.706	4.444	4.529	$p = .753$
Homicide	6.078	5.158	5.784	$p = .116$
$N$	51	54	51	—

crime types were more randomly ordered, the reported risks of arrest are .620 for robbery, .163 for marijuana use, .525 for breaking into a building, .454 for stealing clothes from a store, .425 for driving drunk, .633 for auto theft, and .777 for homicide. The results of the Kruskal-Wallis tests indicate that differences across conditions are significant for four of the seven crime types at  $p < .05$ . There are also significant differences across conditions in the composite average risk perception score ( $p = .002$ ).

The observed differences across conditions are consistent with the findings of prior research indicating that individuals tend to place the initial probabilistic expectations questions closer to the .50 response category and, as a result, overestimate “low” probabilities and underestimate “high” probabilities. This is precisely what is observed in the current data. The “marijuana first” condition reveals higher risks of arrest for marijuana (i.e., closer to .50) and for all subsequent crime types when compared with the “homicide first” condition. Similarly, the “homicide first” condition reveals a substantially lower risk of arrest for homicide (i.e., closer to .50) and all subsequent crime types when compared with the “marijuana first” condition. Thus, the results of this experiment further call into question whether treating reported risk of arrest as a “true” value with ratio interpretation is a tenable assumption.

Table 7 reports the average rank positions of each crime type across conditions, and the results of Kruskal-Wallis tests assessing whether any differences are statistically significant. With all but one exception, the results indicate that the rank order of crimes is consistent across conditions. The exception is that in the homicide first condition, robbery, on average, has a slightly higher arrest risk rank than does auto theft (4.57 vs. 4.44), but in the other conditions, auto theft has a higher average rank score. Nevertheless, the average rank scores for each crime type are similar across conditions. For example, the average rank positions for robbery arrest risk are 4.39, 4.57, and 4.26 for the marijuana first, homicide first, and random order conditions, respectively. The Kruskal-Wallis tests fail to reject the null hypotheses that each crime type has the same rank position across conditions, with the  $p$  values ranging from a low of .12 to a high of .75. This indicates that although the order in which questions appear may affect the absolute risk of arrest that individuals report for a crime, it does not affect the rank position of the crime’s arrest risk.

In summary, the results of the question-ordering experiment indicate that simply altering the order in which crime types appear on an arrest risk scale can affect reported probabilities. Beginning with a crime with a low risk of arrest (e.g., marijuana use) tends

to result in individuals overestimating the risk of arrest for all crime types compared with when a scale begins with a crime with a high risk of arrest (e.g., homicide). This result may occur because individuals have a tendency to report initial probabilities closer to the .50 response category. Once this initial value is selected, respondents in each condition rank order crimes in internally consistent ways, such that marijuana is, on average, reported as having the lowest risk of arrest and homicide, on average, the highest. Taken together, the results corroborate the findings in the previous two experiments: Reported perceptions of arrest risk display “coherent arbitrariness.”

#### SUPPLEMENTARY ANALYSES OF RANK ORDER TESTS

The results of the experiments indicated that the presentation of anchors and question ordering affected the absolute values of reported arrest risk perceptions but not the order at which crimes were ranked in detection risk. Some readers may be concerned about the power of the tests detecting rank-order differences (i.e., we are merely making a type II error). With that said, a sufficiently powerful test should be able to detect differences in rank for *different* crimes between conditions—e.g., we can test whether the rank position of arrest risk for marijuana use in the low-anchor condition significantly differs from the rank position of arrest risk for drunk driving in the high-anchor condition by using a *t* test. We present the results of such tests for all three experiments in the appendix in the online supporting information. The findings indicate that, with a few exceptions, for the pairwise comparisons, we can reject the null hypothesis that the rank position of two different crimes is the same across conditions at a 5 percent level. For example, the rank of marijuana in the low-anchor condition is significantly lower than the rank of all other crimes in either the high- or no-anchor conditions. These findings provide support to the conclusion that the rank ordering of crime types is stable across conditions as it indicates not only that the rank of a specific crime does not differ across conditions but also that that specific crime does significantly differ from the rank of *other crimes* across conditions. Put simply, the supplementary analyses lend considerable confidence to the finding that there is internal stability in the rank ordering of crime arrest risk regardless of the instability in the stated absolute value of subjective beliefs

#### DISCUSSION

In deterrence theory, individuals are responsive to sanction threats. This straightforward idea has invoked some of the field’s most heated debates, mostly stemming from the considerable disagreement over the theoretical meaning of a risk–crime relationship (Braga and Apel, 2016; Kleck, 2016; Nagin, 2016; Pickett and Roche, 2016; Pogarsky and Loughran, 2016). In most tests of deterrence, researchers place a literal meaning on the elicited risk perceptions and, in turn, on the nature of the relationship between arrest risks and offending: Reported probabilities reflect stable beliefs, and a 10 percent increase in risk must be associated with a certain reduction in offending for deterrence to be supported (Loughran, Paternoster, and Weiss, 2012). On its face, such a literal interpretation may seem like a necessary condition for sanction risk perceptions to be a meaningful construct and for deterrence to be falsifiable. In this article, we provide the results of empirical tests that directly challenge the necessity of this interpretation for deterrence to work as hypothesized.

By using three experiments, we found evidence that reported perceptions of arrest risk display “coherent arbitrariness.” Elicited risk perceptions are arbitrary in that they are susceptible to exogenous factors such as anchors (informative and noninformative) and question ordering, which likely means that there is instability in the reporting of probabilistic expectations. When considered on their own, these findings raise questions about the utility of deterrence as a theory as it would suggest that one of the theory’s key constructs may not be “anchored in reality” (Nagin, 1998) and is instead a random response to survey items. This interpretation is challenged by the finding that individuals are locally coherent in the rank order of arrest risk among different crime types. Providing individuals with a high anchor tends to shift one’s subjective probability distribution toward 1, and a low anchor toward 0, but these distributions display impressive consistency in the rank ordering of crimes. Therefore, elicited subjective risk perceptions are likely meaningful values *within persons* and thus anchored in reality (Pogarsky and Loughran, 2016).

To be clear, we believe that individuals do hold subjective beliefs on the likelihood of arrest when committing crime but also that these risks are difficult to elicit in surveys and may be susceptible to exogenous factors that are unrelated to the respondents’ subjective beliefs. We also believe that there are between-individual differences in the perceived risk of arrest, but the ability to assess such differences is constrained by the arbitrary nature in which initial values are selected. The findings of the current study lead us to challenge the assumption that individuals’ reported arrest risk all fall on the same subjective probability distribution. A more tenable assumption is that each individual has his or her own subjective distribution of arrest risk for offending, and that these distributions may be difficult to compare across persons (Pogarsky and Loughran, 2016).

Thus, the implication of coherent arbitrariness is not that deterrence theory cannot be formally tested. Indeed, the findings presented here offer an exciting within-individual research agenda for deterrence scholars. If we believe that risk perceptions are internally coherent within each person and that individuals have their own subjective distribution of risk, then at any given time-point, an individual would have an overall mean level of perceived risk that falls somewhere on that distribution. Around this overall mean fall the risks of arrest for specific types of crime as is evident by the fact that individuals rank order crimes in their arrest risks—i.e., individuals tend to perceive some crimes as more risky than others. *On average*, there was consistency within and between conditions in the rank ordering of crime types, but there was also variation across people within each condition in how crimes were ordered: Some individuals consider drunk driving to be the least risky, and others consider driving drunk to be the most risky. In deterrence theory, at a given time point, individuals should have a greater probability of committing crimes that they consider to be the least likely to result in arrest, as well as have the lowest probability of committing crimes that they consider to have the highest likelihood of arrest. This is analogous to arguing that risk perceptions should be related to one’s offending repertoire and can be tested empirically by relating risk perceptions to statistical models designed to detect specialization in offending (Osgood and Schreck, 2007).

In deterrence theory, it is predicted that the mean subjective arrest risk should change over time—particularly in response to new information—and that within-individual changes in perceived risk of arrest should predict within-individual changes in behavior

(see Pogarsky, 2007). There is now considerable evidence that getting arrested for a crime increases the perceived risk of arrest and getting away with a crime leads to lower risk perceptions (Anwar and Loughran, 2011; Lochner, 2007; Paternoster et al., 1983; Saltzman et al., 1982; Thomas, Loughran, and Piquero, 2013). There is also evidence—both indirect and direct—that within-individual changes in risk perceptions lead to within-individual changes in behavior. Indirectly, there are consistent research findings that policies aimed at temporarily increasing the risk of arrest decreases criminal behavior (e.g., police crackdowns and hot-spots policing; see Braga, Papachristos, and Hureau, 2012; Sherman, 1990). Directly, Loughran and colleagues (2016) found that within-individual changes in perceived risk were associated with within-individual changes in offending, so that as an individual's perceived risk of arrest increased, his or her offending behavior decreased (see also Matsueda, Kreager, and Huizinga, 2006). These study outcomes reflect the idea that subjective beliefs are internally coherent and, furthermore, that reported risk perceptions do not need to be comparable between persons to test whether deterrence works as hypothesized.

The results also question whether calibration studies that are aimed at attempting to link individual subjective beliefs to some “objective” arrest risk are useful in assessing the validity of deterrence theory. This issue was at the heart of a recent debate in which Pickett and Roche (2016) challenged the validity of deterrence theory based on Kleck and colleagues' (2005) findings that macro-level arrest rates are weakly correlated with individual perceptions of arrest risk (see also Kleck, 2016). Beyond the conceptual criticisms put forth by others (Braga and Apel, 2016; Nagin, 2016; Pogarsky and Loughran, 2016; Raphael, 2016), the results of our study provide empirical evidence that casts doubt on the validity of the assumptions necessary to conduct calibration studies. In calibration studies, researchers assume that reported risk perceptions are “real” values, but we suggest that this is not necessary for deterrence-based policies to be tested. Almost a decade ago, Lochner (2007) highlighted that individual perceptions of risk are not congruent with objective clearance rates but individuals nonetheless still update their perceived arrest risk in rational ways. The findings of our study may help to explain the findings of Lochner (2007) and may help challenge whether the weak objective–perceived risk correlation observed by Kleck et al. (2005) is as detrimental to the deterrence perspective as some purport (Pickett and Roche, 2016) or whether it is simply an artifact of individuals' reporting tendencies (Pogarsky and Loughran, 2016).

One potential limitation to this study is our reliance on a college student sample, which may not be generalizable to the overall population. Although this is far from the first deterrence-related study to rely on data from college students (e.g., Nagin and Pogarsky, 1993; see also Paternoster, 2010), the fact remains that many deterrence scholars are interested in—and have used—high-school student (Paternoster, 1987) or offender samples (Loughran et al., 2016). There are strong reasons to believe that the current sample differs from other populations in important ways, and future work is encouraged to explore the measurement properties of reported risk perceptions among a more representative sample. Nevertheless, given our focus on demonstrating the “arbitrariness” of people's answers to probabilistic survey items, one might assume that a college student sample would provide a more conservative estimate of the overall arbitrariness of respondents' answers. College students are generally required to have a basic understanding of mathematics and to maintain a generally higher level of education than that of the general



populace. As such, college students may be *less* likely to provide arbitrary answers to surveys, not more.

In conclusion, given its salient presence in criminal justice policy, it is likely that deterrence will remain a prominent, albeit controversial, theory of criminal behavior. It is our hope that the results of this study will be useful for future work that is focused on assessing and refining the tenets of deterrence. Coherent arbitrariness in the reported risk of arrest should not lead scholars to shy away from the theory's perceptual orientation; rather, it simply purports that individuals imperfectly report their perceived risk of arrests on surveys and, in turn, encourages scholars to be mindful of the assumptions that are made on the interpretability of subjective values.

## REFERENCES

- Alberini, Anna, Maureen Cropper, Alan Krupnick, and Nathalie B. Simon. 2004. Does the value of a statistical life vary with age and health status? Evidence from the US and Canada. *Journal of Environmental Economics and Management* 48: 769–92.
- Andersson, Henrik, and Mikael Svensson. 2014. Scale sensitivity and question order in the contingent valuation method. *Journal of Environmental Planning and Management* 57:1746–61.
- Anwar, Shamena, and Thomas A. Loughran. 2011. Testing a Bayesian learning theory of deterrence among serious juvenile offenders. *Criminology* 49:667–98.
- Ariely, Dan, George F. Loewenstein, and Drazen Prelec. 2003. Coherent arbitrariness: Stable demand curves without stable preferences. *Quarterly Journal of Economics* 118:73–106.
- Becker, Gary S. 1968. Crime and punishment: An economic approach. *The Journal of Political Economy* 76:169–217.
- Braga, Anthony A., and Robert J. Apel. 2016. And we wonder why criminology is sometimes considered irrelevant in real-world policy conversations. *Criminology & Public Policy* 15:813–29.
- Braga, Anthony A., Andrew V. Papachristos, and David M. Hureau. 2012. Hot spots policing effects on crime. *Campbell Systematic Reviews* 8:1–97.
- Bruine de Bruin, Wandi, Paul S. Fischbeck, Neil A. Stiber, and Baruch Fischhoff. 2002. What number is “fifty-fifty”? Redistributing excessive 50% responses in elicited probabilities. *Risk Analysis* 22:713–23.
- Fudenberg, Drew, David K. Levine, and Zacharias Maniadis. 2012. On the robustness of anchoring effects in WTP and WTA experiments. *American Economic Journal: Microeconomics* 4:131–45.
- Gottfredson, Michael R., and Travis Hirschi. 1990. *A General Theory of Crime*. Stanford, CA: Stanford University Press.
- Horney, Julie, and Ineke Haen Marshall. 1992. Risk perceptions among serious offenders: The role of crime and punishment. *Criminology* 30:575–94.
- Hudomiet, Peter, and Robert J. Willis. 2012. *Estimating Second Order Probability Beliefs From Subjective Survival Data*. Washington, DC: National Bureau of Economic Research.
- Hurd, Michael D. 2009. Subjective probabilities in household surveys. *Annual Review of Economics* 1:543–62.

- Jensen, Gary F., Maynard L. Erickson, and Jack P. Gibbs. 1978. Perceived risk of punishment and self-reported delinquency. *Social Forces* 57:57–78.
- Kahneman, Daniel, Ilana Ritov, and David Schkade. 1999. Economic preferences or attitude expressions? An analysis of dollar responses to public issues. *Journal of Risk and Uncertainty* 19:203–35.
- Kahneman, Daniel, Ilana Ritov, Karen E. Jacowitz, and Paul Grant. 1993. Stated willingness to pay for public goods: A psychological perspective. *Psychological Science* 4:310–5.
- Kennedy, David M. 2009. *Deterrence and Crime Prevention: Reconsidering the Prospect of the Sanction*. New York: Routledge.
- Kleck, Gary. 2016. Objective risks and individual perceptions of those risks. *Criminology & Public Policy* 15:767–75.
- Kleck, Gary, Brion Sever, Spencer Li, and Marc Gertz. 2005. The missing link in general deterrence research. *Criminology* 43:623–60.
- Lochner, Lance. 2007. Individual perceptions of the criminal justice system. *American Economic Review* 97:444–60.
- Loughran, Thomas A., Ray Paternoster, Aaron Chalfin, and Theodore Wilson. 2016. Can rational choice be considered a general theory of crime? Evidence from individual-level panel data. *Criminology* 54:86–112.
- Loughran, Thomas A., Ray Paternoster, and Kyle J. Thomas. 2014. Incentivizing responses of self-report questions in perceptual deterrence studies: An investigation of the validity of deterrence theory using Bayesian truth serum. *Journal of Quantitative Criminology* 30:677–707.
- Loughran, Thomas A., Ray Paternoster, and Douglas Weiss. 2012. Hyperbolic time discounting, time preferences and deterrence. *Journal of Quantitative Criminology* 28:607–28.
- Manski, Charles F. 2004. Measuring expectations. *Econometrica* 72:1329–76.
- Manski, Charles F., and Francesca Molinari. 2010. Rounding probabilistic expectations in surveys. *Journal of Business & Economic Statistics* 28:219–31.
- Matsueda, Ross L., Derek A. Kreager, and David Huizinga. 2006. Deterring delinquents: A rational choice model of theft and violence. *American Sociological Review* 71:95–122.
- Mulvey, Edward P., Laurence Steinberg, Jeffery Fagan, Elizabeth Cauffman, Alex Piquero, Laurie Chassin, George Knight, Robert Brame, Carol Schubert, Thomas Hecker, and Sandra Losoya. 2004. Theory and research on desistance from antisocial activity among serious juvenile offenders. *Youth Violence and Juvenile Justice* 2:213–36.
- Nagin, Daniel S. 1998. Criminal deterrence research at the outset of the twenty-first century. In *Crime and Justice*, vol. 23, ed. Michael Tonry. Chicago, IL: University of Chicago Press.
- Nagin, Daniel S. 2016. What we've got here is failure to communicate. *Criminology & Public Policy* 15:753–65.
- Nagin, Daniel S., and Greg Pogarsky. 2001. Integrating celerity, impulsivity, and extralegal sanction threats into a model of general deterrence: Theory and evidence. *Criminology* 39:865–92.

- O'Hagan, Anthony, Caitlan E. Buck, Alireza Daneshkhah, J. Richard Eiser, Paul H. Garthwaite, David J. Jenkinson, Jeremy E. Oakley, and Tim Rakow. 2006. *Uncertain Judgements: Eliciting Expert Probabilities*. New York: Wiley.
- Osgood, D. Wayne, and Christopher J. Schreck. 2007. A new method for studying the extent, stability, and predictors of individual specialization in violence. *Criminology* 45:273–312.
- Paternoster, Ray. 1987. The deterrent effect of the perceived certainty and severity of punishment: A review of the evidence and issues. *Justice Quarterly* 4:173–217.
- Paternoster, Ray. 2010. How much do we really know about criminal deterrence? *The Journal of Criminal Law and Criminology* 100:765–823.
- Paternoster, Ray, Linda E. Saltzman, Gordon P. Waldo, and Theodore G. Chiricos. 1983. Perceived risk and social control: Do sanctions really deter? *Law & Society Review* 17:457–80.
- Pickett, Justin T., Thomas A. Loughran, and Shawn Bushway. 2015. On the measurement and properties of ambiguity in probabilistic expectations. *Sociological Methods & Research* 44:636–76.
- Pickett, Justin T., and Sean Patrick Roche. 2016. A few clarifying comments on Pickett and Roche (2016). *Criminology & Public Policy* 15:831–6.
- Piliavin, Irving, Rosemary Gartner, Craig Thornton, and Ross L. Matsueda. 1986. Crime, deterrence and rational choice. *American Sociological Review* 51:101–19.
- Piquero, Alex R., and Stephen Tibbetts. 1996. Specifying the direct and indirect effects of low self-control and situational factors in offenders' decision making: Toward a more complete model of rational offending. *Justice Quarterly* 13:481–510.
- Pogarsky, Greg. 2007. Deterrence and individual differences among convicted offenders. *Journal of Quantitative Criminology* 23:59–74.
- Pogarsky, Greg, and Thomas A. Loughran. 2016. The policy-to-perceptions link in deterrence. *Criminology & Public Policy* 15:777–90.
- Pogarsky, Greg, Sean Patrick Roche, and Justin T. Pickett. 2017. Heuristics and biases, rational choice, and sanction perceptions. *Criminology* 55:85–111.
- Pratt, Travis C. 2008. Rational choice theory, crime control policy, and criminological relevance. *Criminology & Public Policy* 7:43–52.
- Pratt, Travis C., Francis T. Cullen, Kristie R. Blevins, Leah E. Daigle, and Tamara D. Madensen. 2006. The empirical status of deterrence theory: A meta-analysis. In *Taking Stock: The Status of Criminological Theory*, eds. Francis T. Cullen, John P. Wright, and Kristie R. Blevins. New Brunswick, NJ: Transaction.
- Prelec, Drazen. 2009. A Bayesian truth serum for subjective data. *Science* 306:462–6.
- Raphael, Steven. 2016. Optimal policing, crime, and clearance rates. *Criminology & Public Policy* 15:791–8.
- Saltzman, Linda E., Ray Paternoster, Gordon P. Waldo, and Theodore G. Chiricos. 1982. Deterrent and experiential effects: The problem of causal order in perceptual deterrence research. *Journal of Research in Crime and Delinquency* 19:172–89.
- Sherman, Lawrence W. 1990. Police crackdowns: Initial and residual deterrence. In *Crime and Justice: A Review of Research*, vol. 12, eds. Michael Tonry and Norval Morris. Chicago, IL: University of Chicago Press.
- Siminski, Peter. 2008. Order effects in batteries of questions. *Quality and Quantity* 42: 477–90.

- Stewart, Jennifer M., Eamon O'Shea, Cam Donaldson, and Phil Shackley. 2002. Do ordering effects matter in willingness-to-pay studies of health care? *Journal of Health Economics* 21:585–99.
- Taylor, Kathryn L., Rebecca A. Shelby, Marc D. Schwartz, Josh Ackerman, V. Holland LaSalle, Edward P. Gelmann, and Colleen McGuire. 2002. The impact of item order on ratings of cancer risk perception. *Cancer Epidemiology Biomarkers & Prevention* 11:654–9.
- Thomas, Kyle J., Thomas A. Loughran, and Alex R. Piquero. 2013. Do individual characteristics explain variation in sanction risk updating among serious juvenile offenders? Advancing the logic of differential deterrence. *Law and Human Behavior* 37:10–21.
- Tversky, Amos, and Daniel Kahneman. 1974. Judgment under uncertainty: Heuristics and biases. *Science* 185:1124–31.
- Wright, Bradley R., Avshalom Caspi, Terrie E. Moffitt, and Ray Paternoster. 2004. Does the perceived risk of punishment deter criminally prone individuals? Rational choice, self-control, and crime. *Journal of Research in Crime and Delinquency* 41:180–213.
- Wright, William F., and Urton Anderson. 1989. Effects of situation familiarity and financial incentives on use of the anchoring and adjustment heuristic for probability assessment. *Organizational Behavior and Human Decision Processes* 44:68–82.
- Zimring, Franklin E., and Gordon J. Hawkins. 1973. *Deterrence: The Legal Threat in Crime Control*. Chicago, IL: University of Chicago Press.

Kyle J. Thomas is an assistant professor in the Department of Criminology and Criminal Justice at the University of Missouri—St. Louis. His research interests include offender decision-making, peer influence, and delinquent attitudes.

Benjamin C. Hamilton is a doctoral student in the Department of Criminology and Criminal Justice at the University of Missouri—St. Louis. His research interests include offender decision-making, measurement of criminological constructs, and quantitative methodology.

Thomas A. Loughran is a professor in the Department of Criminology and Criminal Justice at the University of Maryland—College Park. His research interests include offender decision-making and deterrence, illegal market participation, and public policy.

## SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article at the publisher's web site:

**Appendix A.** Balance Statistics

**Appendix B.** Supplementary Analyses