Intuitive Jurisprudence: Early Reasoning About the Functions of Punishment

Jessica Bregant, Alex Shaw, and Katherine D. Kinzler*

Traditional research on lay beliefs about punishment is often hampered by the complex nature of the question and its implications. We present a new intuitive jurisprudence approach that utilizes the insights of developmental psychology to shed light on the origins of punishment intuitions, along with the first empirical study to test the approach. Data from 80 child participants are presented, providing evidence that children expect punishment to serve as a specific deterrent, but finding no evidence that children expect punishment to have a general deterrent or rehabilitative effect. We also find that children understand punishment in a way that is consistent with the expressive theory of law and with expressive retributivism, and we present evidence that an understanding of the value of punishment to the social contract develops throughout childhood. Finally, we discuss the application of the intuitive jurisprudence approach to other important legal questions.

I. INTRODUCTION

Questions about why, when, and how we punish social violations are central to the theory of law. Conceptions of law as old as Plato and Aristotle have incorporated punishment as an explicit or implied mechanism for enforcing “the law” (Tamanaha 2008). Though normative philosophical debates about punishment are still alive today (e.g., Bilz & Darley 2004; Fincher & Tetlock 2015; Hanna 2008; Murphy & Hampton 1990), modern scholars have also employed new approaches and theories from the fields of law, psychology, and economics to the descriptive task of understanding what drives adults’ intuitions about punishment (see, e.g., Bilz 2007; Camerer & Thaler 1995; Cushman 2008; Darley & Pittman 2003; Haidt 2001; Hans & Reyna 2011). In spite of this wealth of research, a key question remains: How do humans develop ideas about the functions of punishment beginning early in life? Asking how children’s earliest intuitions about justice and punishment become adults’ policy preferences brings together descriptive and normative study with basic science on morality and behavior.

*Address correspondence to Jessica Bregant, email: jbregant@chicagobooth.edu. Bregant is Ph.D. student in Business and Psychology at the University of Chicago; Shaw is Assistant Professor of Psychology at the University of Chicago; Kinzler is Associate Professor of Psychology and of Human Development at Cornell University.

This research was made possible, in part, by NICHD grant R01 HD070890, and by support from the University of Chicago Center for Practical Wisdom from a grant provided by the John Templeton Foundation, both awarded to Kinzler.
Understanding the development of punishment intuitions could not only provide insight into the nature of human reasoning, but it could also offer valuable insights into how best to draft policies to maximize effectiveness and perceived legitimacy.

A well-crafted approach to understanding our early intuitions about punishment could provide a valuable tool for understanding how complex legal ideas form and grow. When do children’s earliest moral judgments, for example, start to reflect the adult belief that punishment can accomplish more than simply retribution? How do early notions of right and wrong become more nuanced ideas about the social contract? This article outlines an approach for beginning to answer these questions, and it proceeds in three parts. First, we outline an intuitive jurisprudence approach to law and psychology that incorporates methods from developmental psychology alongside more traditional law and psychology techniques. We then demonstrate the application of this approach to the question of how people begin to understand the functions of punishment. We provide empirical evidence that children can think and reason about punishment in ways that go beyond simple retribution, including reasoning about specific deterrence and the expressive functions of punishment. Finally, we conclude by suggesting further applications of an intuitive jurisprudence approach in the area of punishment and beyond.

II. THE INTUITIVE JURISPRUDENCE APPROACH TO LAW AND PSYCHOLOGY

While social and economic models of legal psychology have produced a wealth of valuable scholarship, they have also left some important issues unaddressed. Experimental psychology and law scholarship in the last decades has generally been driven by classic social psychology principles and methods, seeking to understand the role of individual and environmental factors on legally-relevant behavior and attitudes (e.g., Anderson & MacCoun 1999; Bilz & Nadler 2009; Burns et al. 2012; Hans & Reyna 2011; Hans & Slater 1984; Robbennolt 2003; Sevier 2014; Stoffelmayr & Diamond 2000; Thompson & Hastie 1990). The influence of economics on law has also led to a growing body of work in law and behavioral economics, which draws on elements of psychology, economics, and decision theory to explain behavior at the individual and group levels (e.g., Funk 2007; Jolls et al. 1998; Kessler & Rubinfeld 2007; Rachlinski 1999; Robbennolt & Studebaker 1999; Thrasher et al. 2007). The social psychology and behavioral economics approaches both study behavior and attitudes at a single developmental moment in adulthood—and they therefore tend to overlook these “intuitive jurisprudence questions,” fundamental questions about how our basic intuitions about the law develop, grow, and change. By incorporating methods from developmental science, researchers in law and psychology can begin to answer these questions.

Of course, developmental psychologists have already made important contributions to legal scholarship, but this research has traditionally been limited to considering the role of children in the legal system. These literatures generally seek to describe and improve the treatment of children throughout the legal processes in which they may be
Developmental researchers have produced a wealth of valuable research on topics such as the reliability of eyewitness testimony from children (e.g., Ackil & Zaragoza 1995; Ceci et al. 1987; Goodman et al. 2002; Koriat et al. 2001; Lampinen & Smith 1995; Tobey & Goodman 1992), the formation of false memories in children (e.g., Brainerd & Reyna 1996, 1998; Ceci et al. 1987; Reyna et al. 2002), pathological and environmental factors surrounding child abuse (e.g., Crimmins et al. 1997; Gray et al. 2005; Sternberg et al. 1997), the best interests of children and families in the legal system (e.g., Ezzell 2001; Hughes 1977; Lahey et al. 1988; Melton & Wilcox 1989; Patterson 2009), and the disclosure and reporting of child abuse (e.g., Brosig & Kalichman 1992; Ibanez et al. Terao 2006; Kalichman et al. 1990; Sternberg et al. 1997; Zellman 1992).

In contrast to this traditional approach, the intuitive jurisprudence approach seeks to put the tools of developmental psychology to a different use: one that leverages the insights from children to help explain adult intuitions. In some ways, the merits of this approach are obvious; after all, children grow into adults, and it makes sense to consider how childhood ideas may develop into adult ideas. Beyond that, however, developmental psychology may provide unique access to the origination and progression of many adult notions. In particular, developmental research can address critical questions about how our early intuitions about legal matters emerge, what cognitive capacities are required to understand and endorse legal principles, and how maturity and experience may change these ideas over time.

Investigating the development of intuitions about morality and punishment is a particularly useful place to begin an intuitive jurisprudence inquiry. Moral judgment is of central importance to law and to psychology. Jean Piaget prefaced his seminal book The Moral Judgment of the Child by noting: “In a sense, child morality throws light on adult morality. If we want to form men and women, nothing will fit us so well for the task as to study the laws that govern their formation” ([1932] 1997:12). Modern psychologists have taken Piaget’s task seriously, shedding light on children’s moral psychology, and exploring topics like children’s prosocial behavior (e.g., Bloom 2010; Eisenberg et al. 1983; Hepach et al. 2013; Ibbotson 2014; Malti et al. 2009; Ross A. Thompson & Newton 2013), understanding of moral and conventional rules (e.g., Helwig & Turiel 2002; Rhodes & Chalik 2013; Smetana 1981; Tisak & Turiel 1984), moral judgment (e.g., Danovitch & Bloom 2009; Kenward & Dahl 2011; Powell et al. 2012; Smetana et al. 2012; Ross A. Thompson 2015; Turiel 2008), and reactions to antisocial behavior (e.g., Hamlin et al. 2011; Kenward & Dahl 2011; Moor et al. 2012; Vaish et al. 2011).

In spite of this rich and growing literature on children’s moral reasoning, little work has taken up the questions—implied in Piaget’s call—of how this reasoning develops into the more complex ideas that surround adult judgments. This article begins to address these questions by asking how children’s intuitive notions of moral judgment reflect adult legal theories of punishment. To do so, we employ a theoretical approach that takes into account both basic intuitions and legal doctrines and that looks for connections to bridge the apparent gap between them: intuitive jurisprudence.
Adult justifications of punishment are generally described with reference to established philosophical categories: retributivism and consequentialism. Retributivism holds that moral desert of an offender is a necessary and sufficient basis for punishment (Dressler 1999); in other words, the retributivist holds that punishment “serves as an end (and a good) in itself” (Bilz & Darley 2004). “Eye for an eye” justice is often considered retributive, in that it seeks to answer moral wrong with moral punishment, but retributivism does not require literal equivalence of crime and punishment. Because it makes the moral blameworthiness of the offender a central factor in the determination of punishment, retributivism requires only that punishment be morally proportional to the underlying offense (Carlsmith 2006). Thus, while the intentions of the offender are vitally important, the degree of harm caused, as an independent concern, is generally irrelevant to the retributivist calculus.

In contrast to retributivism’s focus on moral desert, consequentialism focuses on the potential benefits of punishment to society. Consequentialists permit punishment only to the extent that its benefits outweigh its harms (Dressler 1999). Potential benefits of punishment vary, but commonly cited benefits include deterrence (both “specific” deterrence to the offender in question and “general” deterrence to other potential offenders), rehabilitation of the offender, and incapacitation (Carlsmith et al. 2002; Carlsmith 2006; Darley et al. 2000; Dressler 1999). Under pure consequentialist theories, moral desert of the offender is not relevant; consequentialist punishment is forward looking, in the sense that it is concerned only with what can be gained by punishing (Dressler 1999).

While a great deal of scholarship has described retributivism and consequentialism, empirical attempts to compare these philosophies to the actual intuitions of lay people have met with mixed results: despite broad endorsement of nonretributive theories in survey research (Cullen et al. 1990), behavioral research has found a persistent tendency to enact punishment in line with retributive ideals (Carlsmith 2006, 2008). People are capable of using both retributive and consequentialist factors in their punishment determinations, but they seem to require less prompting to take retributive factors into account, and their “default” tendencies seem to strongly favor retributivism (Darley et al. 2000). This apparent preference for retribution has led some to argue that people’s professed endorsement of other punishment goals is nothing more than an inability to articulate the causes of their own behavior (Carlsmith 2008; see also Nisbett & Wilson 1977). The implication of this argument seems to be that people are “truly” retributivists, and that deterrence and other consequentialist justifications for punishment are merely rationalizations. However, the pervasiveness of consequentialist justifications in everyday discussions of punishment makes this a difficult hypothesis to test using traditional methods, at least in adult participants, whose ideas about punishment are often heavily informed by prevailing culture, philosophy, history, or legal matters.
Children, in contrast, provide a perfect opportunity to explore the intuitions underlying adult beliefs that are less “tainted” by complex theories of morality and punishment. The intuitive jurisprudence approach is especially well suited to this task because it affords a way to strip away the complicated rhetoric of policy debates and examine our earliest developing understandings of punishment. This leads to the first of our current research questions:

1. Do early punishment intuitions incorporate consequentialist reasoning about punishment?

If consequentialist ideas really are no more than rationalizations for retributive instincts, we might expect children to show little or no evidence of consequentialist thinking. On the other hand, if children expect punishment to serve as a deterrent or to rehabilitate offenders, that would suggest that consequentialist thinking runs deeper than currently thought.

Certainly, the desire to see punishment meted out exists early in childhood. Infants as young as eight months old prefer those who hinder (punish) an antisocial agent over those who help an antisocial agent (Hamlin et al. 2011), despite a general preference for agents who help over those who hinder (Hamlin et al. 2007). By the time children are two years old, they will enact punishment themselves; toddlers who were given the option to take away a treat from a puppet who previously helped or a puppet who previously hindered a third party chose to punish the hinderer puppet by taking away his treat (Hamlin et al. 2011). Older children—those in early elementary school—will even endure a cost to themselves to punish a perceived wrongdoer, particularly if the transgressor is an out-group member (Jordan et al. 2014; McAuliffe et al. 2015).

Some developmental research has referred to children’s desires for “retribution” or “restoration,” but these labels are generally used to describe the punishment act itself, not the underlying justification for punishment (Hamlin 2013; Riedl et al. 2015). Our research questions seek to push the discussion of children’s punishment ideas beyond simple retribution by exploring the origins of other complex adult ideas about the functions of punishment.

No research to date has systematically explored children’s reasoning about the functions of punishment and compared it to adult justifications, but some prior work provides tantalizing hints that children may think about punishment as being more than simply retributive. Stern and Peterson (1999) asked children in a range of age groups (4–5 years, 7–8 years, and 10–11 years) to assign punishment for a variety of hypothetical moral, safety, and social-conventional offenses committed by another child. Participants were then asked, among other questions, why the punishment was appropriate. Children’s answers were coded into broad categories of response, including “rudimentary” and “prevention” (Stern & Peterson 1999). Some of the rudimentary answers, like “Because what he did was wrong,” may indicate retributive thought processes, but the category also included tautological items like “[Y]ou have to spank him.” Rudimentary answers were equally likely to be given by older and younger children. In contrast, prevention responses seem to reflect a possible expectation of deterrence or
rehabilitation. Responses included, “So that it will help her make a wiser choice next time” and “So that she’ll get it in her head not to disobey,” and they were strongly associated with age; while children 7–8 and 10–11 occasionally used these justifications, children 4–5 years old gave almost no “prevention” answers. Children, especially in the older age groups, also commonly gave “negative consequence” answers, which included, “Because the person will be upset if she can’t have her candy bar” and “Because he’ll get hurt if he falls off the seesaw” (Stern & Peterson 1999).

Sociological evidence also shows that older children may provide consequentialist rationales for punishment. For example, in Ghana, where physical punishment of children is commonplace, older children are sometimes given the authority to punish younger siblings and cousins. In a diary study of Ghanaian children, sociologist Afua Twum-Danso Imoh (2013) found that some of these children offered apparently spontaneous deterrence rationales, such as such as a 13-year-old child who wrote:

When my younger siblings do something wrong, I usually spank them on the buttocks. I do that so that they will not repeat the act again because if I just tell them not to do that again, next time they will forget my warning and do it. (Twum-Danso Imoh 2013)

Another child, age 14, explained: “If a child does something wrong and gets punished for it, he/she will be careful next time; if not the child will do whatever he/she likes and grow up to become a bad person.” However, other participants—even those who supported physical punishment—were less clear about the reasons for punishment, such as the 10-year-old who argued: “It is parents’ duty to train a child to become a good adult and [physical] punishment is one of the ways through which a child can be corrected, so it is good” (Twum-Danso Imoh 2013).

Taken together with Stern and Peterson (1999), these findings suggest that children spontaneously generate retributive justifications for punishment, and that at least by the age of seven or eight, some children also seem to provide rationales that go beyond retribution. These “prevention” responses (Stern & Peterson 1999) may map onto adult notions of consequentialism, but it is difficult to distinguish what specific functions the child expects punishment to fulfill. Do children see punishment as a deterrent, or do they view it as a mechanism for rehabilitation? Moreover, as the tautological responses present in both studies reveal, children may struggle to identify and articulate the reasons for their punishment intuitions. The direct questioning method may therefore not accurately reflect children’s reasoning, just as it often fails to reflect adult reasoning (e.g., Nisbett & Wilson 1977), and it may underestimate children’s understanding of how punishment functions. An indirect method, which could measure children’s implicit beliefs and expectations without relying on their ability to give voice to their reasoning, could produce more reliable evidence of children’s thought processes.

IV. Social and Expressive Implications of Punishment

Of course, adult ideas about punishment are not limited to retribution and consequentialism; punishment serves many other important social functions. Hobbesian
philosophers argue that rational (and self-interested) human beings agree to subject themselves to the authority of the state so that they may live in a civil society. As a mechanism for enforcing the rule of law, punishment is an integral part of that social contract (e.g., Hobbes [1651] 1996). Research showing broad support for punishment policies (see, e.g., Cullen et al. 1990; Tyler & Boeckmann 1997) suggests that, at least when it comes to the punishment of others, adults perceive punishment as a social benefit. On the other hand, young children’s experiences are generally as targets of rule enforcement. This experience could lead them to believe that punishment is nothing but a hindrance to their own desires. By comparing ideas about punishment in society across age groups, we can address a second broad research question:

2 When does a basic understanding of the social contract emerge?

When do humans begin to understand the basic importance of punishment as a tool for keeping social order? By examining the progression of social ideas across childhood, we can better understand the foundations of social order; the development of social ideas in children can serve as a kind of model for the development of social ideas in a culture. This question is one that only a developmental approach can answer, and it is a particularly clear illustration of the value of an intuitive jurisprudence approach.

Punishment can also be an important medium for communicating messages about society’s shared social values. This view, often referred to as the “expressive function” of law, encompasses a wide range of potential messages that law and law enforcement (i.e., punishment) can send (e.g., Feinberg 1965; Funk 2007; Sunstein 1995). At the broadest level, laws signal what is and is not acceptable in society. The sanctions imposed by law may also communicate more nuanced information about the particular kinds of actions and motives that a society permits (Feinberg 1965); for example, a murder statute that subjects “heat of passion” crimes to lesser sanctions expresses a degree of tolerance for those who kill under specific circumstances. The selective enforcement of codified laws also sends an important social message, particularly when that enforcement favors certain groups of people (Harcourt 2002).

In contrast to the normative theories of retribution and consequentialism, the expressive theory of punishment is primarily descriptive, but scholars have used focused expressive arguments to make normative claims as well.¹ In Jean Hampton’s theory of “expressive retributivism,” punishment of wrongdoing sends a social message about the status of the victims and the wrongdoers (see, e.g., Hampton 1992; Hanna 2008; Murphy & Hampton 1990). Hampton argues that punishment allows society to reject the offender’s “false moral claim” of right over the victim (Murphy & Hampton 1990). In

¹Others have described expressive functions as any way in which law can influence behavior other than deterrence (Dharmapala & McAdams 2003). Though this description coincides with consequentialist functions of punishment, for our purposes it is sufficient to highlight expression as a distinct function from more traditional consequentialist concerns like rehabilitation and incapacitation. We have also chosen to treat expressive functions as orthogonal to retribution and consequentialism to better allow for the inclusion of expressive retributivism in the discussion.
other words, while crime acts to lower the status of the victim relative to the offender, punishment may act to restore the victim’s social standing (see also Bilz 2010; Kahan & Nussbaum 1996).

Developmental research can also help inform our understanding of the expressive functions of punishment. If punishment does serve a meaningfully communicative or expressive function, young children, who are just beginning to learn about the organization of society, might be particularly sensitive to its message. This leads to our third major research question:

3 Does punishment serve expressive functions by communicating meaningful social information to children?

Prior research has generally concluded that, for at least some kinds of laws, the communication of social norms through laws can influence group-level behavior (see, e.g., Cooter 2000; Funk 2007; McAdams 2000), but while some scholars have put forth theories about how laws use expressive messages to shape behavior (e.g., Geisinger 2002; Kahan 1997), little research has empirically evaluated the particular messages communicated by punishment. By adopting a simple and easy-to-understand paradigm, we can explore precisely what children learn about the social world through the implementation of punishment.

V. THE PRESENT STUDY

By examining the ways in which children reason alongside the more traditional ways in which legal thought is organized, the intuitive jurisprudence approach provides a new method for understanding the fundamental psychology of punishment justifications. In spite of the wealth of research on adult ideas about punishment and the growing body of work on children’s moral reasoning, ours is the first study to bring these two programs together in a systematic way. In our paradigm, children watch a slideshow with two stories about unfamiliar fictional worlds. In each world, one of three characters steals a favored treat from another character. The worlds differ in punishment policy; in one world, it is explained that residents are “never punished for stealing,” while in the other, residents are “always punished for stealing.” Children are then asked to make predictions about the future stealing behavior of each character, as well as to rate how much they like each character. At the end of each world, children are asked whether stealing is “bad” in the world. Finally, after viewing both stories, children are asked in which world they would prefer to live.

By comparing children’s responses across the two stories, we can test whether early beliefs about the functions of punishment go beyond simple retribution. This paradigm allows us to address our research questions in detail. Our first question can be broken down into two subparts:

1. Do early punishment intuitions incorporate consequentialist reasoning about punishment?
a. Do children expect punishment to deter (either specifically or generally)?
b. Do children expect punishment to rehabilitate?

If children expect punishment to have a specific deterrence effect, they should predict that a thief who is punished is less likely to steal in the future than a thief who is not punished. If they expect punishment to have a more general deterrent effect, they should similarly predict that the bystander in the punishment world, who merely sees punishment occur, is less likely to steal in the future than the bystander who lives in the no-punishment world.

Our paradigm also allows us to see whether the thief himself is restored in the eyes of the child; that is, whether punishment rehabilitates the offender. If a thief is rehabilitated by punishment, we would expect children to not only predict a lower likelihood of recidivism, but also to like the punished thief more than the nonpunished thief.

Second, this paradigm allows us to probe children’s understanding of the broader functions of punishment in a just society. Armed with this knowledge, we can begin to answer our second research question:

2. When does a basic understanding of the social contract emerge?

To explore the extent to which children believe the social benefits of punishment outweigh the costs, we asked all children which world they would prefer to inhabit. Although we expected many children to be reluctant to choose the punishment world, we were particularly interested in whether their preferences would remain stable across the age range of our sample, or whether their responses may begin to change with age, reflecting a growing understanding of the social contract—that punishment can be good because it protects one from being harmed by others.

Our final research question can be evaluated from multiple angles:

3. Does punishment serve expressive functions by communicating meaningful social information to children?
   a. Do children receive a message about the acceptability of an action from the presence or absence of punishment?
   b. Do children receive a message about the social status of the victim from the presence or absence of punishment?

If, as the expressive theory of law holds, children learn about the acceptability of an act based on its punishment status, they should respond that stealing is “bad” in the world with punishment for stealing, but not in the world without punishment for stealing. Given that children may already have strong ideas about the wrongness of stealing, our paradigm presents a conservative test of this expressive hypothesis; children may respond, based on their own experience, that stealing is bad regardless of whether it is punished in a particular world. Our paradigm further allows us to test the expressive retributivism idea that punishment communicates information about the social status of a victim because we can compare how much children report liking the victim in both stories. If punishment restores some social standing to the victim of a harm, then children might report liking a victim more when the thief who harmed her is punished.
Importantly, and in contrast to previous work, our paradigm relies on simple and implicit expectations to evaluate participants’ reasoning. Children are asked simple questions about their expectations and liking, and they respond on simplified and illustrated scales. This child-friendly method allows us to measure the naïve theories of our participants without requiring children to articulately verbalize their own thoughts about punishment, and it thus may provide clearer access to their early developing reasoning and understanding than more complicated procedures.

A. Method

1. Participants

Participants were 80 children (42 females) ages 5–8 (M_{age} = 6.9, range = 5.0–8.9); the sample included 20 5-year-olds (7 females), 20 6-year-olds (11 females), 20 7-year-olds (14 females), and 20 8-year-olds (10 females). Based on an a priori rule, we excluded from analysis 11 additional children who answered at least one recall question incorrectly (M_{age} including excluded children = 6.8, range = 5.0–8.9).

2. Materials

Stimuli were vignettes presented in animated slide shows, labeled “Square World” and “Triangle World.” In each show, three anthropomorphic shapes (squares in Square World and triangles in Triangle World) appeared. Each character was a different color, and the experimenter labeled each according to their color (i.e., “a red triangle, a yellow triangle, and a blue triangle”). The experimenter then narrated a story in which one of the characters stole from another character. In the beginning of the story, the characters received favorite treats (cupcakes or ice cream) and then went to play before snack time, leaving the screen. While the other characters were absent, one of the characters (the thief) returned to the screen and consumed another character’s treat. The experimenter said: “Look. The [red triangle] stole the [yellow triangle’s] [ice cream] and ate it up.” The thief then left the screen and returned with the other two characters, while the experimenter explained that it was “snack time.” During “snack time” the thief and the bystander consumed their treats, while the victim of the theft frowned. To emphasize the theft and the roles, the experimenter said: “But the [yellow triangle] is sad, because it doesn’t have [an ice cream] to eat, because the [red triangle] stole the [yellow triangle’s] [ice cream] and ate it up. Poor [yellow triangle].”

At this point, the stories diverged. In the no-punishment version, the experimenter explained that, in that world, “[triangles] are never punished for stealing. So now the [red triangle, the yellow triangle, and the blue triangle] are going to go play.” After this line, the victim returned to smiling, and the characters bounced happily off the screen. In the punishment version, the experimenter instead explained, “[triangles] are always

\[2\] Throughout this article, we will use the terms “thief,” “victim,” and “bystander” when describing the characters. However, the children never heard these words. Instead, the characters were always referred to by their color and shape, i.e., “the red triangle.”
punished for stealing. So now, the [red triangle] is in time out, but the [yellow triangle and the blue triangle] are going to go play.” After the words “time out,” a small set of cartoon jail bars appeared in front of the thief character, who frowned. The other two characters (now both smiling) bounced off the screen.

3. Design and Procedure

Children were first familiarized with two four-point answer scales, as shown in Figure 2; one scale measured liking, and the other measured perceived probability. During familiarization, the experimenter explained what each scale point represented. These simplified scales were adapted from prior research, and they were designed to allow children to easily grasp their meanings and to facilitate children’s responses (see, e.g., Eston et al. 2009; Killen et al. 2011; Olson et al. 2006; Olson & Shaw 2011). The experimenter

Figure 1: Example stimuli from “Triangle World” (left) and “Square World” (right).

Figure 2: Two four-point scales used by participants.

Note: (a) “Liking” scale. Faces were said to represent, L–R: “Don’t like at all,” “Don’t like a little,” “Like a little,” and “Like a lot.” (b) “Yes-No” scale. Responses were, L–R: “Definitely no,” “Probably no,” “Probably yes,” and “Definitely yes.”
demonstrated the use of the scale (e.g., “My favorite food is spaghetti. If you asked me if I like spaghetti, I would choose ‘like a lot!’”), and then children were asked practice questions about their food preferences (e.g., “What’s something you really like to eat? If I asked if you like [that food], what would you pick?”). This practice was repeated for a disliked food, and then again for a food that the experimenter and child said was “just okay.” After these three rounds of practice questions, the experimenter asked additional questions of the same type until the participant had responded using both endpoints of the scale and at least one “inner” point on the scale.

To familiarize participants with the probability scale, the experimenter again modeled its use by referring to foods, for example: “If you asked if I’m going to have spaghetti for dinner this week, I would say, ‘Spaghetti is my favorite food, and I eat it almost every week,’ so I would choose ‘Definitely Yes’.” Then the experimenter asked the child: “Do you think you’ll eat [your favorite food] this week?” As with the first scale familiarization, the experimenter continued to ask questions about what the child might eat “this week” until the child responded using both endpoints and at least one nonendpoint.

For the test trials, each participant then saw a story about Triangle World and a story about Square World. For each participant, one world had punishment for stealing and one world had no punishment for stealing. The order of the worlds, as well as which world (Triangle or Square) had punishment and which world did not, was counterbalanced between participants. After each world, participants were asked four or five recall questions to ensure that they understood and remembered the role of each character and whether or not punishment had occurred.3

At the end of the first story, children were presented with images of all three characters, in the same position they had been on the screen at the beginning of the story. Children were then asked, for each character: “Do you like [the red triangle]?” They responded on a four-point scale from “Don’t like at all” to “Like a lot,” as shown in Figure 2(a). Children were then asked, again for each character: “Do you think [the red triangle] will steal the next time they play together?” They responded on the four-point scale from “Definitely No” to “Definitely Yes” shown in Figure 2(b). Finally, children were asked, “Is stealing bad in [Square World]?” and they responded on the same “Definitely No” to “Definitely Yes” scale.

After the second story, participants answered the same questions as they had after the first story, as discussed above. When those questions were completed, children saw a new screen with both sets of characters, and the experimenter asked: “Which world would you rather live in? Square World or Triangle World?” When children responded, the experimenter followed up by asking why. Responses to the “why” question were coded as being about punishment or not about punishment.

3The questions traced the events of the story: “Did one of the [squares] steal?”; “Which one stole?”; “Was one of the [squares] punished?”; “Which one?”; “Who did the [red square] steal from?” The questions were tailored to the story the participant had just seen, so that children were not asked who was punished if, in fact, no character had been punished. Correct answers were given positive feedback (e.g., “That’s right!”). If an incorrect answer was given, the experimenter corrected the child.
In summary, the primary independent variable of interest was whether the world in question was a “punishment” world or a “no-punishment world,” and it was manipulated within subjects. The order of the two stories, as well as which world was the “punishment” world, were varied between subjects.

B. Results

All repeated measures (i.e., all measures except which world the child would rather live in) were tested with a two-way analysis of variance (ANOVA) (within-subjects factor: punishment world vs. no-punishment world, continuous variable: participant age). Analysis of the world choice question is described below. Two additional factors, presentation order and participant gender, were also tested using 2 (world, within-subjects) × 2 (presentation order or participant age, between-subjects) ANOVAs. Neither had any significant interaction with our variable of interest, that is, the within-subjects punishment
manipulation, and we therefore omit any further discussion of order or gender in these results.4

1. Evaluations of the Thief

For children’s ratings of whether a thief would steal again in the future, there was a significant main effect of punishment, $F(1,78) = 5.10, p = 0.03$, a marginally significant main effect of age, $F(1,78) = 3.32, p = 0.07$, and a significant interaction between whether the thief had been punished and the age of the participant, $F(1,78) = 4.72, p = 0.03$. Across all ages, participants rated the thief who was punished as less likely to steal next time than the thief who was not punished ($M_{\text{punished}} = 1.19$, $M_{\text{nonpunished}} = 1.53$), as shown in Figure 3. When the thief was not punished, there was a significant positive correlation5 between age and reported likelihood of stealing, $r = 0.31, p < 0.01$; older children reported that the unpunished thief was more likely to steal in the future than did the younger children. There was no effect of age on predictions about the thief who was punished, $r < 0.01$. Collapsing across the punishment condition, there was a small positive correlation between age and predicted likelihood of future theft, $r = 0.15, p = 0.05$; older children were generally more likely than younger children to report that a thief would steal again, regardless of whether the thief was punished.

There was no significant difference between liking ratings for thieves who were punished compared to those who were not, $F(1,78) = 0.90, p = 0.34$ ($M_{\text{punished}} = 0.64$, $M_{\text{nonpunished}} = 0.76$). There was a main effect of age on liking of the thief, $F(1,78) = 4.20, p = 0.04$, such that younger children reported liking the thief more than older children did, $r = -0.18, p = 0.02$, but there was no interaction between age and punishment status, $F(1,78) = 0.08, p = 0.78$.

2. Evaluations of the Bystander

There was no main effect of punishment on participants’ ratings of bystanders’ likelihood to steal, $F(1,78) = 0.59, p = 0.44$; that is, bystanders were rated as equally likely to steal next time regardless of whether they lived in a world where people are always punished for stealing or one in which people are never punished for stealing, ($M_{\text{punished}} = 0.68$, $M_{\text{nonpunished}} = 0.78$). There were no significant main effects of age $F(1,78) = 0.05, p = 0.82$, and no interaction between age and punishment world, $F(1,78) = 0.03, p = 0.87$.

4Presentation order (world with punishment first vs. world without punishment first) had no significant effect (main effect or interaction) on any of the measured variables, and analyzing only the first world seen by each participant does not change the results of any significance tests. Gender significantly affected two measures; there was a main effect of participant gender on liking of the thief, $F(1,78) = 11.02, p < 0.01$, such that boys rated the thief more highly than did girls ($M_{\text{girls}} = 0.97$, $M_{\text{boys}} = 0.45$), but there was no interaction between gender and punishment status, $F(1,78) = 0.02, p = 0.89$. In addition, girls reported liking the bystander significantly more than did boys, $F(1,78) = 6.70, p = 0.01$ ($M_{\text{girls}} = 2.67$, $M_{\text{boys}} = 2.33$), but there was no significant interaction between gender and world, $F(1,78) = 2.94, p = 0.09$.

5All reported correlations are Pearson’s $r$. 


Similarly, for liking of the bystander, there was no main effect of punishment, $F(1,78) = 0.15, p = 0.70$, no main effect of age, $F(1,78) = 2.23, p = 0.14$, and no interaction, $F(1,78) = 0.27, p = 0.61$ ($M_{punished} = 2.53, M_{nonpunished} = 2.49$).

3. Evaluations of the Victim

Children’s ratings of the victims’ likelihood of theft did not differ based on the presence of punishment, $F(1,78) = 0.38, p = 0.54$ ($M_{punished} = 1.26, M_{nonpunished} = 1.18$), or based on the age of the participant, $F(1,78) < 0.01, p = 0.94$, and the two variables did not interact, $F(1,78) < 0.01, p = 0.95$.

There was a main effect of our manipulation on liking of the victim, $F(1,78) = 4.83, p = 0.02$; victims in the world without punishment were liked less than victims in the world with punishment ($M_{punished} = 2.54, M_{nonpunished} = 2.30$). There was no significant main effect of age, $F(1,78) = 0.33, p = 0.57$, and no interaction between age and punishment world, $F(1,78) = 1.68, p = 0.20$.

4. Personal and Social Value of Punishment

Approximately half ($n = 44, 55.0$ percent) of all participants responded that they would prefer to live in the no-punishment world. The effect of age on world choice was analyzed with a logistic regression, and there was a significant effect of age on world choice, $b = 0.60, p < 0.01$, such that older children were less likely to choose the world without punishment. Among 5-year-olds and 6-year-olds, the no punishment world was preferred nearly three to one (75.0 percent for 5-year-olds and 70.0 percent for 6-year-olds).

Figure 4: World choice by age year.

Note: For all four age groups, $n = 20$. 

![World choice by age year](image.png)
However, 7-year-olds were evenly split, with 45.0 percent choosing the world without punishment, and 8-year-olds showed a strong preference for the punishment world, choosing the no-punishment world just 30.0 percent of the time. These data are summarized in Figure 4.

When asked why they had chosen the world they indicated, 72.5 percent ($n = 58$) of all participants referred to punishment in their explanations. Selected responses are shown in Table 1. There was no significant difference in likelihood of referring to punishment based on world selected, $\chi^2(1, N = 80) < 0.01$, $p > 0.99$.

### Table 1: Sample Explanations Given by Children for Their World Choice

<table>
<thead>
<tr>
<th>Answers Coded as “Not About Punishment”</th>
<th>Answers Coded as “About Punishment”</th>
</tr>
</thead>
<tbody>
<tr>
<td>“I like purple.”</td>
<td>“You don’t get punished.”</td>
</tr>
<tr>
<td>“I love cupcakes.”</td>
<td>“People need to get in time out.”</td>
</tr>
<tr>
<td>“Because the yellow triangle is friendly.”</td>
<td>“If somebody steals from me, they get punished.”</td>
</tr>
<tr>
<td>“Because [a square] has four sides, and that’s close to how old I am.”</td>
<td>“Because it has a law when you steal you get punished, and I would never steal.”</td>
</tr>
</tbody>
</table>

5. Is Stealing Bad?

Participants rated stealing as significantly worse in the world with punishment than in the world without punishment, $F(1,78) = 179.84$, $p < 0.01$ ($M_{punished} = 2.86$, $M_{nonpunished} = 1.15$), as shown in Figure 5. Age had no effect on this measure, $F(1,78) = 0.23$, $p = 0.63$, and there was no interaction between age and world, $F(1,78) = 0.23$, $p = 0.63$.

C. Discussion

Our paradigm allows us to examine the way in which punishment intuitions develop and to evaluate whether early reasoning about punishment reflects adult theories about the functions of punishment. We can now return to the research questions posed earlier:

1. Do early punishment intuitions incorporate consequentialist reasoning about punishment?
   a. Do children expect punishment to deter (either specifically or generally)?
   b. Do children expect punishment to rehabilitate?
2. When does a basic understanding of the social contract emerge?
3. Does punishment serve expressive functions by communicating meaningful social information to children?
   a. Do children receive a message about the acceptability of an action from the presence or absence of punishment?
   b. Do children receive a message about the social status of the victim from the presence or absence of punishment?
1. Consequentialist Reasoning

While prior research has suggested that children (and adults) think little of punishment beyond retribution, our findings illustrate that some facets of more complex reasoning are occurring in young children.

\textit{a. Specific and general deterrence.} Our results provide some evidence that, as young as six years old, children expect punishment to function as a deterrent; however this effect of deterrence is limited. Participants reported that a thief was marginally less likely to steal in the future when that thief was punished than when the thief was not punished, and this expectation is consistent with a specific deterrence effect of punishment. In contrast, we find no evidence that children expect punishment to serve as a general deterrent to others. Bystanders and victims were rated as equally unlikely to steal regardless of whether they saw a thief punished.

\textbf{Figure 5:} Likelihood that stealing is bad in punishment and nonpunishment worlds.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure5.png}
\caption{Likelihood that stealing is bad in punishment and nonpunishment worlds.}
\end{figure}

\textbf{Note:} Error bars indicate standard error of the mean.
Taken together, these results suggest that while children expect the personal experience of punishment to deter future wrongdoing by the punished individual, they do not generalize that deterrent effect to those who merely live under the threat of punishment. However, the lack of evidence for general deterrence is not conclusive. Notably, children seem very skeptical that the bystanders would steal in either world; their overall likelihood ratings on this question were quite low, which could indicate a floor effect. Such a floor effect might hide any differences, obscuring true expectations about general deterrence. Our results do not eliminate the possibility that children expect the threat of punishment to have a general deterrent effect, but future work is necessary to reach such a conclusion. For example, if given a forced choice between a character who lives in a world with punishment and a character who lives in a world without punishment, children might expect that the latter is more likely to steal.

b. Rehabilitation. While children may expect punishment to change the future behavior of a thief, there is no evidence that children believe the punished thief is rehabilitated. We found no differences in children’s reported liking of a thief, regardless of whether the thief was punished. Thus, while thieves are less well-liked than bystanders and victims, thieves who are punished do not seem to be restored, at least in terms of their social standing, in the eyes of our participants. This may suggest that children do not believe that punishment rehabilitates, but this finding, too, warrants further investigation. Although the question of how much participants like the thieves does provide a sense of whether the social standing of the thief was rehabilitated, it does not address rehabilitation in a more personal sense. A different way of conceptualizing rehabilitation could be to ask whether a punished thief is less likely to steal even after the threat of punishment is removed, distinguishing rehabilitation from specific deterrence. For example, if a thief in our paradigm moves from the world with punishment to the world without punishment, do children still predict that the thief is less likely to steal than the thief who has never been punished?

The lack of difference between likability of the punished and the unpunished thieves may also seem somewhat surprising in light of prior research that children tend to dislike “unlucky” individuals (Olson et al. 2006). In our paradigm, the only substantial difference between the thieves is the apparent accident of which world they live in, so children could have construed the presence of punishment as bad luck. However, the punishment difference here was implemented at the policy level, not at the individual level. Children may be less likely to infer any luck differences between the individual thieves in the presence of information that the difference in their outcomes was based on a uniform policy. As we discuss below, our results suggest that children noticed the policy-level differences in punishment between worlds, a finding that would be consistent with this explanation.

2. Punishment and the Social Contract

Our results also provide striking evidence of children’s emerging understanding of the role of punishment in the social contract. When asked which world they would prefer to live in, 5- and 6-year-olds reliably chose the world without punishment, while 8-year-old children overwhelmingly chose the world with punishment. Though it is not surprising that young children might prefer a world in which they cannot be punished, it is remarkable to see such a clear shift in thinking as they age. Anecdotally, many children
who chose the world without punishment gave qualitatively different explanations for their choice than those who chose the world with punishment. As one might expect, the former group tended to explain, “Because you don’t get punished,” while the latter tended to argue that people should get punished for stealing. Moreover, when the children referred to themselves in their explanations, those who chose the no-punishment world often identified themselves as potential targets of punishment (e.g., “Because I won’t get punished”), while those who selected the punishment world tended to identify as potential victims (e.g., “So if someone steals from me, they get punished”).

Perhaps as children age through this period, they begin to see themselves less as potential wrongdoers and more as potential victims. Our sample began at five years old, roughly the age at which children typically begin school, but it was only the 8-year-olds, who had several years of school experience, who reliably chose to live in the world with punishment. Maybe school experiences, and social experiences with other children more generally, teach children that they are more likely to witness punishment than to receive it. Older children probably have more experience being the victim of their peers than do children just starting school. Or perhaps as children age, they begin to see themselves more as authority figures, responsible for ensuring that punishment is carried out and “justice is done.” Alternatively, the change in children’s preferences may reflect an increasing tendency (and ability) to consider the causes of others’ behavior. If punishment deters, then older children might expect that others are less likely to harm them in the world with punishment as opposed to the world without punishment.

Whatever the reason, it is clear that children in our sample began to see punishment as more beneficial than harmful by approximately age eight. Moreover, these children seem to be willing to subject themselves to the threat of punishment in exchange for the benefits it provides. This emerging understanding of the social contract is provocative, and it deserves further study. In particular, future work should explore the mechanisms that support this developmental shift, and the depth of children’s understanding of the broader implications.

3. Expressive Functions of Punishment

Our findings also demonstrate that children are learning important social information from punishment policies.

a. Learning about the wrongness of acts. In the world without punishment for stealing, children indicated that stealing was significantly less “bad” than it was in the world with punishment for stealing. This finding is consistent with theories about the expressive function of law, in which laws are said to communicate a social message. In our simplistic fictional worlds, the social message of punishment seems to be that “stealing is bad,” and despite the fact that these children have presumably had a chance to learn that stealing is bad in their own world, they get this message differentially in our paradigm. Notably, the experimenter also called children’s attention to the harm caused by stealing in both worlds by pointing out that the victim was “sad” and expressing pity for the victim. Nonetheless, if the perpetrator was not punished, children reported that stealing was much less bad.
Our finding that stealing is perceived as worse in the world with punishment also raises additional questions about our finding that thieves are equally well-liked (or, equally poorly liked) in both worlds. In light of the former, the latter finding is potentially perplexing because it suggests that children like (or dislike) the thieves equally well despite the fact that the thief in the punishment world has committed a worse crime. One potential explanation for this contradiction is that dislike of the thieves is so strong that it obscures any difference based on severity of the crime; that is, our participants may have been at floor on liking of the thief, leaving no room to decrease liking.

A more substantive explanation for this discrepancy, however, is that children fail to apply differing “local” standards or norms in the two worlds to the question of how much they like the thief. Our data reveal that children are aware that differing policies (i.e., punishment for stealing vs. no punishment for stealing) may reflect differing community standards; this is evident from the dramatic difference in children’s rating of whether stealing is “bad” in each world. However, children were asked how much they liked the thief immediately after viewing the story and before they were prompted to think broadly about this community standard. Thus, it is plausible that participants’ evaluations of the thief were driven by the strength of the anti-stealing norm at work in the child’s own world, prompting low ratings across the board for thieves.

b. Expressive retributivism and the victim’s social status. Our results also provide some evidence for the “expressive retributivism” theory that punishment communicates a message about victims of harm. Children reported liking the victim significantly more if the thief was punished than if the theft went unpunished. This suggests that, in the eyes of our participants, the punishment of the thief may restore the social standing of the victim, just as expressive retributivism predicts (Murphy & Hampton 1990). Recent research with adults has identified a similar pattern; adults, too, seem to view the punishment of an offender as a reflection on the victim (Bilz 2016).

These findings also raise the possibility that children engage in a form of victim blaming, similar to the way adults sometimes discredit crime victims (see, e.g., Hafer 2000; Lerner & Miller 1978; Savani et al. 2011; van Prooijen & van den Bos 2009), and they therefore add to a mixed body of evidence on children’s attitudes toward victims. As previously noted, children sometimes express increased disliking of “unlucky” individuals, including the victims of harm (Olson et al. 2006), but some subsequent research has found no evidence of victim-blaming behaviors in toddlers (Hamlin et al. 2011). Our results indicate that, at least by the age of five, children may make social inferences about the victims of crime based on whether punishment occurs.

VI. CONCLUSION

Our results provide evidence that complex ideas about punishment develop relatively early in life. While we have not shown that children’s desire to punish stems from a

---

6We are grateful to an anonymous reviewer for this suggestion.
consequentialist notion, we have shown that even 5-year-old children expect at least one consequentialist aim—specific deterrence—to be accomplished through punishment. Our findings also provide evidence that punishment communicates important social information; children in particular seem to learn a great deal from punishment policies, including what actions are considered “bad” and who in society is worthy of esteem. Finally, we have shown that understandings of punishment are evolving throughout childhood, and that by the age of eight, children are beginning to understand the importance of punishment to social order.

In a broader sense, our results also provide suggestive evidence about the nature of punishment reasoning. Our results indicate that the roots of such reasoning run deep, though many adult notions about punishment do not seem to be central to early intuitive reasoning. Further research is necessary, but our findings suggest that people are capable of complex social reasoning about punishment from a very young age, and that reasoning influences our expectations about the future behavior of others. Importantly, these results do not suggest that children lack retributive instincts; our experiment was not designed to elicit retributive responses, and we do not view consequentialism and retributivism as mutually exclusive. Indeed, it seems likely that children and adults harbor intuitions about deterrence and retribution simultaneously. Future work could further describe the way in which childhood concepts of punishment evolve into adult beliefs.

Our results demonstrate the promise and potential of the intuitive jurisprudence approach to law and psychology. This approach can provide a richer and more detailed understanding of fundamental questions of human behavior that are relevant to the law, such as how people justify punishment and how we assign blame. The study of lay legal intuitions has proven to be very useful in understanding how people interact with the legal system (e.g., Hans 2003; Malle & Nelson 2003), and intuitive jurisprudence provides a unique new way to consider these questions.

Because children develop social intuitions over time as they develop, this method also allows us to isolate different facets of social reasoning, giving us a more precise understanding of the psychology of moral and legal thought. By comparing children across development, we can learn more about the deepest psychological influences on our beliefs and behavior. In a broader sense, children may serve as models for the way a society comes to agree on social and moral rules; their understanding evolves over time as they develop more nuanced and comprehensive ideas. By incorporating the methods and foundational knowledge of developmental psychology, researchers interested in law can address deeper and more complex issues involving the origins and development of legal intuitions.

REFERENCES


Intuitive Jurisprudence 717


