INTUITIONS OF BLAMEWORTHINESS AS A HEURISTIC THAT EVALUATES THE PROBABILITY OF THE OFFENDER COMMITTING FUTURE ANTISOCIAL ACTS

BENJAMIN BUMANN\textsuperscript{1,2} AND DAVID M. EAGLEMAN\textsuperscript{2,3}

I. INTRODUCTION

In an economic model of crime, the costs and benefits that are associated with committing a crime can be partitioned into a series of factors, such as social costs, material gain from the act, fear of retribution, state punishment, and several others. An understanding of the values that an offender places on the underlying variables would tell a great deal about how likely an offender is to recidivate. However, because these variables are private, they can only be estimated by inference. We argue that people have evolved behavioral heuristics to roughly estimate the utility functions of norm-violators in our societies and that the output of the heuristic is our sense of blameworthiness. In other words, the degree of blameworthiness serves as an unconscious estimate of another actor's assumed utility function; those with a high likelihood to recidivate induce feelings of higher blameworthiness. In this way, blameworthiness has served a crude but effective evolutionary role in directing punitive action towards offenders in proportion to their recidivistic potential. In this article, we present evidence from the behavioral sciences and from analysis of the American legal system that support this model. Additionally, an alternative to our theory is put forth, but is shown to fail at explaining people's intuitions of blameworthiness.

II. MODELING THE UTILITY FUNCTION FOR COMMITTING A CRIME

The expected utility an individual receives for committing a crime can be modeled and used to predict his likelihood of committing the crime. The

\textsuperscript{1} University of Pennsylvania Law School.
\textsuperscript{2} Initiative on Neuroscience and Law, Baylor College of Medicine.
\textsuperscript{3} Department of Neuroscience and Department of Psychiatry, Baylor College of Medicine.
utility function is built from the subjective costs and benefits of performing the act.\textsuperscript{4} The benefits include the material, social, or psychic gains that would result if the crime were successfully completed—for example, the wealth, social prestige, or satisfaction that may result from completing the crime. Each individual has his own value for each of those factors, which may be greater than, less than, or the same as the average member of society. The realization of these benefits are in part dependent on the crime being initiated and/or completed, events which are themselves in part dependent on the situation and the characteristics of the criminal.\textsuperscript{5} To arrive at the expected value for the benefits of committing a crime, one multiplies the value of each benefit by the probability that the benefit will be realized if the crime is attempted. The costs include those factors that are dependent, to varying degrees, on being discovered, caught, and/or convicted. These include retaliation from the victim, state-imposed punishment or compensation, moral stigma, and loss of social standing. There are also psychic costs that depend on the individual’s self-realization that he has committed (or desires to commit) a crime—for example, the shame that flows from non-conformance to the social group.

All of the above costs and benefits are subjective, making their values difficult to ascertain. Nonetheless, estimates of these values for an individual can give insight into future behavior.

We will use the following equation as a basic illustration of the preceding intuitions:

\[
\text{Utility (crime)} = \text{Expected Benefits} - \text{Expected Costs}
\]

Where

\[
\text{Expected Benefits} = (\text{Probability of Material Gain}) \times \text{Value of Material Gain} + (\text{Probability of Social Gain}) \times \text{Value of Social Gain} + (\text{Probability of Psychic Gain}) \times \text{Value of Psychic Gain}
\]

\[
\text{Expected Costs} = (\text{Probability of Material Cost}) \times \text{Value of Material Cost} + (\text{Probability of Social Cost}) \times \text{Value of Social Cost} + (\text{Probability of Psychic Cost}) \times \text{Value of Psychic Cost}
\]


\textsuperscript{5} \textit{Id.}
If the utility for this particular crime is higher than the alternative action of highest utility, the crime should be undertaken.

Increasing or decreasing any of the subjective factors in the function affects the perceived utility of the act, and thus the likelihood that the individual will undertake it. Let us assume that the majority of individuals in society have utility functions that would cause them to not commit a crime in a certain situation. If an individual commits the crime in that situation, we can assume his utility for stealing the bread is outside the normal range. For example, imagine that the average person walks into a small grocery store and notices a loaf of bread. He would reject stealing the bread because the net costs to him are greater than the expected gain. If an individual steals the bread, we can infer that his utility function at that moment differed from the one that would have been held by the majority of society. We will argue that in the context of crimes, it is important to know which of the factors an individual abnormally values, the extent of the abnormality, and the cause of that abnormality—because these factors can imply different likelihoods of recidivism. We suggest that the larger the abnormality of a criminal's value, and the more transferable that value is to other situations in which the criminal may find himself, then the greater his probability of recidivism. Once we have set this bedrock, we will then argue that the sense of blameworthiness exists to give members of a society a rough estimate of the utility function of antisocial actors.

A. Transferability

An abnormal value associated with a cost or benefit may mean that the offender holds an abnormal utility just for that particular situation; on the other hand, it may imply the offender has abnormal values that apply to many situations, modifying the output of many of the offender's utility calculations. Knowledge of this matters for determining how likely the offender is to commit future offenses. For example, imagine John has an abnormally low sensitivity for defrauding an innocent person. This may simply mean that he is relatively unconcerned that an immoral act occurred when that act is fraud, and this would be a very specific class of this low sensitivity. Or, given other data, we may be able to infer that he is relatively unconcerned when he commits any nonviolent criminal act, a

---

6. We acknowledge that an atypically prosocial person would be considered to have an abnormal value scheme, but in this article we will only be concerned with statistical abnormalities in the antisocial direction.
somewhat more general class. Or most globally, we may be able to infer that he is relatively unconcerned by committing any crimes.

The more general the abnormality we infer, the more likely it will be that an offender's utility calculations will be similarly abnormal in other situations. We will use the term "transferability" to denote the extent to which one can transfer an abnormal value from the utility function at hand into other utility functions the actor has or will have.

For illustration, we will now consider how changing each variable from a normal range, even while keeping the others within a normal range, can increase the likelihood that an individual commits a crime. For each variable change, three different causes will be presented. Each cause lies on a spectrum from completely transferable to completely specific to that situation. While the completely transferable cause implies the individual's variable will be abnormal in many situations, the completely specific cause implies the variable will only be abnormal in the specific situation at hand. The causes are presented from transferable to specific. Once again, keep in mind that cause transferability is a spectrum, and not just three points as we have used here for illustrative purposes. The transferability of a variable's value may lie at any point from completely transferable to completely specific.

An individual walks into a small grocery store and sees a loaf of bread. His likelihood of stealing it is increased if he is an individual with an:

1. Abnormally high psychic pleasure from committing the crime:
   
   Highly transferable cause: Is addicted to breaking the law in order to feel "tough."
   Less highly transferable cause: Is a kleptomaniac with an "irresistible impulse" to steal.
   More specific transferability: Is a kleptomaniac whose condition is only triggered in grocery stores.

2. Abnormally low cost of state-induced punishment:
   
   Highly transferable cause: Abnormally low cost for all state punishment.
   Less highly transferable cause: Abnormally low cost for punishment related to stealing only.
   More specific transferability: Abnormally low cost for stealing bread only.
3. Abnormally low cost of victim retaliation:

Highly transferable cause: For all victim retaliation.
Less highly transferable cause: For victim of theft retaliation.
More specific transferability: For victim of bread theft retaliation.

4. Abnormally low social network costs:

Note: A low social network cost may be due to the fact that the social group the subject most closely ascribes to has norms different than the mainstream society's norms. As a result, an individual often has lower expected social costs than would members of the larger society. This phenomenon is also generalizable, as illustrated by the following example:

Highly transferable cause: Ascribes to a social network that condones all crimes.
Less highly transferable cause: Ascribes to a social network that condones most petty crimes.
More specific transferability: Ascribes to a social network that condones stealing but not other crimes.

Of course, a subject may also have an abnormally low cost within his most ascribed social group:
Highly transferable cause: Is insensitive to social pressure for all antisocial acts.
Less highly transferable cause: Is insensitive to social pressure for stealing.
More specific transferability: Is insensitive to social pressure for stealing bread.
Similarly, a subject may have an abnormally high (or low) probability estimate of successfully realizing potential costs and benefits. For example:

5. Abnormally high probability estimate of realizing benefits:

Highly transferable cause: Believes he has the ability to successfully complete any act.
Somewhat highly transferable cause: Believes he has the ability to successfully steal anything.
More specific transferability: Believes he can steal any loaf of bread.
B. Presumptive Transferability of Variables

As the reader may have noticed, some of the abnormal values above seem presumptively, if not inherently, transferable while others seem presumptively nontransferable. This is because many of the variables above are situation-dependent to a greater or lesser extent.

As one example, if a kleptomaniac feels a powerful compulsion to steal, we do not infer from this that he must have a similarly high subjective benefit for committing rape. Clearly, facts may surface demonstrating that the kleptomaniac stole in this instance not because of his kleptomania, but because he loves the feeling of committing all antisocial acts, and thus his value should be transferable to other situations. But this seems like an unlikely state of affairs, and the presumption should be against generalizing an abnormally high value into other situations without evidence.

Other costs may be assumed to change situation by situation. For instance, if an individual believes that robbing others is okay and that the larger society is wrong in its condemnation of the act, he may have an abnormally low social influence cost. However, it is possible that he agrees with society’s moral and norm schedule on all other points and has normal utility function associated with these other acts. Nevertheless, it is intuitive to assume that if an individual has abnormal values for this cost in one situation, he is likely to diverge from the norm in other situations as well. In other words, transferability can be presumed: his theft is not likely to stay confined to this store, but is likely to occur elsewhere.

To summarize so far, the decision to commit a crime depends on many variables in the form of subjective costs and benefits. Because crimes are usually acts which the average person does not commit, we can infer that an individual who does commit a crime has a different utility function for that situation than would the majority of society’s members. An individual’s variables may have different values due to different causes, some of which are specific to the situation in which he committed the crime, and some of which are transferable to other situations. Fact finding, along with presumptions of transferability, may help us decide which kinds of causes are responsible for his abnormal variable values. Ceteris paribus, an abnormal value’s likelihood of causing recidivism is dependent in part on the magnitude of its difference from the normal range and its transferability.
III. SOLVING FOR A LAW BREAKER’S ABNORMAL UTILITY FUNCTION

The utility function that results in a criminal action will not necessarily be due to a single abnormal value. Indeed, a plurality of abnormal values, each of different magnitudes and transferability, may be at play. To determine a criminal’s recidivism, it is important to determine as much of the utility function he had for committing the crime as possible, including its causes.

A. Fact-finding and Inferences Can Be Used to Solve for the Subject’s Utility Function

As an illustration, we will assume John and Valerie are coworkers at a publishing company. One day, during an office meeting, Valerie disparages John’s work product in front of five others. Later that day, John shoots Valerie to death in her office. Based on these facts, we can begin to construct an initial abnormal utility function for John. First, we must construct the approximate utility function that the majority of society would have for the decision of killing Valerie. The average person may derive a small psychic benefit from satisfying some feelings of revenge. However, the sum of expected costs would be so large that the average person would not commit the crime. The psychic, material, and social costs associated with murder (and being a murderer) are likely considered large by the average person, as are the employment opportunities lost by being classified as a murderer. The potential retaliation from Valerie or from non-state actors adds to the overall costs of the action.

What value abnormalities can we infer about John? At this point, many different abnormal value schemes would cause the result. We can get rid of a few, however. Because the magnitude of the difference in value schemes from an average person, we can get rid of value schemes that cannot explain the difference. For instance, assume John had no fear of retaliation from Valerie, but was normal on all other values. Intuitively, we would conclude that a zero cost on that variable still does not make killing Valerie a good move for most people, as other costs would still be too high.

An additional zero cost on state punishment likely will still not make the crime effective: The average person, even if miraculously not convicted or charged by the state, would still not kill another, due to the social costs involved. Thus, we assume that low social influence costs or high value to commit the murder play a causative role in John’s decision to kill Valerie.
Now assume that John feels truly horrible and guilty about committing the murder as it was planned and after it was committed. Assume he feels a drive to kill Valerie that is so intense he has to relieve it, even though he hates himself for it. Here, we may be able to say that his value for committing the act was very high, and his social influence values were not necessarily abnormally low.

Now, assume that John does not feel the irresistible impulse to kill Valerie. Assume he instead has the same benefit to killing Valerie that the average member of society would have, i.e. a somewhat small one associated with satisfying revenge impulses and reasserting himself. We must assume that his social influence values are lower than average in this situation, but which ones, exactly? Even with no desire to conform to the larger society, we intuitively believe that the average person would still not commit the crime. Even assuming there is no state or victim-derived punishment; average people tend toward not committing antisocial acts that cause them to suffer social damage or psychic discomfort. Although all the aforementioned value changes would increase the likelihood of the average person killing Valerie, we infer that low total values on social and/or psychic costs are needed as well.

Indeed, we may be able to infer the general magnitude of John’s psychic discomfort difference from normal: Imagine the facts were changed and John killed Valerie two days after she insulted him, planning the murder the whole time. From this we may infer the opportunities for him to feel psychic discomfort related to his plan are greater than they are in the first situation when he reacted more immediately. We may infer a low psychic discomfort for killing Valerie attended a variety of situations and was deliberated. From this, we may more safely presume that the psychic discomfort value present in this example is transferable to many situations.

Suppose further that during those two days, John interacted a lot with Valerie, even seemingly working things out with her over a pleasant lunch, all the while planning to kill her. The psychic discomfort of guilt and shame would normally be enough of a pain to the average person otherwise planning to kill Valerie. In fact, this discomfort would most likely cause the average person to avoid the lunch. However, if the psychic discomfort of guilt and shame did not cause John to avoid the situation, we may infer an even greater deviation from the normal range on psychic discomfort.

What if since childhood John has belonged to a peer network that did not believe that killing someone who insulted you was immoral? Assuming that John belonged to that social network, as well as to the larger society,
we should investigate the proportional importance of each social network to him. All else equal, if John ascribes to a peer network that finds the act acceptable, he will have a lower social cost of committing the act than if he heavily ascribed to a social network that is, or behaves exactly like, the larger society.

Furthermore, note the recidivism potential is different between the following situations: If the cause of John's low social costs is due to low concern about any social networks or any conforming, then this abnormal variable is transferable to many situations. If John wholly ascribes to a peer group that allows murdering those who insult you, but on all other crimes has a morality schedule similar to the larger society, his abnormal value is only present in situations involving murdering someone who insults you. What if John knowingly joined this group as a mature adult, fully aware that it condoned murdering in response to insult? Then we may look to his decision to join such a peer group in order to calculate his utility function. In that case, we may infer a more transferable abnormal schedule than we can in a situation where he was raised and socialized by this group.

Suspend disbelief and assume John was raised from birth by a mad scientist that socialized him to think that breaking every one of our society's laws was the right thing to do. Thus, one may say that his abnormal schedule is transferable in all situations. However, it is still predicted that we would blame him less than if he did not have this socialization, because it is not clear that he does not attach a strong subjective value to social influences generally. Indeed, it is possible that his abnormal schedule is not transferable to a situation in which he must conform to our society's norms, and if we freed him from the madman and bathed him in our socialization, he may have offense rates approaching normal. Indeed, this mechanism may also be involved in reducing culpability for the gang members described above.

What if John is thirteen years old? Juveniles have less ability to control impulses than do adults.\(^7\) Consistent with this, the prefrontal regions of adolescents are not fully developed, and these areas are recruited

---

for deliberative decision making. Similarly, it is known that juveniles are more influenced by, and dependent on, their environment than are adults. Therefore, even if his utility function differs markedly and generally from an average person’s utility function, or even an average thirteen year old’s utility function, we cannot be sure the abnormal schedule will transfer to adulthood. Indeed, we cannot even be sure that the schedule will transfer to situations that are more straightforward, easier to understand, or are less sensitive to environmental factors. What if John is mentally ill to the point of not being able to understand the nature of his act? At that point, we cannot infer anything about his utility function because we are not sure what situation he is reacting to (as it exists in his mind only), and thus cannot know whether the decision he made (given the situation he believed he was in) is different than the decision an average member of society would make in the same situation. If John is mentally ill to the point that he does not understand that killing Valerie is considered wrong by society, then we must admit that we cannot infer that he has low social influence costs, only that he believes society has value judgments much different than it has in reality. In neither type of mental illness can we infer that his abnormal valuations for the action are transferable outside of his mental illness. Additionally, mental illness may not even be transferable to other situations that the mentally ill person comes across while ill. The illness may be sporadic, or triggered only by certain situations, and thus we cannot infer transferability as easily as we can for the equivalent, sane individual.

What if John attempts to shoot and kill Valerie by shooting her at close range, but right before the bullet hits her chest, a police sniper fired-bullet miraculously knocks John’s bullet out of the way? What inference can we make about the John who kills Valerie that we cannot make about the John in this situation? We have no idea if the John in the attempt situation, would not have immediately regretted shooting her and attempted to save her life by performing CPR or rushing her to the hospital. We have no idea if the John in the attempt situation would have broken down crying in guilt and shame. We do know that the John in the completed scenario does not show any of these redeeming features. Either way, we have more facts to


infer great and transferable value abnormalities from the completed situation than we do from the attempt situation.

What if John was mentally slow and monumentally underestimated the probability of being caught? Compared to a John who estimated being caught like the average member of society would, we must infer his predicted social influence costs to be lower than normal, but not because of a general unconcern with social influence. Problems estimating probability of an event occurring are likely transferable to many situations with similarly complicated scenarios. However, scenarios where the probability of being caught are very clear, such as when an authority figure is in proximity, may be simple enough that a mentally challenged brain will be able to make the same predictions as a normal brain. Furthermore, John’s mental retardation may impair his ability to understand whether something is a crime, or the magnitude of its benefit. If this occurs in a relatively complex situation, we cannot infer that the abnormality is transferable to simpler situations. Thus, low social costs seem to be much more transferable than low probability estimations.

B. Members of a Society Solve for a Criminal’s Abnormal Utility Function by Using a Heuristic in the Form of Blameworthiness Intuitions

1. What Motivates a Need to Identify Utility Functions?

A society may derive significant benefits by knowing which of its members has abnormal utility functions. Compared to people living in the counter-factual world, the overall transaction costs for society members would decrease in most aspects of their life for two related reasons. First of all, the number of potential antisocial actors in the world would go down, assuming they were warehoused or eliminated upon discovery. This would reduce the probability of encountering them, so the expected cost from their harm would decrease. Second, because many of these abnormal individuals are difficult to detect and avoid, they increase the uncertainty of any transaction. Reducing the frequency or magnitude of deviance of abnormal schedules in the population would ameliorate this uncertainty cost.

A related and less obvious benefit of identifying abnormal utility functions involves reducing the overall cost of enforcing the law. Individuals with abnormal schedules force state punishment to be more expensive than it would be if the population was composed of people with low variance around a “normal” utility function. This is because people
with approximately normal utility functions do not need to be threatened with as much, if any, state induced punishment to deter them as compared to individuals with abnormally low social influence costs or high action benefit. In order to deter these latter individuals, the state must threaten and make good on more punishment than it would have to with the former group. This poses potential problems.

If the state wants to deter crime at high enough levels that it is necessary to generally deter some of those abnormal individuals, it will have to punish the crime at a higher level than if it just needed to deter those individuals in the normal range. If the state was not able to identify and separate each group beforehand, it would end up over-punishing many individuals who require less state punishment to be generally deterred. Identifying the utility functions of different offenders can help with this problem. First of all, some of the increased cost could be defrayed by incarcerating some offenders less than the majority. This way, the deterrence of the crime would still be stronger and centered around the more severe punishment but would cost less to implement.

Of course, when deciding which offenders to incarcerate less, we would want to minimize expected cost. The expected cost of a higher risk offender is greater for each day not incarcerated. Thus, at sentencing, being able to differentiate offenders by their likelihood of recidivism is important.

Identifying and punishing those with abnormal values is justified from an evolutionary perspective as well: A species that depends on its ability to adapt to new circumstances as a group and engage in social cooperation would benefit from a mechanism that helps ensure its members all march to the sound of the same (likely to change) drum beat.

Of course it would be very costly, even if possible, to go through a society and identify who had abnormal utility functions. Instead, a preselection process can efficiently deliver the majority of these individuals to the law, giving the opportunity to warehouse (or even eliminate) them. This may be manifested in the punishment phase of a criminal proceeding. Abnormal schedules are probably disproportionately represented as convicted criminals. Furthermore, although many individuals who fall within the normal range on their utilities, or are weakly abnormal, may commit moderately antisocial acts, and the overall risk they pose can be inferred from small infractions. These infractions are relatively low and will likely be more than offset by the benefits they render as otherwise productive members of society. Indeed, only sentencing those that have committed a crime ensures a large proportion of the offenders will have
recidivistic utility functions worth addressing. Thus, the sentencing phase is an opportunity for society to determine the extent to which these individuals deviate from the norm, and how transferable their deviation is. It also situates a sentencer to effectuate his determination through differential punishment.

2. Summary So Far:

Considering the benefits, we should not be surprised if people had a mechanism to solve for and differentially incarcerate or reduce the fitness of abnormal utility functions. With limited resources, we should also expect the heuristic to differentially direct blameworthiness impulses depending on the offender’s likelihood of committing a future antisocial act. This likelihood in turn should be based on a combination of the magnitude of their deviance from the norm and the transferability of that deviance. Keeping in mind the difficulty of gathering this information, we should expect people to do this when most efficient, such as when evaluating an individual who was determined to have committed an antisocial act. We will now argue that people do this by employing a heuristic in the form of their sense of blameworthiness, what we will call the “Utility Heuristic.”

C. An Alternative to the Utility Heuristic: The Act Heuristic

It should now be noted that there is a potential alternative explanation for much of the evidence we will present below. It is possible that blameworthiness is a heuristic by which individuals learn from their community and/or are predisposed by evolution to regard certain acts to be more harmful, and to punish those most severely. We will call this the Act Heuristic, as it holds that blame targets the act itself and holds individuals who commit it blameworthy as a tool to deter the occurrence of the act.

The Utility Heuristic recognizes some of the wisdom of the Act Heuristic. For instance, it is quite likely that individuals find certain general types of crimes (i.e. murder, robbery, arson) more blameworthy than others because society established the offense as more prohibited. The Utility Heuristic, however, would focus on the fact that someone who commits the act can be inferred to have a more abnormal utility function than someone who commits what society considers a less serious offense. For instance, murder is seen as more blameworthy than robbery, all else
equal. In order to accept the Utility Heuristic, one does not need it to explain why society considers one general type of crime more serious than another. This is because the heuristic is meant to identify the type of people whose value scheme is most abnormal and transferable. While this does involve investigating the type of offense committed, we will see that many other factors are relevant. For instance, a kleptomaniac who steals a diamond necklace from the mom-and-pop jeweler may be found less blameworthy than a non-afflicted individual who shoplifts a pair of socks because he does not want to pay and has low social influence or state punishment expected costs. Furthermore, society often finds offenders who committed the same crime to be differentially blameworthy. Both of these points weigh against blameworthiness being explained by the Act Heuristic.

Nevertheless, advocates of the Act Heuristic may say that a diamond necklace stolen by a kleptomaniac is a less serious or harmful event than the sock-stealer's offense. They may also say that it can explain why offenders who commit the same offense are found differentially blameworthy. The heuristic may be attempting to deter harmful acts surrounding the crime that one offender committed and the other did not. Admittedly, much of the evidence below appears to be predicted as well by the Act Heuristic as by the Utility Heuristic. However, a significant amount is much more difficult to explain by this alternative. Indeed, it seems that a large part of blameworthiness and the legal doctrines that spring therefrom serve to identify and eliminate abnormal utility functions most likely to recidivate.

When possible, we will point out evidence that the Act Heuristic hypothesis has difficulty explaining and that the Utility Heuristic hypothesis does not. Thus, while intuitions of blameworthiness may have in part an act-deterrence function, we will demonstrate that they seem to be more focused on identifying abnormal schedules and reducing recidivism.

1. Referencing Norms

The Utility Heuristic holds that when people are deciding blameworthiness it is crucial to reference the norms and morals of the relevant social groups. In a study conducted by Gollwitzer, subjects who self-identified with their social network were more likely to display a retributive sentencing philosophy than subjects who did not identify with

---

their social network. Similarly, in another study, researchers found that placement on the psychopathy spectrum were negatively correlated with the likelihood of retributive judgment-making. A possible explanation for this is that psychopaths have deficits in applying society’s concepts of right and wrong in conducting their own behavior. It would be difficult for psychopaths to calculate how far an offender’s psychic discomfort during and before a crime must have varied from that of the average member of society, as they have little basis to know how a normal-range individual should feel about the situation.

Determinations of blameworthiness often depend in part on the society in which the act occurs. For instance, Kahan shows that societies differ in their condemnation of individuals who engage in acts such as marijuana use and prostitution; with blameworthiness increasing as that society’s norm differs from the activity. Additionally, Kahan discusses research which showed that Kuwaitis and Americans differ markedly on the punishment they believed different offenses deserved. For instance, Kuwaitis believed that someone who throws a rock at another, intending to lightly injure, yet kills the person is far less blameworthy than an American would hold the offender to be. This may very well be due to the fact that rock throwing is a much smaller deviance from Kuwaiti norms than it is from American norms. Similarly, Kuwaitis hold female adulterers to be more culpable than are men who kill cheating spouses.

Even in America, we see differences in cultural norms correlating with differences in expressed culpability. Kahan presented a study in which both Midwesterners and Southerners judged the culpability of a man who used severe violence against another man who had made advances on his wife.

11. Id. at 471–72.
15. Id. at 1558.
16. Id.
17. Id.
18. Id. at 1582.
To a significant degree, Southerners found the man to be less culpable than did Midwesterners.\textsuperscript{19} Kahan attributes these differences to Southerner’s placing greater average value on defending one’s honor than do Midwesterners.\textsuperscript{20}

As one would predict, an individual is more likely to commit an act if it is inline with the norm of his social network. Patachini and Zenou studied data on juvenile offenders who moved among different cities and found that they began committing the crimes that the new social network they moved into found normal.\textsuperscript{21} The authors controlled for confounding variables, such as common environment, and determined that the evidence supported the finding that the change in criminality was caused by moving in-line with the new group norms.\textsuperscript{22}

Additionally, Kelly et al. found that people were more likely to find “bad” actors—such as captains who whipped sailors, slave owners, and others—much less blameworthy if told the actor lived centuries before when the action was in line with community norms.\textsuperscript{23}

Is it possible that these examples demonstrate the Act Heuristic? For example, perhaps to figure out how much an act should be deterred and avoided, subjects in the above examples had to reference the norms of the social group. There is reason to doubt that possibility. Proponents of the Act Heuristic must accept that if perceived social norms seem to control blameworthiness judgments, then perceived social norms must strongly influence what actions individuals find necessary to avoid and deter. The above studies show that the Act Heuristic, if true, is fragile. If momentary changes in perceived social norms can cause people to perceive an act to be less costly, as in the Kelley experiment\textsuperscript{24} above, then it seems that they could all too easily stop avoiding and deterring a still-dangerous activity. Similarly, new dangers an individual faces from the environment may very well outpace the ability of his group’s social norm to regard it as blameworthy. While individuals probably do rely in part on social norms to determine the dangerousness of an action, one would expect that individuals would not rely on this information so completely.

\textsuperscript{19} Id.
\textsuperscript{20} Id. at 1583.
\textsuperscript{22} Id. at 19 n.31.
\textsuperscript{23} Daniel Kelly et al., \textit{Harm, Affect, and the Moral/Conventional Distinction}, 22 MIND & LANGUAGE 117, 123 (2007).
\textsuperscript{24} Id.
On the other hand, the Utility Heuristic’s goal is to determine the antisocial potential of others in society. To do this, modifying blameworthiness for an actor based on the social norms that apply to him, and thus being sensitive to a temporary change in perceived norms such as in the Kelley experiment, is appropriate. Additionally, if the social norms of society do not keep pace with the dangers of the environment, referencing them is still relevant to identifying antisocial actors. Furthermore, the error that can flow from a false negative would lead to missing the presence of an actor with a high chance of committing future crimes. While this is a costly error, it seems less serious than dropping one’s guard for an entire class of dangerous action. In conclusion, the Utility Heuristic seems more suited to use social norm information than is a heuristic that uses this information to infer the inherent potential harm of different acts.

Of course, there are mechanisms by which individuals hold different acts as differentially dangerous and thus worth deterring. Our position is just that intuitions of blameworthiness, given the above argument as well as evidence below that show the blame for an actor depends on his behaviors surrounding the crime, appears ill-suited for that role. Intuitions of blameworthiness, on the other hand, seem well designed for the role our hypothesis believes it plays.

2. Calculating the Abnormal Utility Functions’ Deviance from the Normal Range and Transferability

In a study by Garberg, subjects read the following hypothetical: Four robbers planned out the robbery of a small store. They each armed themselves, then drove to the store. The getaway driver stayed in the car, while the lookout stood outside the store’s door, and two of the robbers went inside. During the robbery, the store owner was accidentally shot and killed by one of the robbers inside. They were all charged with felony

25. Id.
27. Id. at 560.
28. Id.
29. Id.
30. Id.
murder. As expected, subjects convicted the actual shooter of the felony murder more than they did the other three accomplices. Further, subjects convicted, in order from highest to lowest rate, the other robber who was inside the store, the lookout, and the getaway driver. Our heuristic may infer greater abnormal valuations for the lookout than for the getaway driver because the lookout was willing to get out of the car, which provided anonymity, safety, as well as a quick escape, and put himself closer to where the crime was taking place. The result here is not surprising, considering that subjects report that one of the most important factors when deciding culpability is the amount of effort a criminal puts forth during the crime. This may also help explain why subjects considered a thief who went into his neighbor’s unlocked detached garage to steal an electric drill was less culpable than a thief who broke into his neighbor’s house to steal a microwave. Although both crimes are similar episodes of burglary, the extra effort involved in breaking into the kitchen, along with the fact that the second thief was undeterred by invading a more personal space, allows us to infer a more recidivistic utility function.

A pair of hypotheticals from Robinson and Kurzban are also illustrative. In the first situation, a woman highly offends John. A day after he finds out where she lives, he goes to her house, waits for her to come home, and shoots her to death. In the second situation, a woman tells John’s boss about unspecified misdeeds. John then devises a plan to kill her. A week later he forces her into his car at knife point, then drives her to a remote place where he shoots her to death. Subjects rated John more culpable in the second scenario, as the blameworthiness heuristic would cause us to predict. Both crimes are heinous, premeditated murders. However, in the second crime, John spends a longer time thinking

31. Id.
32. Id. at 562.
33. Id.
36. Id.
37. Id.
38. Id.
39. Id.
40. Id.
41. Id. at 1876.
about the murder, and interacts much more with his victim. Thus, we can infer lower psychic discomfort in the second case than we can in the first.

Previously in this article, it was suggested that prior criminal history might be useful for inferring the transferability of an abnormal schedule. As expected, subjects rate individuals who have committed crimes in the past as more culpable than those who committed the same crime yet have no criminal history. It is possible that this intuition is codified in the many states that use sentencing guidelines to sentence offenders. These guidelines often use previous criminal history as a main factor to determine the recommended sentence. The Act Heuristic does not seem to explain this evidence, as an offender's previous convictions should have no bearing on the harm of his crime.

Indeed, the Act Heuristic seems to have difficulty explaining much of the above studies. Admittedly, it is possible in theory that the act of being a lookout or getaway driver may have differential harm to society. However, it is unclear what principles may explain these differences. The Act Heuristic cannot make predictions like the Utility Heuristic. Until it can, the Act Heuristic seems only able to make post-hoc explanations of data and is thus a less useful model than the Utility Heuristic for predicting intuitions of blameworthiness.

The Utility Heuristic may also explain some of the reasons why judges employ downward departures from sentencing guidelines. Kramer and Ulmer analyzed Pennsylvania courts to investigate patterns of downward departures among the state's judges. They found that young Hispanic men were less likely to receive downward departures than any other group, and speculated that this was due to their being perceived as more dangerous. Similarly, men were half as likely as women to receive downward departures from judges. Furthermore, judges were interviewed

44. Id.
45. Id. at 916.
46. Id. at 915.
regarding their reasons for giving downward departures. One important factor was the
victim’s relationship to the offender. Offenders who had close relationships with the victim had a higher chance of receiving a
downward departure. This may be explained by making the reasonable
inference that crimes committed against a loved one do not imply abnormal
values as transferable to other situations as would be implied by crimes
against a stranger. Other reasons that judges gave for the downward
departures included the perceived dangerousness of the offender, a
perceived drug problem, or a perceived mental problem. The last two
reasons are in line with the Utility Heuristic, as we will see, because an
offender’s utility function cannot be inferred as strongly if mental illness or
drugs seem to be playing a causative role in the offense. Furthermore,
downward departure rates differ by county. The authors suspected this
was in part due to different norms in different counties finding offenders
less blameworthy, as the blameworthy hypothesis would predict. The
results found in this study would be predicted if judges used a Utility
Heuristic when deciding whether to depart from a crime’s sentencing
guideline.

As mentioned earlier, those with high values for committing a crime
will not be inferred to have low costs without further evidence on the fact.
Many of us may feel less retributive anger toward a murderer if we are
convinced that he was suffering from irresistible impulses to commit his act
and he felt that if he did not do it something terrible could happen. As
previously discussed, his value for the crime is presumed to not be as
transferable to other situations as low costs would be. If the murderer also
expresses deep guilt and shame about his action (and we believe him), we
may not infer that he has generally low social and moral costs for
committing the crime, but rather a high value. Indeed, when certain kinds
of sex offenders, such as child molesters, are viewed as responding to a
compulsion they cannot control and feel self-hatred about, we may have

47. Id. at 918–21.
48. Id. at 921.
49. Id.
50. Id.
51. Id.
52. Id. at 923.
53. Id.
54. See generally Mark D. Alicke, Culpable Control and the Psychology of Blame, 126
55. Id. at 556.
less desire to hold them blameworthy and more of a desire to chemically castrate, counsel, warehouse, or otherwise reduce the danger of the abnormally high value. If the sex offender does not seem driven by an abnormally high value as much as low costs, because those are more transferable, we may feel confident in our blame and punitive instincts towards him.

The Utility Heuristic predicts even less retributive impulses towards an offender whose mental illness is such that he does not understand that his act was wrong. As explained above, we cannot infer that the offender’s abnormal value scheme is transferable to situations not burdened by his mental illness, or even to other situations while he is mentally ill. Though we desire that the offender be potentially warehoused in a mental institution, if it is ever decided that the offender is no longer a danger, he will be released.56 This is a relatively efficient outcome, as there is no reason to waste money detaining someone with utility functions that are now likely normal. Indeed, the principles of rehabilitation and parole for the non-insane offender may be appealing for the same reason.

Like the severely mentally ill, juveniles are generally considered less culpable for bad acts than sane adults.57 Because the juvenile brain is not fully matured, and because the juvenile’s decisions often have a greater potential to be influenced by the environment (such as family and peers), the blameworthiness heuristic may not transfer a juvenile’s abnormal schedule into adulthood or situations where these deficiencies are not factors.58 Of course, if a juvenile deviates greatly from the norm, or is close to age of majority, it may be safe to assume a recidivistic utility function is transferable to adulthood and most other situations.59 As expected, the law works this way, generally reducing the culpability of juvenile offenders, and often releasing them from detention when they reach the age of majority.60 Similarly expected, the law is harsher on juveniles who commit truly heinous acts, especially if committed by a juvenile close to age of majority.61 In these cases, juveniles are often held culpable enough to be tried as adults.62

58. Id. at 238.
59. Id.
60. Id. at 230.
61. Id. at 231.
62. Id.
Similarly, most would agree that the mentally retarded are less blameworthy than their intellectually normal counterparts who commit the same crime, as long as the mental retardation affected their ability to evaluate the consequences of the act. As discussed earlier, this may be due to the fact that their probability estimates for costs and benefits are abnormal, which is presumptively less transferable than other factors, such as low social costs. It may also be because, like the insane, they are not aware of the nature of their act, or that their act is considered wrong, a problem that may not translate to other, simpler to understand crimes and circumstances.

As discussed previously, the Utility Heuristic finds that low social influence costs leading to an act, if less transferable, are seen as less blameworthy than their more transferable counterpart. If a petty-theft offender is raised in a social group that condones petty thefts but no other crimes, we should not transfer those abnormal value differences to all crimes, but only petty crimes. Intuitively we believe that someone who grew up in a society that tolerated petty crimes but not more serious ones, such as someone who was raised in an inner city gang, is less morally culpable for petty theft than an equivalent actor whose social network does not tolerate the theft. Furthermore, it seems intuitively clear that most people would hold men whose only social network is the larger society to be more blameworthy for “marrying” and having intercourse with a consensual thirteen-year-old girl than the exact same offender that was raised in and ascribes to a religion which promotes such an act.

In addition to these offenders’ schedules being less transferable, we cannot be sure that the offenders will not socialize with our group once they are taken away from their other group, and afterwards must calibrate to our group, as they live among us. The heuristic should find this offender less blameworthy than his counterpart raised in the larger society. Indeed, U.S. courts have used a defendant’s background as mitigating factors in sentencing. 63 However, we find an actor more blameworthy if he chose to enter a religion so that he could commit pedophilia. In that case, we use the opportunity to evaluate his utility function by looking at the utility function that led him to join the religion. We may reasonably believe he joined the religion because he had low social influence costs for the larger society,

which is a transferable condition that cannot be inferred if he was simply raised in the religion and had little choice to accept it as his social network.

Once again, the Act Heuristic has trouble explaining much of the above disparities in blameworthiness. This is because it is supposed to only care about differentially deterring events, and should not care whether the event was produced by a low or high transferability schedule. But above we see that the transferability of an abnormal schedule affects the blameworthiness of offenders who committed the same crime. For instance, should not theft by the above gang member raised in a condoning social network be just as serious as theft by an offender who was not so raised? Similarly, a kleptomaniac's theft seems just as serious as a non-kleptomaniac's. A building burned down by a fifteen year old is just as destroyed as if done by someone of the age of majority.

On the other hand, there may be plausible reasons why the Act Heuristic would differentially blame the above examples. For instance, an adult committing a crime may be more harmful than a juvenile committing the crime because adults need to set good examples for children. Or maybe when adults commit crimes, it is more likely to disrupt the social order. Therefore, we would blame the adult more than the juvenile. Then again, a similarly plausible opposite position could be taken: Juveniles should perhaps be deterred more than adults, because they are at the age when they are learning how to be members of society. Also, juveniles are often more impulsive and less concerned about lost employment opportunities than adults, and thus may need an increased threat of deterrence.

Similarly, the Act Heuristic might blame the above petty-theft criminal less harshly than someone who grew up in Greenwich, Connecticut, or someone with kleptomania who steals less than someone without the disorder. This results potentially because there is less of a threat to social order in the first cases and because those individuals are already outsiders and should not influence members of the larger society to act. However, other marginal groups, such as transvestites, do not receive reduced culpability. Forgive us if we have constructed straw-man arguments for the Act Heuristic. The point is simply that there is no clear reason why the Act

---

64. See Grisso, supra note 57, at 231.
65. Id.
66. Id.
67. Id.
68. Id. at 232.
Heuristic would differentially blame offenders who committed the same offense but have utility functions of different transferability. Indeed, many of the offender classes above seem to need more deterrence than the average person in order to not commit the same antisocial act. Thus, as mentioned above, the Act Heuristic presently has little predictive value.

The Utility Heuristic may also explain culpability in attempt crimes. Moral philosophers are often perplexed by the fact that people impute less culpability to an individual who attempts a crime, but fails, than to a criminal who completes the crime. This is so even if it is clear that the former offender tried just as much, and had as much intent, as the latter. Seemingly, luck should not bear on culpability. As mentioned above, however, in the latter case we know that the criminal did not immediately regret his action, or did not immediately try to fix his harm, or otherwise redeem himself. This allows the heuristic to more easily infer an abnormal schedule that it cannot infer with the criminal whose attempt was not completed. As expected, the criminal justice system also generally punishes attempt crimes less severely than their completed versions, possibly reflecting lay intuitions formed by the Utility Heuristic.

If the Utility Heuristic is an important component of our intuition of justice, we should expect our intuitions of justice to be offended if the law tracks away from the heuristic. A study by Robinson et al., found that among the biggest variations between people’s intuitions of just deserts and what the law actually sentenced were the examples of statutory rape where the offender reasonably believed the victim was of age, and selling drugs. In the first example, it is very hard to impute an abnormal utility function to the offender. Indeed, well-known lines of cases from the earlier part of American law used to not find a man guilty of crimes that occurred due to a reasonable mistake as to a female’s minority, unless immoral acts were associated with the event, in which case the judge found the offender should be incarcerated, as the blameworthiness heuristic would predict. In the Robinson et al. study, subjects rated the statutory rape case as deserving 2.9 years. Subjects were then shown how a state court handled that exact

70. Id.
71. Id. at 1940–2033.
72. Id.
73. Id. at 1964.
74. Id. at 1978.
75. Id. at 1972.
scenario in the real world. The court, treating this as a strict liability crime, punished the same offense from 40 to 60 years.\textsuperscript{76} In another example, subjects had to sentence a man that, during a routine traffic stop, was found to have a soda-can-and-a-half worth of cocaine in his trunk, but he had no prior record.\textsuperscript{77} Subjects on average sentenced him to 4.2 years.\textsuperscript{78} The actual sentence given by the court was life without parole; the same sentence as given for first degree murder.\textsuperscript{79} The law in these situations is likely tracking away from the blameworthiness heuristic.

In the Robinson paper, there are many more examples of the law giving offenders harsher sentences than the subjects did.\textsuperscript{80} There are also examples of the law not punishing when subjects determined that punishment was in order.\textsuperscript{81} Subjects believed that both posting negative false comments on a Facebook page and not reporting a neighbor who violated an ordinance limiting the amount of dogs in a house were acts worthy of punishment.\textsuperscript{82} The law, however, generally gives no punishment for these scenarios.\textsuperscript{83} The authors showed in the same study that when subjects are shown that the law incarcerates the perpetrator either much more or much less than they would, they are less willing to cooperate with the law.\textsuperscript{84} When subjects were presented with evidence that the law’s sentence comports with the sentence they gave, they were then much more willing to cooperate with the law.\textsuperscript{85} These results may be due to the fact that if people see that the law is not tracking with the Utility Heuristic, it is not doing a good job of prioritizing and eliminating the most dangerous abnormal utility functions.\textsuperscript{86} Thus, the law would be a dangerous institution to follow.

\textsuperscript{76} Id. at 1974.
\textsuperscript{77} Id. at 2030.
\textsuperscript{78} Id. at 1972.
\textsuperscript{79} Id.
\textsuperscript{80} Id.
\textsuperscript{81} Id.
\textsuperscript{82} Id. at 1999.
\textsuperscript{83} Id.
\textsuperscript{84} Id. at 2002–03.
\textsuperscript{85} Id. at 2003.
\textsuperscript{86} Id. at 2007.
IV. Conclusion

We suggest that blameworthiness is a heuristic that serves the function of identifying the individuals who are most likely to commit future antisocial acts. Going forward, the Utility Heuristic can be further tested. It predicts that many of our intuitions of blameworthiness, and legal doctrines that spring therefrom, are all ways to identify different abnormal utilities. Specifically, more blame is predicted to attach to those utilities whose abnormality can be inferred to have the greatest deviance from the normal range, as well as the greatest transferability to other situations. Additionally, our intuitions of blameworthiness and legal doctrines are predicted to reflect a desire to not want to incarcerate an abnormal schedule that has returned to normal.

Some other questions should make for further research. The question of which factor, magnitude of deviance from normal or transferability is more sensitive to eliciting blameworthiness is an important one that needs resolution. Another issue that needs further testing is the possibility of the Act Heuristic being responsible for the intuitions of blameworthiness in question. As of now, the Act Heuristic as a model for differential blameworthiness for the same act predicts little, explains few judgments and legal doctrines, and requires that people determine the harmfulness of events based on a social consensus (which seems like a particularly death-inviting way to determine what is dangerous in a world where new hazards spring up faster than a social consensus could seemingly be rendered).

Scientists and engineers tackling difficult problems from space flight to combating disease have long turned to nature to understand the general premise of what solutions work, and then improve upon those solutions for the need at hand. This same process can be used in the behavioral and neural sciences. Here, we suggest that human brains roughly solved the problem of identifying those most likely to recidivate by evolving a heuristic in the form of blameworthiness. How can we now improve upon that heuristic? First, as stated above, it will be necessary to better understand what cost and benefit factors are more important in recidivism calculations, which values are most likely to be transferable to other utility functions, and which normally have the biggest weight in a utility function. Once we understand these factors, it may be possible to estimate the utility functions of different offenders by using neuroimaging techniques, and thus approximate their recidivism potential. At the danger of oversimplification, scientists can use neuroimaging techniques, such as fMRI, to
gain insight into whether subjects feel pleasure for a particular event or fear or disgust for that event. It may be possible, at least in a rough manner, to piece together the benefits and costs that fMRI reveals a subject to have, and apply them to a framework that calculates recidivistic potential using those values. This method could help mitigate malingering and also would allow preferences to be established without waiting for them to be revealed through behavior, which may be too reliant on probabilistic inferences and take too long. If constructing utility functions is what humans are attempting to do through our intuitions of blame, then perhaps abandoning reliance on our intuitions for a more precise tool is a just decision.

Of course, simply showing that intuitions of blameworthiness seem to be heuristics for a consequentialist process does not mean that deontological retributive theories of punishment should be abandoned for a consequentialist framework. Retributive theories can be justified as what one ought to do by rationale independent of the ontogeny of the intuitions. However, we believe that demonstrating the intuitions of blame serve as a crude consequentialism will cause people to be less willing to give undue credence to their intuitions. This should encourage many to undergo a deliberation on whether to adopt retributive or explicitly consequentialist punishment distribution schemes, or even stick with whatever their intuitions happen to be. Nevertheless, the conclusion arrived at, whatever it may be, will likely be more efficient and humane as a product of the deliberative process.